TASO Transforming Access and Student Outcomes in Higher Education

Research protocol

TASO Summer Schools Evaluation

VERSION	DATE	REASON FOR REVISION/NOTES
Any changes to the design to be agreed between the implementation partner(s), evaluator and TASO. Note any agreed changes in the table below.		
1.3 [*original*]	19/03/2021	NA
Pre-registration	19/03/2021	This design has been pre-registered on the Open Science Framework (OSF) registry.
1.5	04/10/2021	Adding names to project personnel (Table 1) Adding definitions of compliance to CACE analysis Updating intervention descriptions (Section 4 and Appendix III) Adding sub-project protocol for HEPs not participating in RCT (Appendix V)

The QA rating system is based on the Evaluation Security tool presented in the TASO Monitoring and Evaluation Framework.

QA	Comments	Rating (out of 5)
Design	RCT. Whilst there are some quirks with the randomisation, they do not compromise the validity.	5
Sample size	Cohen's h is between 0.1 and 0.2.	4
Outcome measure	The outcome measure is a direct measurement of the desired outcome. It is unambiguously defined and the data is expected to be of high quality.	5
Attrition	Expected to be low for the primary and secondary outcomes. Non-compliance could be more of an issue, though there is a CACE analysis specified to measure this.	5
Validity	The only question is how representative the 2021/22 year group will be of future outcomes given the educational shocks that Covid has caused.	5
Overall		**4.8**

BIT protocol ref: 20200

TASO Transforming Access and Student Outcomes in Higher Education

Contents

1. Summary

Background

This project is a collaboration between the Centre for Transforming Access and Student Outcomes in Higher Education (TASO), eight Higher Education Providers (HEPs) and the Behavioural Insights Team (BIT). In summer 2021, a series of summer schools will be delivered with the aim of widening participation in higher education (HE) among participants. Three types of evaluation will be conducted with these summer schools: an impact evaluation, a cost evaluation, and an implementation and process evaluation (IPE). This protocol comprehensively covers the first two of these evaluations, as well as a specific contribution to the IPE. TASO is leading the IPE and is developing a separate protocol for this.

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. There is currently limited evidence on this topic.

Intervention

This study will evaluate a collection of interventions. Eight HEPs will deliver their own summer schools, either for students in pre-16 or post-16 education.

Design

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

Outcome measures

The primary outcome is whether or not the individual enters HE in the 2022/23 academic year. The secondary outcome is whether or not the individual enters HE at his/her summer school host institution.

Analyses

A combination of logistic and OLS regressions are used, as appropriate, to estimate effects on the primary, secondary and exploratory outcomes.

2. Background

This project is a collaboration between the Centre for Transforming Access and Student Outcomes in Higher Education (TASO), eight Higher Education Providers (HEPs) and the Behavioural Insights Team (BIT). In summer 2021, a series of summer schools will be delivered with the aim of widening participation in HE among participants. Three types of evaluation will be conducted with these summer schools: an impact evaluation, a cost evaluation and an implementation and process evaluation (IPE). This protocol comprehensively covers the first two of these evaluations, as well as a specific contribution to the IPE.

BIT is responsible for:

- design, analysis and reporting for the impact evaluation;
- randomly assigning participants to the treatment or control group for the impact evaluation;
- design, analysis and reporting for the cost evaluation; and
- collecting covariate data from the National Pupil Database (NPD), if this is deemed necessary and feasible.[1]

TASO is responsible for:

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

The eight HEPs are responsible for:

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

[1] Whether it is necessary to access the NPD will depend upon what data TASO is able to access from the Higher Education Access Tracker (HEAT) and the Higher Education Statistics Agency (HESA). At the time of writing the protocol, TASO is still in discussion with HEAT and HESA about this. Whether it is feasible to access the NPD will depend upon the ease of accessibility at the time. Access to the NPD is currently subject to substantial challenges and delays.

A research assistant (RA) will be placed by TASO in each HEP to support them with their evaluation responsibilities. The table below summarises the key project personnel for each organisation.

Table 1. Project personnel

Organisation	Name	Role and responsibilities
BIT	Patrick Taylor	Evaluation Manager
	Kim Bohling James Lawrence	Evaluation QA
	Dr Giulia Tagliaferri	Evaluation Supervisor
	Pujen Shrestha	Data Analyst
	Sarah Breathnach	Data Analyst
TASO	Dr Helen Lawson	Research Programme Manager. Responsible for the day-to-day management of the study.
	Sarah Chappell	Research Officer. Supporting the team on the day-to-day management of the study.
	Dr Eliza Kozman	Deputy Director (Research). Responsible for overseeing the implementation of the study.
	Jessica Hunt	Maternity cover for Deputy Director (Research).
University of Surrey	Katherine Sela	Project lead at the University of Surrey. Responsible for implementing randomisation and data collection there.
	Karla Lopez Murillo	Supporting data collection.
University College London (UCL)	Shireen Quraishi	Project lead at UCL. Responsible for implementing randomisation and data collection there.
	Emily Burchell	Supporting data collection.
University of Leeds	Liz Hurley	Project lead at the University of Leeds. Responsible for implementing randomisation and data collection there.

	Rebecca Talbot	Supporting data collection.
University of Suffolk	Marianna Stella	Project lead at the University of Suffolk. Responsible for implementing randomisation and data collection there.
	Owen Evans	Supporting data collection.
University of Gloucestershire	Fiona Curry	Project lead at the University of Gloucestershire. Responsible for implementing randomisation and data collection there.
	Liz Gray	Co-project Lead.
University of Kent	Marta Almeida	Project lead at the University of Kent. Responsible for implementing randomisation and data collection there.
	Amy Burt	Co-project Lead.
Nottingham Trent University (NTU)	Peter Cassidy	Project lead at NTU. Responsible for implementing randomisation and data collection there.
	Laura Hope	Co-project Lead.
University of East Anglia (UEA)	Rosie Hannant	Project lead at the UEA. Responsible for implementing randomisation and data collection there.
	Ed Penn	Supporting data collection.

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

3. 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. 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 focussing 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)?[2]

To answer these questions, outcomes will be compared between the participants in the trial summer schools (the treatment group), and eligible applicants who are not selected to participate (the control group). The eligibility criteria applied by HEPs will ensure that the trial sample is composed solely of disadvantaged or under-represented groups (see 'Sample selection' below for more detail on this).

4. Intervention

4.1. Introduction

This study will evaluate a collection of interventions. Participating HEPs will deliver their own summer schools, either for students in pre-16 or post-16 education. Each summer

[2] To support the IPE, effects will also be 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.

school will have its own specific characteristics, but all have the same broad aims and involve similar activities. Below, we present TASO's brief descriptions of the pre-16 and post-16 programmes. 'Appendix III: Intervention descriptions by HEP' contains a description of each summer school, broken down by provider.

4.2. Pre-16 summer schools

These summer schools are focused on Year 9 or Year 10 students from underrepresented / disadvantaged backgrounds to help them decide whether higher education is the right option for them. They also allow students to experience different university subjects to discover what subject options exist outside their current school curriculum. The experience generally lasts from 3-5 days. Students experience a range of sessions including subject tasters, student life, student finance, study skills, campus tours, and evening social activities. They also have the opportunity to work with, and ask questions of, current students at the university, either in small groups or via one-to-one mentoring.

4.3. Post-16 summer schools

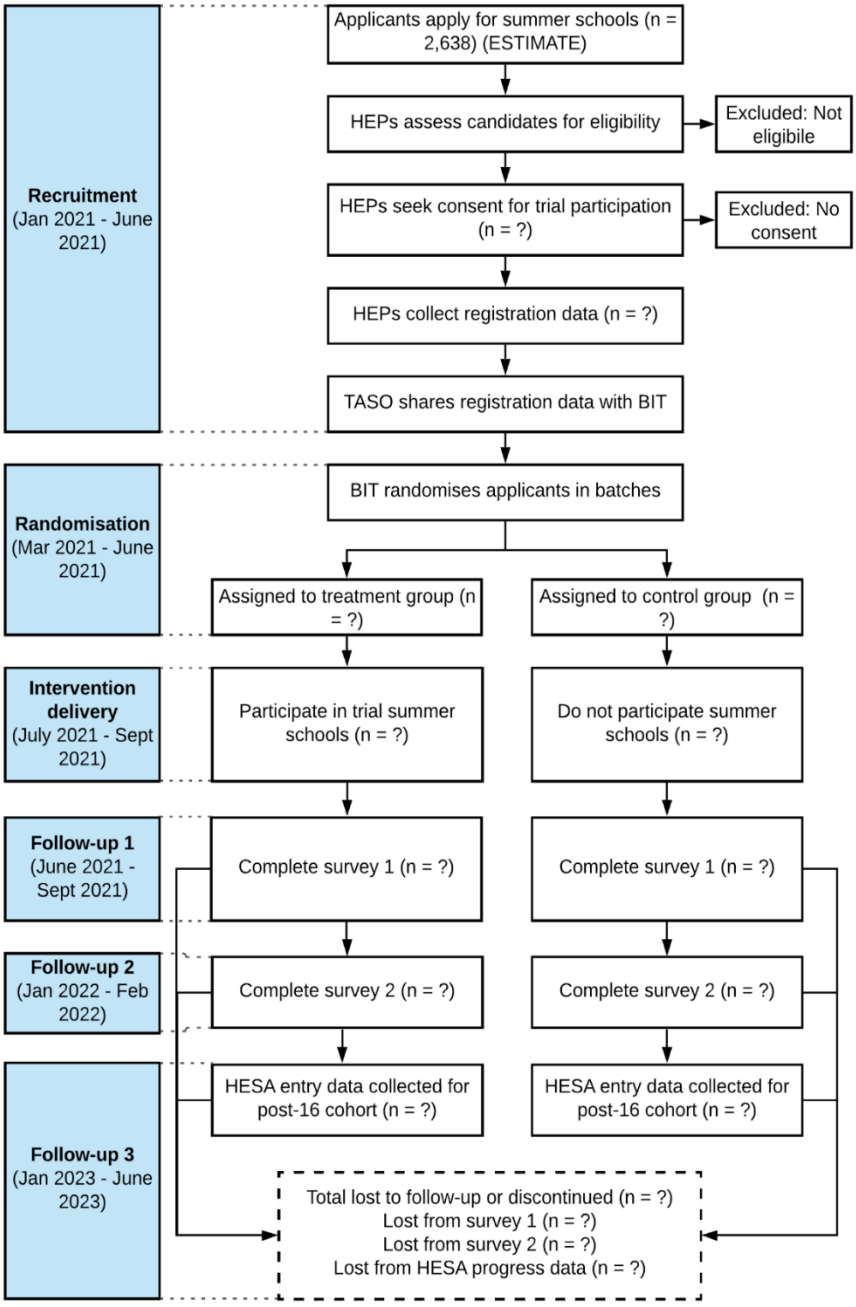
These summer schools aim to support Year 12/ First year of post-16 (and occasionally Year 13) students from underrepresented / disadvantaged backgrounds in their future decisions, including whether university is the right path for them and what subject they could study. Students will have virtual tours of accommodation and the university campus. Students will experience subject tasters and are usually required to complete a project or assignment in the subject area of their choice. Other sessions aim 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.

5. 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 will be randomly assigned to either the treatment or control group. Each summer school programme has a different number of places available, a different number of eligible applicants, and a different set of quotas that they wish to fulfil in their participant pool, so the ratio of assignment will differ by programme. See 'Randomisation' below for details of the assignment procedure.

Study activities will take place between January 2021 and November 2023[3] (including final reporting). Figure 1 gives an overview of the study flow and timeline up to the point of final data collection. A wider project timeline is given in 'Procedure', below.

Figure 1. Study flow diagram



Notes: 'Registration data' includes baseline survey data.

[3] This is an estimate based TASO providing final outcome data with BIT by the end of July 2023.

6. Outcome measures

The outcomes to be measured 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 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 is 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 will be the estimated effects on the primary and secondary outcomes. The proximal outcomes will be used for interim reporting (as early indicators), and the mediating mechanisms will be reported as part of the implementation and process evaluation.

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 will be available for both cohorts. The final column of Table 2 indicates which cohort the relevant data will be available for and, therefore, defines the sample to be used for analysing each outcome.

Table 2. Outcome measures

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 (June 2023)	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 (June 2023)	Post-16 only
EXPLORATORY 1 (PROXIMAL): Application to HE	Survey 2: Have you applied to university? Binary: yes/no	NA	After endpoint (January 2022)	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 (August 2021)	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 (August 2021)	Pre-16
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”	NA	Baseline After endpoint (August 2021)	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 (August 2021)	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 (August 2021)	Both

[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.

7. Sample selection

The study sample will be made up of all applicants to the trial summer schools who meet the HEPs' eligibility criteria. These criteria vary slightly by HEP, but the following list covers all criteria used across providers in the study. To be eligible for consideration, an applicant must have one or more of the following characteristics:

- identify as coming from a black or minority ethnic background;
- live in an area of deprivation (as defined by the most deprived quintile (Q1) of the Index of Multiple Deprivation (IMD) and/or the participation of local area in higher education (POLAR) classification);
- be in care or a care-leaver;
- be a young carer;
- have a disability;
- be the first in her/his family to attend HE;
- attend a school that partners with the HEP;
- be eligible for free school meals;
- indicate an interest in a subject offered by the HEP;
- indicate an interest in studying close to home; and/or
- be a refugee or asylum seeker.

One HEP (UCL) also requires the students to be on track to achieving the qualifications and grades required to attend the relevant degree at UCL.

The sample is divided into two age groups: a pre-16 and post-16 group. The pre-16 group will contain individuals from Years 9 and 10. The post-16 group will contain individuals from Year 12/First year post-16 education.

Recruitment of study participants will be carried out by the HEPs. The size of the sample is determined by the number of eligible applicants to the summer schools. The size of the treatment group is determined by the number of places available in each summer school. The estimated sample sizes, based on figures provided by the HEPs are given in Table 3 below.

Table 3. Estimated sample size by cohort[5]

Cohort	Estimated sample size	Size of treatment group
Pre-16	750	300
Post-16	2,640	970
Combined	3,390	1,270

8. Randomisation

8.1. Introduction

Four practical constraints are imposed by the programme that affect the randomisation:

i. Some HEPs guarantee places for applicants meeting certain criteria (e.g. care leavers)
ii. Most HEPs have quotas that they need to fill in the intervention group (for example, a 50/50 male-female split), and these quotas vary by HEP;
iii. Applicants have to be randomised in batches; and
iv. It is possible that some students will apply to more than one summer school.

These constraints add complexity to the randomisation, so a detailed step-by-step process is provided below. See Appendix IV for further information on the quotas and guaranteed places, broken down by summer school. The key thing to note from these details, for the randomisation, is that the post-16 providers wish to impose one type of quota, whereas most pre-16 providers wish to impose two types.

Randomisation will be conducted at the individual level and will be blocked, with the block influencing the probability of assignment. The characteristics of the blocks are defined by each summer school, based on the characteristics of their applicant pools, and on the quotas that they wish to meet. Individuals in the same block have the same probability of assignment. These differences in probabilities of assignment 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). As randomisation will be conducted *within* blocks (and not *across* blocks), this is a stratified randomisation, in which each block is a strata. The randomisation strategy differs from a standard stratification strategy in that we are not randomly allocating *half* candidates to the treatment and control group, but we are

[5] All figures rounded to nearest 10.

allocating the required number of candidates to the treatment group (corresponding to the available summer school places) and the remainder to the control group.

Stratified randomisation is advisable only when the average size of blocks is not too small. The next section explains what 'too small' means in this context, and applies these conditions to the expected characteristics of the sample.

8.2. Decision rule for stratified randomisation

In order to meet the quotas specified by providers, we are using stratified randomisation. However, stratified randomisation should not be performed if: (i) the average block/stratum size < (the number of arms * 10); and (ii) there are 10+ blocks/strata containing ≤ (the number of arms * 2). So, for this randomisation strategy to be valid:

i. the average size of the strata/blocks we create should not contain fewer than 20 participants; and
ii. we must not have 10 or more blocks containing 4 or fewer participants.

In this case, a strata/block is the combination of summer school x quota (for example, UCL English x female).

Table 4 below shows that even in extreme situations, such as 10% fewer candidates than expected applying to summer schools, these conditions are likely to be met for the sample of post-16 providers.

Table 5 summarises whether the conditions for stratified randomisation would be met based on the information provided by pre-16 providers. It shows that, if these estimates are accurate, we will not meet both stratification conditions for the analysis of the outcomes that include the pre-16 cohort in the sample. The figures indicate, however, that if we ignore the feeder school quotas and allocate the places using the same approach employed for the post-16 HEPs (i.e. keeping the sex and deprivation quotas only), we will likely meet the conditions for stratified randomisation. As the randomisation will take place in batches, and the schedule for this is not yet confirmed, we have to make the decision now as to which quotas we fulfil. This means that we will not be able to engineer the randomisation to meet the feeder school quotas requested by HEPs. We will only randomise to meet the sex and deprivation quotas.

Table 4. Checking block sizes for the post-16 sample[6]

[6] All figures rounded to the nearest 10.

Scenario	Condition (i) What is the average block size?	Is condition (i) met? (>20)	Condition (ii) How many blocks contain n ≤ 4?	Is condition (ii) met?
Expected n =2,600[7]	50	✔	0	✔
Expected n =2,470 (accounting for 5% attrition)	50	✔	0	✔
Expected n =2,340 (accounting for 10% attrition)	40	✔	0	✔

Table 5. Checking block sizes for the pre-16 sample[8]

Scenario	Condition (i) What is the average block size?[9]	Is condition (i) met? (>20)	Condition (ii) How many blocks contain n ≤ 4?	Is condition (ii) met?
Expected n = 740[10], gender AND feeder quota apply	40	✔	Both male and female blocks in Kent (80 blocks)	X
Expected n = 740[11], only gender quota applies	110	✔	None	✔

[7] This expected n excludes the applicants who will receive a guaranteed place at the summer school.
[8] All figures rounded to the nearest 10.
[9] These estimates assume that the HEPs will receive the same number of applicants from each feeder school, and that the applicant pool in each school has the expected proportion of female applicants.
[10] This expected n excludes the applicants who will receive a guaranteed place at the summer school.
[11] This expected n excludes the applicants who will receive a guaranteed place at the summer school.

BIT will communicate the results of the randomisation to TASO who will enrol participants in the trial. Trial participants will not be blind to the study.[12] Balance checks will be conducted on all of the control variables used in the primary analysis.

8.3. Randomisation procedure

TASO will provide BIT with a series of Excel spreadsheets containing a list of all eligible applicants for each individual summer school. BIT will provide a template for data submission. The variables used for randomisation will be as follows.

- Name of summer school
- TASO unique ID. The same applicant must be identifiable with the same ID across different summer schools / spreadsheets.
- Sex (M/F)
- Guaranteed place (Y/N)
- Low-SES (Y/N) (Surrey only)
- School provider (Pre-16 providers only, if stratification conditions are met)

These spreadsheets will be sent to BIT in batches via a pre-agreed secure method. BIT will allocate applicants to treatment/control conditions on a rolling basis in these batches, as follows.

First batch

If this batch includes more than one summer school:

1. Append applicant lists from different summer schools.
2. Assign guaranteed places. All applicants with a characteristic that guarantees them a place will be assigned to participate in the summer school, *but not included in the trial analysis*.
3. For each applicant applying to more than one summer school in the batch, randomly select for which summer school they are to be considered using a random number generator. We will create a variable (ENTERRAND) taking value 1 if the applicant enters randomisation for that summer school, 0 otherwise. This strategy implies that if two applicants in the same batch apply to the same set of summer schools, they might not be selected to participate in the randomisation for the same summer school.
4. For each summer school in the batch, assign applicants with ENTERRAND = 1 to treatment/control. This will be done as follows.

[12] Both post-16 and pre-16 participants will have to read and sign a consent form.

a. Split the applicant list according to the quota variable (e.g. sex). Using the 50/50 sex quota as an example, assign females a computer-generated random number.
b. Sort the random numbers in ascending order.
c. Allocate 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 will receive a place at the summer school.
d. Allocate all remaining female applicants to the control group.
e. Repeat steps (a) to (d) to allocate the remaining 50% of places available to males on the list.

If the batch covers 1 summer school only:

1. Assign guaranteed places. All applicants with a characteristic that guarantees them a place will be assigned to participate in the summer school, *but not included in the trial analysis*.
2. For each summer school in the batch, assign applicants with ENTERRAND = 1 to treatment/control using steps 4a to 4e above.

Second/third/n-th batch

1. Check if any applicants appear in a previous batch using TASO's unique ID.[13] If so, assign ENTERRAND=0 to the applicant for the summer schools in the current batch (so that they cannot be assigned to either the treatment OR control group in this batch). This does not apply to participants with guaranteed places, who will be given places on all summer schools to which they apply.
2. Repeat steps 3 and 4 outlined above for batch 1.

This strategy means that the order in which a batch comes in may affect the number of students who can enter the randomisation for those summer schools (in the case where some students do apply for more than one summer school). In later batches, every applicant who applied to a summer school in a previous batch is automatically excluded from entering randomisation. It is unlikely that a sex or SES quota will not be able to be fulfilled. However, in the event that this is the case, the quota will be dropped and

[13] NB: Prior to sending applicant data to BIT, TASO will identify duplicate applicants by checking unique HEAT IDs, and flagging these duplicates in the dataset (with a new variable) before sharing with BIT.

randomisation will be conducted within the provider to fill the number of places available on the summer school, with the remainder allocated to the control group.

9. Data collection

Data will be collected for the following five purposes.

1. For project management.
2. For randomisation (including ensuring quotas are met and checking covariate balance).
3. For estimation of treatment effects.
4. For assessment of the external validity of estimated treatment effects.
5. For estimation of costs.

Data will be collected from the following eight sources.

1. TASO's HEP staff contact list
2. HEP participant registration forms
3. Outcome survey 1, administered by TASO
4. Outcome survey 2, administered by TASO
5. HEAT
6. HESA
7. The NPD
8. Cost evaluation survey, administered by TASO

All individual items of data to be collected are listed in Table 6 below, with more detailed descriptions of the purpose of each item. The table also indicates who collects each data item. For all data except that accessed from the NPD, TASO will be responsible for sharing the data with BIT. Some variables are collected twice from different sources to support interim report writing deadlines, to ensure that we collect the variable, and to improve data quality.

Table 6. Trial data

Data item	Purpose	Collection point	Collector	Sample
Student data				
TASO unique ID	Matching datasets shared by TASO	Baseline	TASO	Both

Sex	Meeting treatment group quota (stratification) Balance checks Control variable Assessing external validity	Baseline After endpoint	HEPs HESA	Both
Ethnicity	Balance checks Control variable Assessing external validity	Baseline After endpoint	HEPs HESA	Both
Postcode-level marker of disadvantage	Meeting treatment group quota (stratification) Control variable Assessing external validity	Baseline	TASO	Both
Free School Meal (FSM) status	Meeting treatment group quota (stratification) Balance checks Control variable Assessing external validity	Baseline After endpoint	HEPs NPD (TBC)	Both
Whether anyone in the family has been to university	Balance checks Control variable Assessing external validity	Baseline	HEPs	Both
Disability status	Assessing external validity	Baseline	HEPs	Both
Experience of children's social care	Meeting treatment group guaranteed places Assessing external validity	Baseline	HEPs	Both
Whether from an underrepresented group (Young carer, estranged, Gypsy, Roma, Traversal communities, refugees, children of military families)	Meeting treatment group guaranteed places Assessing external validity	Baseline	HEPs	Both
First name	Uniqueness check Accessing HESA and NPD data	Baseline	HEPs	Both

Last name	Uniqueness check Accessing HESA and NPD data	Baseline	HEPs	Both
Date of birth	Uniqueness check Accessing NPD data	Baseline	HEPs	Both
Postcode	Uniqueness check Accessing HESA and NPD data	Baseline	HEPs	Both
Academic year group	Balance checks Control variable Accessing HESA and NPD data Assessing external validity	Baseline	HEPs	Both
School name	To identify School ID	Baseline	HEPs	Both
School location	To identify School ID	Baseline	HEPs	
School ID (URN)	Uniqueness check Control variable Accessing HESA and NPD data	Endpoint	TASO	Both
Summer school applied to	Subgroup analysis Control variable	Baseline	TASO	Both
Pre or post-16 programme	Subgroup analysis	Baseline	TASO	Both
Summer school attended	Estimating effects of intervention Compliance check	Endpoint	TASO	Both (treatment group only)
Summer school attendance	CACE analysis	Endpoint	TASO	Both (treatment group only)
Attainment at Key Stage 2 Maths and English	Control variable Assessing external validity	Baseline After	HEPs NPD	Both

		endpoint	(TBC)	
Attainment at Key Stage 4 (Attainment 8 score)	Control variable Assessing external validity	Baseline After endpoint	HEPs NPD (TBC)	Post-16 only
Progression to HE	Primary outcome	After endpoint	HESA	Post-16 only
Progression to host university	Secondary outcome	After endpoint	HESA	Post-16 only
Application to HE	Exploratory outcome	After endpoint	TASO	Post-16 only
Likelihood of going to HE	Exploratory outcome	Baseline After endpoint	TASO	Both
Likelihood of progressing to academic study post-16	Exploratory outcome	Baseline After endpoint	TASO	Pre-16 only
Desirability of HE	Exploratory outcome	Baseline After endpoint	TASO	Both
Self-efficacy relating to HE	Exploratory outcome	Baseline After endpoint	TASO	Both
Compatibility of HE with social identity	Exploratory outcome	Baseline After endpoint	TASO	Both
Perception of practical barriers to HE	Exploratory outcome	Baseline After endpoint	TASO	Both
Cost data				
Intervention cost estimates	Estimating cost per participant	Endpoint	TASO	NA

HEP staff data				
First name	Project management	Baseline	TASO	NA
Last name	Project management	Baseline	TASO	NA
Work email address	Project management	Baseline	TASO	NA
Work telephone number	Project management	Baseline	TASO	NA

The two outcome surveys have been developed by TASO, taking items from a range of sources and creating some items from scratch. A summary of the constructs measured in these surveys, along with their source and notes on validity and reliability is provided in Table 7. (See the outcomes section above for the full questions and methods of aggregation).

Table 7. Validity and reliability of survey items

Construct	Source	Notes on validity and reliability
Applied to HE	NA	No testing performed. Created by TASO for this evaluation. A direct question about past behaviour.
Likelihood of going to HE	Next Steps	Item adapted from Next Steps; therefore cognitively tested. Aspirations found to be highly correlated with actual HE progression (Anders & Micklewright 2015).
Likelihood of progressing to academic post-16 study	NA	No testing performed. Created by TASO for this evaluation.
Self-efficacy relating to HE	Next Steps	Scale adapted by TASO from Next Steps, which was cognitively tested. However, TASO's version reduces a 4-item scale to 2-items and alters the wording of the items that are kept for this evaluation. Aspirations found to be highly correlated with actual HE progression (Anders & Micklewright 2015).
Compatibility of HE with social identity	Adapted from Uni Connect and University of Gloucestershire in-house survey	No validation evidence found, but full scale developed by sector (so some face validity). TASO's version reduces a 5-item scale to single item for this evaluation.

Perception of practical barriers to HE	Adapted from Uni Connect and University of Gloucestershire in-house survey	No validation evidence found, but full scale developed by sector (so some face validity). TASO's version reduces a 4-item scale to 2-items and alters the wording of the items that are kept for this evaluation.
COVID impact	Pearson global learner survey	No validation evidence found. Items have been adapted. Previous items were "The COVID-19 pandemic has made me rethink my career path" and "I'm worried that I may have to change industries or career fields because of the COVID-19 pandemic."

TASO will promote data quality and security through the following measures.

- A data sharing template, including details on the variables and their required coding, will be given to all providers to ensure consistent and reliable data collection across all universities.
- All data shared with TASO will be processed in line with its data protection policy.
- Before sharing with BIT, all data received by TASO will be checked and cleaned by the Research Programmes Manager and the Research Officer.

All data shared with BIT will be processed in line with its data protection policy. A summary of this policy can be found in Appendix II. In the analysis, BIT will promote data quality and security through the following measures.

- All variables will be clearly named, coded and labelled before analysis.
- Checks on the data received will be carried out for valid values, range, and consistency against already held data.
- Any modifications to datasets will be recorded in the analysis code, which will be well-annotated.
- Original raw datasets will never be amended.
- Access to the project data will be restricted to project personnel.
- All data stored by BIT will be backed-up.

10. Procedure

A high-level project timeline is given below.

Table 8. Trial timeline

Timeframe	Action
October 2020 - June 2021	• Complete trial protocol • Set up data sharing processes and agreements • Recruit participants and assign to treatment or control group • Collect baseline registration data
July 2021 - January 2022	• Deliver summer schools • Collect outcome data through survey 1 (knowledge and attitudes) • Analyse data and complete interim report 1
January - April 2022	• Collect outcome data through survey 2 (HE applications) • Analyse data and complete interim report 2
June - December 2022	• Collect NPD data (TBC)
January 2023 - November 2023	• Collect HESA outcome data • Analyse data and complete final report

11. Power calculations

11.1. Introduction

Power calculations have been conducted for the primary outcome only (i.e. progression to HE). This means that only the 27 post-16 summer schools are included in the calculations. We do not have control over the size of the sample, so these calculations estimate the minimum detectable effect size (MDES), given the estimated sample. Each provider has supplied us with the following estimates for each summer school individually.

- Number of expected eligible applicants
- Number of places available
- Number of places guaranteed for applicants meeting certain criteria (e.g. care leavers)
- Desired proportion of places granted to those with certain characteristics (the 'quota', e.g. a 50/50 sex split)
- Expected proportions of applicants meeting the quota criteria (e.g. 65% of applicants will be female)

Power calculations are based on this information. A table containing this information can be found in Appendix IV. There is a different number of expected eligible applicants and places available for each summer school. Each summer school has also specified a quota to be met. For all summer schools except for Surrey, the quota is a 50/50 split by participant sex. For Surrey, the quota is 65% low-socioeconomic-status (SES) / 35% non-low SES.[14]

The total applicant pool is estimated to be 2,638. There are approximately 929 places[15] available in the treatment group, excluding the 37 places that are guaranteed to applicants meeting the relevant criteria.

11.2. Baseline progression to HE

To estimate the minimum effect size that is detectable with a sample of 2,638, we need to know what proportion of the control group (i.e. those who apply but are not invited to attend a summer school) will progress to HE. We will refer to this as 'baseline progression'. Our baseline estimates for HE progression are based on figures reported in two quasi-experimental studies. These studies report data on the proportion of widening participation (WP) students that progress to HE.

Study 1 examined the effect of the Sutton Trust's Summer Schools on subsequent higher education participation (Hoare & Mann 2011). To do this, those applying to and attending summer schools were matched with and compared against a comparison group made up of 'inner controls' and 'outer controls'. Inner controls were students who applied for a summer school place unsuccessfully and 'outer controls' were students with similar characteristics to the Trust's WP eligibility criteria, but who did not apply for a summer school. For the outer control group, applicants were included if they met all of the following criteria: they attended a school with low HE progression, they attended a school with low-attainment, and neither of their parents experienced higher education. In terms of personal characteristics, the study matched on WP indicators such as residence in a low participation neighbourhood, as measured by The Higher Education Funding Council for England's (HEFCE) participation of local area in higher education (POLAR) classification, and ethnicity (white/non-white). This study reported that 76.3% of the applicant pool that did not attend the summer school registered for HE.

[14] SES will be judged by the summer school provider based on indicators such as participation of local area in higher education (POLAR) classification and free school meals (FSM) eligibility. The provider was unable to say which exact indicator they wished to use at the time of writing.

[15] This is an approximation because UCL specifies a range of places available for each of their summer schools. We have taken the midpoint of the range in these cases.

Study 2 investigated whether engagement in Aimhigher interventions (a range of interventions such as mentoring, campus visits, subject masterclasses and attendance at summer schools) increases the likelihood that disadvantaged learners progress to HE (Horton & Hilton 2020). Disadvantaged learners were defined as learners that live in wards funded by the Office for Students' National Collaborative Outreach Programme (NCOP). These wards are characterised by lower HE participation rates. Of those that did not engage with the programme (i.e. didn't take part in any of the activities), 38.7% progressed to HE.

These studies indicate that baseline progression to HE could fall between 38.7% and 76.3%. We have used 76% as the default baseline proportion for progression to HE in the power calculations. This is because the intervention and the characteristics of the sample in Study 1 better match the characteristics of the expected applicant pool in this trial.

11.3. Procedure power calculations

The following approach was taken to estimate the MDES for the primary outcome.

1. We simulated a dataset that contained all trial participants and reflected the characteristics of the expected applicant pool using the estimates supplied by providers. This dataset also included the primary outcome variable "progressed" in which 76% of applicants were randomly allocated to receive a "yes" because this is the expected baseline proportion progressing to HE. Each of these datasets contained the following variables:
    - Name of summer school (e.g. UCL English)
    - Application ID
    - Female (yes/no)
    - FSM (yes/no)
    - Guaranteed place (yes/no)
    - Treated (yes/no)
    - Progressed to HE (yes/no)
    - Block (to indicate the quota, e.g. female x UCL_English).
2. We fitted the logistic regression that will be used to analyse the data post-trial (see section 12.1), excluding all control variables except the randomisation block fixed effect. Control variables were excluded because we do not have estimates for the distribution of these characteristics and how they relate to HE progression. This means that the estimated MDESs are conservative.
3. We estimated the MDES using the following formula (Cohen 2013):

$$MDES = [\phi(1 - \frac{\alpha}{2}) + \phi(power)] \times SE$$

where,

- ϕ is the cumulative density function of the normal distribution;
- α is the probability of rejecting the null hypothesis when it is true (set at 0.05);
- $power$ is the probability of rejecting the null hypothesis when it is false (set at 0.8); and
- SE is the standard errors estimated from the simulated dataset.

The information supplied by providers for these calculations are estimates, so we have repeated the process outlined above a number of times, each time varying one of the following, to estimate a range for the MDES.

- **Baseline progression to HE:** Given that the actual baseline proportion of progression to HE for trial participants might rest somewhere between the figures reported in studies 1 (76%) and 2 (39%), we have performed an additional calculation in which baseline progression is set to 57% (the midpoint between 39% and 76%).
- **Applicant pool numbers (attrition)**: We have conducted two calculations in which we reduced the numbers in the applicant pool by 5% and 10% to account for a potential reduction in sample size due to some of those in the expected applicant pool not meeting the eligibility criteria.
- **Quotas:** We have conducted two calculations in which we changed the expected proportions for quota characteristics to 10 percentage points (pp) less than expected and 10pp more than expected. For example, if the provider estimated that 60% of applicants would be female, we ran power calculations when the pool was also 70% female and 50% female.

Based on these assumptions, our estimates indicate that the MDES may range from 4.7pp to 5.7pp. This is equivalent to saying that, in the worst case scenario estimated, we think that the trial would be powered to detect an increase in progression to HE from 76% to ~82%.

Table 9. MDES estimates

Alpha	**0.05%**
Power	**80%**

Baseline		Attrition			Quota characteristic						
76%	57%	0%	5%	10%	As expected	10% more than expected	10% less than expected	Sample size	n of treated group	n of control group	MDES pp
✔		✔			✔			2,600	925	1,675	4.8
✔			✔		✔			2,465	925	1,540	5.0
✔				✔	✔			2,348	925	1,423	5.0
✔		✔				✔		2,600	925	1,675	5.2
✔		✔					✔	2,600	925	1,675	4.7
	✔	✔			✔			2,600	925	1,675	5.7

pp = percentage points; MDES = minimum detectable effect size.

12. Analytical strategy

12.1. Primary outcome

The following model will be used to estimate the effects of the intervention on the primary outcome. Analysis will be 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 the function $logit$ is defined as the log-odds ratio

$$logit(p) = log(\frac{p}{1-p})$$

and,

- Y_i is a binary indicator of whether the individual enters HE in the academic year 2022/23 (1 if they enter, 0 if not);
- p_i is the probability that the individual enters HE in the academic year 2022/23;
- 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).[16]

12.2. Secondary outcome

The following model will be used to estimate the effects of the intervention on the secondary outcome. Analysis will be 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 the function $logit$ is defined as the log-odds ratio

$$logit(p) = log(\frac{p}{1-p})$$

and,

- Y_i is a binary indicator of whether the individual goes on to study at the HEP that delivers the summer school applied to[17] (1 if they do, 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

[16] Note that the attainment control variable varies depending upon the sample. KS4 scores will not be available for the pre-16 cohort so, when this cohort is included in the analysis of other outcomes, KS2 scores are used instead.

[17] In the event that an individual applies to more than one summer school, the summer school with which they are randomised will be considered the 'summer school applied to'.

- 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).

12.3. Exploratory outcome 1

The following model will be used to estimate the effects of the intervention on exploratory outcome 1. Analysis will be 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 has applied to university by January 2022 (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).

12.4. Exploratory outcome 2

The following model will be used to estimate the effects of the intervention on exploratory outcome 2, using ordinary least squares (OLS) regression. Analysis will be 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 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); and
- ϵ_i is the heteroskedasticity robust residual error term.

12.5. Exploratory outcome 3

The following model will be used to estimate the effects of the intervention on exploratory outcome 3, using ordinary least squares (OLS) regression. Analysis will be 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 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);
- 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.

12.6. Exploratory outcome 4

The following model will be used to estimate the effects of the intervention on exploratory outcome 4, using ordinary least squares (OLS) regression. Analysis will be 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.

12.7. Exploratory outcome 5

The following model will be used to estimate the effects of the intervention on exploratory outcome 6, using ordinary least squares (OLS) regression. Analysis will be 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 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.

12.8. Exploratory outcome 6

The following model will be used to estimate the effects of the intervention on exploratory outcome 7, using ordinary least squares (OLS) regression. Analysis will be 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.

12.9. Exploratory subgroup analysis

For all binary outcomes (primary and secondary), heterogeneous effects by summer school will be estimated by testing for interactions using the following model.

$$Y_i \sim bernoulli(p_i)\ ;\ logit(p_i) = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 Z_i + \beta_4 T_i \cdot Z_i$$

where,

- Y_i is the outcome of interest;
- p_i is the probability of the outcome for the individual;
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control);
- X_i is the vector of pre-treatment covariates used for the analysis of the whole group of HEPs (excluding summer school applied to); and
- Z_i is a categorical variable indicating which summer school the individual applied to (where they applied to more than one, we select at random following the procedure in the "randomisation" section).

For all continuous outcomes (primary and secondary), heterogeneous effects by summer school will be estimated by testing for interactions using the following model.

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

where,

- Y_i is the outcome for the individual;
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control); and
- X_i is the vector of pre-treatment covariates used for the analysis of the whole group of HEPs (excluding summer school applied to);
- Z_i is a categorical variable indicating which summer school the individual applied to; and
- ϵ_i is the heteroskedasticity robust residual error term.

For the outcomes that are measured for both cohorts, heterogeneous effects by cohort (pre-16 vs. post-16) will be estimated by testing for interactions using the same procedures as above.

12.10. Descriptive statistics on the impact of COVID-19

Outcome survey 1 includes two questions that asks 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 will be reported by treatment condition to aid interpretation of the results in the IPE. A formal test for heterogeneous effects will not be carried out in this case because it is not possible to recover an unbiased estimate when the moderating factor is realised post-intervention (as in this case).

12.11. Multiple comparisons

This study includes a large number of statistical tests. This increases the chance that a finding will appear to be statistically significant when there is no real effect. If all of these tests were given the same status in the analysis, then it would be necessary to adjust the p-values of some estimates to ensure that they reflect the true probability under the null hypothesis. Exactly how many p-values need to be adjusted, and in what way, is disputed in the literature. BIT's standard operating procedures, to guard against this problem of false discoveries, work on the following three principles.

1. Have as few outcomes as possible.

2. Have as few treatment arms as possible.
3. Make as few comparisons as possible.

In situations where a large number of comparisons are made, BIT uses the Benjamini-Hochberg step-up procedure to correct for this (Benjamini & Hochberg 1995). The below table shows when we use this procedure. The procedure is applied separately for primary and secondary outcomes, but does not apply to exploratory outcomes.

Table 10. When to correct for multiple comparisons

Should I use multiple comparisons? Orange = yes					
	Number of outcomes				
Number of treatment arms (i.e. trial arms excluding control)		1	2	3	4+
	1				
	2				
	3				
	4				
	5+				

This study has one primary outcome and one secondary outcome, so no adjustments will be made for multiple comparisons in this case. The categorisation of primary, secondary and exploratory analysis made here has important implications for the interpretation and reporting of the results. The exploratory analysis will be reported as such, and these findings will be described as less secure as a result. The exploratory analysis will be used as follows.

- **Effects on proximal outcomes** will be used for interim reporting. This will give an early indication of the effects, before the primary and secondary outcome data has been collected. The results from the primary and secondary analyses, when available will supersede these interim results.
- **Effects on potential mechanisms** will be used in the IPE to help us to understand how the observed effects are created (or why they are not).
- **Heterogeneous effects** will be used in the IPE to help us to understand the factors that moderate the effects of the intervention on the primary and secondary outcomes.

The headline findings from this study will be in relation to the primary and secondary outcomes only.

12.12. Compliance

In the case of one-sided non-compliance (where some individuals who are assigned to treatment do not participate), we will use an instrumental variables approach to estimate the Complier Average Causal Effect (CACE) for the primary outcome. In the context of the trial, to be considered as minimally compliant with the treatment, a participant must have attended a certain number and type of sessions. We do not know the true minimal dosage needed to generate a treatment effect, so the cut-off chosen for compliance is based on the providers' best estimates. There is variation in these estimates because there is variation in what constitutes a summer school in each case. The definition of compliance with treatment for each summer school is given in the table below.[18]

Summer school	Definition of compliance with treatment
Kent	≥ 50% attendance overall and completing 'time capsule' assignment
NTU Ellis	100% attendance
NTU Nottingham	100% attendance
UCL Architecture	≥ 30% attendance overall
UCL Health and Wellbeing Data Science	≥ 30% attendance overall
UCL Biosciences	≥ 30% attendance overall
UCL Chemical Engineering	≥ 30% attendance overall
UCL Astrophysics	≥ 30% attendance overall
UCL History	≥ 30% attendance overall
UCL Natural Sciences	≥ 30% attendance overall
UCL Economics	≥ 30% attendance overall

[18] The list of summer schools in this table is different to those in Appendix IV because these definitions were collected after the trial had started. At this point, the list of participating summer schools had changed.

Leeds German	≥ 60% attendance overall
Leeds French	≥ 60% attendance overall
Leeds Spanish	≥ 60% attendance overall
Leeds Psychology	≥ 60% attendance overall
Leeds Maths	≥ 70% attendance overall
Leeds Biological Science	≥ 65% attendance overall, including 2 academic project sessions
Leeds Social Sciences	≥ 60% attendance overall

The instrumental variable that we will use is treatment assignment, which is assumed to influence participation in the programme but not the outcome variable in its own right.

Two key assumptions need to hold for this approach:

1. Being assigned to the treatment increases participation in the treatment. In this instance, individuals may only participate in the programme if they are assigned to treatment. This is a safe assumption as BIT will define assignment and HEPs will have control over participation.
2. Assignment does not, in itself, have an effect on the outcome of interest. We have no reason to believe that the offer of the programme would influence entry to HE on its own, but instead believe that any effect will be achieved through participation in the programme.

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

$$T_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \eta_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 \hat{T}_i + \beta_2 X_i + \epsilon_i \quad (2)$$

where:

- Z_i is a binary indicator for treatment assignment (1 if the individual is assigned to treatment and 0 if they are assigned to control);
- T_i is whether a student meets the minimal compliance threshold;

- 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);
- η_i is the error term in the first stage;
- ϵ_i is the error term in the second stage;
- $\hat{T}_i$ are the predicted levels of compliance with the programme from (1); and
- Y_i is a binary indicator of whether the individual enters HE in the academic year 2022/23 (1 if they enter, 0 if not).

12.13. Missing data

All analysis described above will be conducted on complete cases only. Missing data analysis will then be conducted on the primary outcome only as follows. First, the number of complete observations (those without any data missing) will be reported. If fewer than 5% of observations contain missing values, then little bias is likely to be introduced by listwise deletion (Shulz & Grimes 2002, p.784), so no further analysis will be conducted. If more than 5% of observations have missing values, then we will aim to establish whether the data is missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). If we think data is MAR or MCAR, we will test this by running a logistic regression; creating a binary indicator for whether values of a variable are missing, then examine whether any of the covariates are significant predictors of this missingness. If the data appears to be MCAR or MAR, the following procedure will be followed.

1. Multiple imputation will be carried out.
2. The relevant analysis to re-estimate effects will then be performed separately on each imputed dataset.
3. The results from these estimates will be pooled into a single set of parameter estimates and confidence intervals using ‘Rubin’s rules’.

If the data appears to be MNAR sensitivity analysis will be conducted. This will investigate the sensitivity of the point estimate of the treatment effect to changes in model specification (and hence sample definition), through the inclusion and exclusion of variables for which observations are missing.

12.14. Robustness checks

The quota imposed by HEPs means that a large number of blocks will be used in the randomisation for this study, and that different individuals will have different probabilities of assignment. To account for the differential probability of assignment, an indicator of randomisation block is included as a covariate (a block fixed effect) in the models used to estimate treatment effects. Under these conditions, for binary outcome variables, it is possible that some blocks may contain all zeros or all ones. In this case, these blocks will not contribute to the effect estimate when using logistic regression, thus affecting its accuracy and precision. To account for this, all effects based on binary outcomes (which are analysed using a logit in the main analysis) will be re-estimated using OLS regression as a robustness check. If there are blocks that contain all zeros or all ones, and the results differ between logit and OLS, then the OLS results will be preferred.

13. Cost evaluation

The cost evaluation will provide an estimate of the cost of the intervention per participant. This estimate will focus on cost from the perspective of an HEP and will be based on the direct, marginal financial costs of implementing the intervention. This includes anything which the HEP needs to pay for beyond business as usual costs. Time spent by HEP staff in preparing and delivering the summer schools will be reported separately from the financial costs.

A cost questionnaire will be conducted with the member of staff in each HEP who is responsible for managing the summer school. This questionnaire will be created by BIT, in consultation with the project team at TASO, and the data collection will be coordinated by TASO. The questionnaire will be conducted through structured interviews with a sample of six HEPs (three pre-16 and three post-16 providers). These interviews will be conducted separately to any carried out for the IPE. Taking an interview-based approach with a small sample (rather than using an online questionnaire with a larger group) will allow us to probe the level of detail required for an accurate estimate.

14. Ethical considerations

TASO has carried out an ethical review of the study that has been approved by the King's College London (KCL) ethics committee. See separate document for details.

15. Risks

Table 11. Risk analysis

Part of evaluation	Risk	Mitigation strategy	Risk owner
Participant recruitment	Data processing agreements (DPAs) are further delayed, so notification of applicants is delayed, so applicants and/or HEPs drop out.	TASO to prioritise setting up and signing DPAs. TASO to maintain regular communication with HEPs to address any issues caused by the delays.	TASO
Randomisation	BIT will not be able to randomise applicants in time to meet HEPs' applicant notification deadlines.	TASO to prioritise setting up and signing DPAs. TASO to maintain regular communication with HEPs to address any issues caused by the delays. BIT to continue to replan project resourcing to try to be as flexible as possible.	TASO and BIT
Data collection	Survey-based outcome measures may yield small samples and be subject to differential attrition.	TASO has funded RAs in every HEP to facilitate data collection. HEPs are funded to take part in the project – so there is buy-in.	TASO, HEPs

16. Bibliography

Anders, J., & Micklewright, J. (2015). *Teenagers' expectations of applying to university: how do they change?*. Education sciences, 5(4), 281-305.

Barnett, E. A., Hare Bork, R., Mayer, A. K., Pretlow, J., Wathington, H., & Weiss, M. J. (2012). *Bridging the Gap: An Impact Study of Eight Developmental Summer Bridge Programs in Texas*. National Centre for Postsecondary Research. https://bit.ly/3jx9saK

Benjamini, Y. and Hochberg, Y. (1995). *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), pp.289-300.

Bowes, L., Tazzyman, S., Sandhu, J., Moreton, R., Birkin, G., McCaig, C., Madriaga, M., Kozman, E. and Wright, H. (2019). *The National Collaborative Outreach Programme. End of phase 1 report for the national formative and impact evaluations*. Office for Students. https://bit.ly/3lYCG4D

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

Hoare, T., & Mann, R. (2011). *The impact of the Sutton Trust's Summer Schools on subsequent higher education participation: A report to the Sutton Trust*. University of Bristol, Widening Participation Research Cluster. https://bit.ly/34rU3o6

Horton, M., & Hilton, G., (2020). *Does Engagement in Aimhigher Interventions Increase the Likelihood of Disadvantaged Learners Progressing to HE?* UniConnect evaluation. https://bit.ly/3hbMCpM

Montacute, R., & Cullinane, C. (2018). *Access to advantage: The influence of schools and place on admissions to top universities*. The Sutton Trust. https://bit.ly/34baZ27

Robinson, D., & Salvestrini, V. (2020). *The impact of interventions for widening access to higher education*. Education Policy Institute. https://bit.ly/34uBXSv

Williams, D. (2006). *On and off the 'Net: Scales for social capital in an online era.* Journal of computer-mediated communication, 11(2), 593-628.

17. Appendix I: Outcome surveys

TASO pre-16 summer school survey items

1. How likely is it that you will study at school or a sixth form after you've finished Year 11? [5-point Likert scale from Extremely likely to Extremely unlikely]
2. How likely are you to apply to university? [7-point Likert scale from Extremely likely to Extremely unlikely]
3. How confident are you that you could make a successful application to university? [5-point Likert scale from Extremely confident to Not confident at all]
4. How confident are you that you could succeed at university? [5-point Likert scale from Extremely confident to Not confident at all]
5. How much do you agree with the following: "University is for people like me"? [5-point Likert scale from Strongly agree to Strongly disagree]
6. How confident are you that you could afford to go to university? [5-point Likert scale from Extremely confident to Not confident at all]
7. How confident are you that you know how to apply to university? [5-point Likert scale from Extremely confident to Not confident at all]
8. How much do you agree with the following: " The COVID-19 pandemic has made me rethink my future plans"? [5-point Likert scale from Strongly agree to Strongly disagree]
9. How much do you agree with the following: "I'm worried that I may have to change my study or career plans because of the COVID-19 pandemic"? [5-point Likert scale from Strongly agree to Strongly disagree]

TASO post-16 summer school survey items

1. How likely are you to apply to university? [7-point Likert scale from Extremely likely to Extremely unlikely]
2. How confident are you that you could make a successful application to university? [5-point Likert scale from Extremely confident to Not confident at all]
3. How confident are you that you could succeed at university? [5-point Likert scale from Extremely confident to Not confident at all]
4. How much do you agree with the following: "University is for people like me"? [5-point Likert scale from Strongly agree to Strongly disagree]
5. How confident are you that you could afford to go to university? [5-point Likert scale from Extremely confident to Not confident at all]
6. How confident are you that you know how to apply to university? [5-point Likert scale from Extremely confident to Not confident at all]
7. How much do you agree with the following: " The COVID-19 pandemic has made me rethink my future plans"? [5-point Likert scale from Strongly agree to Strongly disagree]
8. How much do you agree with the following: "I'm worried that I may have to change my study or career plans because of the COVID-19 pandemic"? [5-point Likert scale from Strongly agree to Strongly disagree]
9. Have you applied to university? (yes/no) (**asked January 2022**)

18. Appendix II: BIT data protection policy summary

The General Data Protection Regulation (GDPR) imposes certain obligations upon Behavioural Insights Limited (BIT), and other companies within the group, as Controllers and / or Processors in relation to processing Personal Data.

BIT takes these obligations seriously. BIT is committed to respecting the rights of all individuals whose personal data it processes:

1. **In relation to data security,** BIT has implemented appropriate measures to ensure the secure storage and handling of Personal Data, including obtaining a Cyber Essentials Plus certification and developing a comprehensive Data Handling Protocol.
2. **In relation to data protection and privacy rights,** our data processing activities are conducted according to the principles relating to the processing of Personal Data set out in the GDPR, including that Personal Data shall be processed lawfully, fairly and in a transparent manner, and in a manner that ensures the security of the Personal Data. BIT has policies and procedures in place to ensure compliance with these principles.

More information on how we handle Personal Data in relation to projects we are working on is detailed below.

BIT is registered with the UK ICO under the terms of the Data Protection Act 2018. Our registration number is ZA038649.

Privacy by design

BIT conducts all trials and research projects with a privacy by design approach to protect and maintain the privacy and security of research participants' and research subjects' data. We work closely with clients, government departments and research partners when designing interventions to ensure that a privacy by design approach is implemented and respected.

Our data protection and data security policies and procedures reflect necessary legislative requirements and set out the standard to which BIT staff should work when dealing with Personal Data, including:

- Attendance at mandatory data protection training for all employees;
- Identifying data requirements from the outset of each project;

- Minimising use of Personal Data where possible and ensuring we have the right to handle any Personal Data where successful project delivery is reliant on using it;
- Putting in place data processing agreements with all clients and suppliers to clarify data handling arrangements ahead of any data being transferred;
- Complying with all relevant data residency requirements and implementing appropriate technical and organisational measures, to protect data and avoid unauthorised access, internally and externally;
- A clear internal reporting process in the event of a data breach, to consider the nature of the breach and identify any necessary action, including whether the breach should be reported to the relevant authorities, i.e. the Information Commissioner's Office in the UK or the Office of the Australian Information Commissioner;
- Clear procedures on retention and destruction of Personal Data to avoid keeping hold of Personal Data longer than necessary for the purposes of each project; and
- Implementing robust investigation and reporting procedures in relation to any data breach or security issues that arise both within our own systems and those of our clients, partners and suppliers.

Data Protection Officer

The BIT group of companies has appointed a Data Protection Officer (DPO) who is the first point of contact for any issue regarding data protection and data security. The DPO can be contacted via email at dpo@bi.team or by writing to us at:

Data Protection Officer, Behavioural Insights Limited, 4 Matthew Parker Street, London, SW1H 9NP, United Kingdom.

19. Appendix III: Intervention descriptions by HEP

The following descriptions summarise the activities in each summer school in the trial. This appendix was amended after the trial had launched, so the list of summer schools below does not match that in Appendix IV, which was the expected list pre-launch (used for power calculations). This is due to low applicant numbers meaning some HEPs could not participate in the trial. TASO will be conducting their own evaluation with HEPs not participating in the RCT and a research protocol for this sub-project has been added as Appendix V.

UCL

Eight subject specific summer schools will run. Several information and guidance (IAG) sessions will be common across all the subjects. The sessions will be delivered virtually through both synchronous and asynchronous modes of delivery, via virtual platforms including Blackboard Collaborate, Zoom and a bespoke platform developed by the Sutton Trust.

Common elements

This includes sessions covering personal statement writing, finances and careers, as well as information sessions for parents. There will also be social sessions, including an online escape room, quiz and a takeaway evening. The week opens and closes with two large group events to welcome pupils and celebrate their completion of the summer school.

Subject specific sessions

Architecture

The theme of the Architecture UCL Summer School Online is "Patterns of Living". Students will discover each other's daily rituals and undertake a critical investigation of their own daily routines in an architectural context through walking, looking, drawing, and making. Students will navigate the week through a series of lectures, podcasts, readings, demonstrations and hands-on tasks.

Astrophysics

Subject specific sessions include virtual tour of the UCL Observatory; lectures (i.e., on coding using Python, space weather and earth management) with supporting Q&A sessions; support sessions; practical workshop (develop a research proposal for a space mission) to then be presented on at the end of the week.

Biosciences

Subject specific sessions include introduction to the staff and summer school; lectures (i.e. preparation for plastic pollution in the oceans and molecular machines); team work on presentations for the end of the week, supplemented by presentation skills and support sessions.

Chemical Engineering

Subject specific sessions include introduction to the teaching team and departmental staff; conducting experiments themselves, both on the computer and in their homes, with supporting results discussion and Q&A; live demonstrations; research skills lecture; designing own experiment to be presented to the group through a poster; talks with both current students and admissions tutor.

Economics

Subject specific sessions include introduction to the staff and the summer school along with an economics walk; lectures (i.e., economics, sports and social media and economists save the world); guest talks on economics graduate experiences; poster creation and presentation; meeting the career and admissions tutors.

Health and Wellbeing Data Science

Subject specific sessions include introduction to the summer school and staff; lectures including recommended reading; seminar sessions to discuss readings and do tasks; practical sessions (i.e., Excel).

History

Subject specific sessions includes general introduction to the staff and summer school, including an icebreaker session, and then daily intro sessions; lectures (i.e. place, space and material culture); seminars (i.e. material culture); independent study with tasks (presentation preparation) and miscellaneous events (tour of British Museum); admissions talk from tutors and student ambassadors.

Natural Sciences

Subject specific sessions include introduction to staff and the programme; lectures; workshops discussing material from lectures; project work sessions, working towards a presentation at the end of the week; admissions and career talk.

Leeds Languages

A five-week programme with two hours of sessions each week focused on either French, German or Spanish, depending on the student's choice. The online workshops are designed to give a taster of studying languages at the University of Leeds, expose students to the career opportunities available to graduates, and to provide the opportunity to meet current languages students. Further sessions include Joint Honors taster sessions, life at university as a languages student, informal networking/social sessions, applying through UCAS, writing a personal statement, Q&A with Leeds alumni and a final celebration to showcase student learnings.

Leeds Social Science

The focus of this five-day event is on the 2021 United Nations climate change 'conference of the parties' (COP26). Students will take part in a range of activities and workshops to tackle and understand how various Social Science subjects engage with climate change, learn about the upcoming COP26 conference, and consider how we can save the planet. There will be six interactive academic workshops, giving students an insight into studying various subjects at university, and how these engage and respond to the topic of climate change. There will also be an opportunity to speak to current students at the University of Leeds regarding what it is like as a student, moving away from home, finances, the transition from school or college to university, clubs and societies, and anything else to do with student life.

Leeds Bioscience

This is a three-day online summer school for year 12 students who are interested in exploring biological sciences at the next level and finding out where it can lead. As well as taking part in lectures, students will be involved in a project of their preference, getting to experience what researching is really like from start to finish. Students will have the opportunity to meet academics and current students from the Faculty of Biological Sciences and take part in a Q&A session to find out more about Leeds undergraduate degree programmes and future career pathways. Sessions will also focus on enhancing a UCAS application, with an opportunity to speak to admissions staff, and get advice on the best way to use what students have learnt at the event in their personal statements.

Leeds Maths

This short Zoom course for students in their first year of sixth form will provide an introduction to mathematics at the University of Leeds. Students will be offered a

preview of university life plus an insight into how mathematics develops at degree level. The University tutors will extend and enrich student's existing study of mathematics at A-level (or equivalent) through lectures and interactive workshops. They will also give students an invaluable insight into the structure of mathematics degrees, the courses, admissions procedure and how to make the most of their application to university. Other sessions will focus on careers specific to this degree, and application support for writing a personal statement and student finance.

Leeds Psychology

The Leeds Psychology Summer School is designed to give Year 12 students an insight into life studying Psychology at university, and the science behind why we behave like we do. Taking place over four days online, participants will take part in subject masterclasses on different areas of psychology, hear from current students about their experiences studying psychology, and learn more about the process of applying to university. This will include sessions specific to UCAS and writing a personal statement and applying for student finance.

Kent

The Kent summer school will be a four-day online event for Year 10 students, using both Zoom and Thinkific. Day one will involve ice-breakers and an introduction to current University of Kent students, followed by activities which focus on barriers to happiness and approaches to positive psychology. The second day will focus on academic barriers. Students will undertake a practical research session to enable them to answer questions around this topic, drawing on examples from multiple disciplines. The third day will focus on the 'Big Question' around community barriers – 'How can we make our communities better places in the wake of Covid-19?' which will involve students creating a submission for a digital time capsule. Students will work in small groups to brainstorm ideas for focuses and mediums for their response and will work independently engaging with relevant materials. The final day will be a launch and celebration event, which parents/carers are invited to. Whilst most activities will be live, students can carry out independent activities at any time using Thinkific.

NTU

Due to Covid-19, the NTU summer school will be a one-day programme, held in school, designed to give pupils a miniature experience of university life. Year 9 pupils will take part in a number of activities alongside a mini research project, for which they will receive a university style criteria and grading upon completion. Pupils will select which mini-lecture (out of a possible 6 courses) they would like to 'attend' and will be given a

question to answer in the form of an academic poster. They will be given time during the day to complete their poster and will be able to utilize the support of the NTU Outreach Ambassadors and member of staff to do this. Pupils will be supervised and supported by NTU members of staff and student ambassadors throughout this process but pupils themselves will be in control and responsible for producing their academic poster. Attendees will also have the opportunity to take part in an activity based on a society or club that NTU offers, for example, knitting. Finally, pupils will receive some information, advice, and guidance on the main aspects of university (e.g., courses, finance, and extracurricular activities.)

20. Appendix IV: Expected characteristics of the applicant pool, places available and quotas specified by HEPs

Table 12. Expected characteristics of the applicant pool, places available and quotas specified by HEPs

HEP	Expected n of Applications	Places Available	Guaranteed places	Quota Characteristics	Proportions for Quotas	Expected proportions of applications with these characteristics
Pre-16 HEPs						
Gloucestershire	148 (all eligible)	60	Children in Care (Estimated to be 2% of applicants)	1.Sex 2.School Provider	1. 50/50 2. 9 schools. 6 places per school.	75% female
Kent	200	60	Children in Care and Young Carers (Estimated to be 2% of applicants)	1.Sex 2.School Provider	1. 50/50 2. 40 schools - even spread from across schools but diff no. of apps from schools	None Specified
NTU	150	80	None Specified	1.Sex 2.School Provider	1. 50/50 2. 8 schools, 10 places per school	50% female
UEA	250	100	None Specified	1. Sex	1. 50/50	More females apply - not specified
Pre-16 total	**748**	**300**	**7**			
Post-16 HEPs						
Surrey	858	250	Care-leavers, GRTSB, Refugee, Asylum Seekers (Estimated to be n=~35; 14%)	1. IMDQ and or/POLAR4 Q1, FSM	60%-70%	60% low-SES 40% non-low-SES
UCL Architecture	99	37	Care-leavers and estranged students (1-3 per year across all subjects for UCL)	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M

UCL Population Health	99	37	0	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M
UCL Biosciences	99	37	0	1. Sex	1. 50/50 (as close to as possible)	60% F and 40% M for STEM subjects
UCL Chemical Engineering	99	37	0	1. Sex	1. 50/50 (as close to as possible)	60% F and 40% M for STEM subjects
UCL Philosophy	99	37	0	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M
UCL Astrophysics	99	37	0	1. Sex	1. 50/50 (as close to as possible)	60% F and 40% M for STEM subjects
UCL Fine Art	99	37	0	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M
UCL Geography	99	37	0	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M
UCL History	99	37	0	1. Sex	1. 50/50 (as close to as possible)	roughly 65% F, 35% M
UCL Nat. Sciences	99	37	0	1. Sex	1. 50/50 (as close to as possible)	60% F and 40% M for STEM subjects
Suffolk	200	70	0	1. Sex	1. 50/50	60% female
Leeds Broad	200	100	0	1. Sex	1. 50/50	65% female
Leeds Languages	65	30	0	1. Sex	1. 50/50	65% female
Leeds Psychology	65	30	0	1. Sex	1. 50/50	65% female
Leeds Maths	65	30	0	1. Sex	1. 50/50	60% F and 40% M for STEM subjects

Leeds Food Sciences	65	30	0	1. Sex	1. 50/50	65% female
Leeds Biological Science	65	30	0	1. Sex	1. 50/50	60% F and 40% M for STEM subjects
Leeds Social Sciences	65	30	Care-leavers (1 applicant last year)	1. Sex	1. 50/50	65% female
Post-16 total	**2,638**	**966**	**37**			

21. Appendix V: Sub-project research protocol – pre-test post-test analysis

Summary

Background

Five of the eight HEPs participating in the summer school evaluation did not receive enough applicants to be able to be involved in the RCT. Subsequently all applicants for these summer schools received a place. Students responded to survey items at application stage and will do so again post-summer school. HEPs will therefore have baseline and post-event survey data for all applicants allowing for pre-test post-test analysis of survey responses. This analysis will form a discrete sub-project designed by TASO and only including HEPs unable to participate in the main trial.

Aims

To conduct pre-test (pre-summer school) post-test (post-summer school) analysis and assess the change in student survey outcomes during the time that they participate in the summer school.

Intervention

Five HEPs will deliver their own summer schools, either for students in pre-16 or post-16 education.

Design

This study is a matched pairs pre-test post-test design.

Outcome measures

The change in student survey responses post-intervention compared to pre-intervention.

Analyses

The analysis will establish whether the mean difference between pre-test and post-test scores in applicants are different from each other. (Note: the assumptions supporting this approach will need to be checked using the data.)

Background

The five HEPs involved in this sub-project did not receive enough applicants to be able to participate in the Randomised Controlled Trial (RCT) so will instead be conducting their own evaluation of their summer schools using pre-test post-test analysis of the survey responses.

The overall project, and therefore this sub-project, is funded by TASO, and TASO is funded by the Office for Students (OfS), the independent regulator of higher education in England.

Some HEPs have used the funding to hire a research associate (RA) to support them with their evaluation responsibilities. Others will be conducting the evaluation using current staff in the Evaluation and Widening Participation (WP) teams. The table below summarises the key project personnel for each organisation.

Table 13. Sub-project personnel

Organisation	Name	Role and responsibilities
TASO	Sarah Chappell	Research Officer and Sub-project Lead. Day-to-day support for HEPs, conducting aggregate analysis, and write-up of final report combining the reports from all universities.
	Helen Lawson	Research Programmes Manager. Overall Project Lead. Sub-project support.
	Jess Hunt	Head of Research. Project support and QA.
University of Surrey	Katherina Sela	Project lead at the University of Surrey. Responsible for data collection and production of final report.
	Karla Lopez Murillo	Research Associate. Support with data collection, analysis, and final report.
University of Suffolk	Marinna Stella	Project lead at the University of Suffolk. Responsible for data collection and production of final report.

	Owen Evans	WP practitioner. Support with data collection including survey responses and cognitive interviews. Support with final report.
University of Leeds	Liz Hurley	Project lead at the University of Leeds. Responsible for data collection and production of final report.
	Rebecca	Research Associate. Support with data collection, analysis, and final report.
University of East Anglia (UEA)	Rosie Hannant	Project lead at UEA. Responsible for data collection and production of final report.
	Ed Penn	Research Associate. Support with data collection, analysis and final report.
University of Gloucestershire	Liz Gray	Project lead at the University of Gloucestershire. Responsible for data collection and production of final report.

Aims

There is evidence to suggest that summer schools are positively correlated with an increase in confidence and aspirations related to higher education.

Evaluations of summer schools including residentials in the East Midlands found that secondary school participants had significantly improved their knowledge on how to apply to university, of the difference between course options, and of the best university for their subject of interest (Hayes et al., 2018; Church, 2018). Similarly, following a two-night residential summer school, students reported increased aspiration to go to university and a better understanding of the steps that needed to be taken to achieve this goal (Aspire to HE: Higher education for all, 2018). Hatt et al (2009) issued pre and post summer school questionnaires to Year 11 students and found that they felt significantly more confident in their ability to fit in at university and to successfully enter HE, after the summer school.

As well as increasing confidence, summer schools are also associated with a reduction in fears around enrolling in HE. Students with autism reported a significant reduction in concerns relating to transition to university, specifically, support, leaving home, and academic issues, after they attended a residential summer school (Lei et al 2020).

However, due to the COVID-19 pandemic, summer schools this year are taking place online with no residential element. Whether online summer schools can have the same effect as face-to-face delivery is unknown. Through analysing survey responses before and after five virtual summer schools, we will have some idea as to whether online delivery can replicate the benefits of residential summer schools.

It is important to note that this approach will not provide causal evidence as to the impact of the university summer schools, due to the lack of comparator group. It therefore does not eliminate the possibility that positive outcomes might have occurred regardless of the intervention. Despite this, the sub-project intends to produce valuable guidance for the sector in carrying out analysis on survey responses captured before and after their interventions. Furthermore, HEPs will be carrying out an implementation and process evaluation (IPE) alongside this sub-project to support in interpreting any findings.

Intervention

This sub-project will look to evaluate a collection of interventions. Five HEPs will deliver their own summer schools, either for students in pre-16 or post-16 education. Each summer school will have its own specific characteristics, but all have the same broad aims and involve similar activities.

Due to the COVID-19 pandemic, these interventions have been redesigned for online delivery:

University of Surrey

The Surrey Virtual Summer School is a 5-day online programme for Year 12 students. The activities include interactive subject-specific lectures and seminars, university information, such as on personal statements, finance, interviewing and choosing a university and course, and sessions to support the development of the participants' study skills to prepare them for the transition to university.

In addition, the participants’ will work in groups to complete a research project in their specialised subjects, which they will present virtually to their peers, university academics, and staff. Additionally, a variety of evening entertainment including quizzes, clubs and society tasters and well-being activities will be offered.

University of Suffolk

The Suffolk summer school is focused on Year 12 students and will be completed over four days using the online platform UniBuddy. The intervention will consist of

activities on various topics focusing on university life, studying at the University of Suffolk, support on offer, how to apply, and more. Participants will also get a chance to work together on collaborative projects.

University of Leeds (Leeds)

The Leeds Broad 'Reach for Excellence' summer school is targeted at Year 12 students. The activities will 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 such as Team building activities to encourage engagement with the event and each other, and 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. There will also be a Q&A and social time with current UG students.

University of East Anglia (UEA):

The UEA summer school targets Year 10 students and will be delivered entirely online. Participants will take part in five synchronous group sessions, accompanied by asynchronous individual follow-up tasks intended to encourage reflection on the activities participated in. Students will work in groups to produce a marketing campaign to promote the non-academic benefits on university to their peers, exploring the topics of belonging, independence, agency and exploration.

University of Gloucestershire

The Gloucestershire summer school targets Year 10 students and takes place across three days using online platforms eLearn, Teams and Zoom. Activities are split into four sections: information on HE, subject specific, social, and wellbeing. HE information sessions include an online campus tour, Q&A with current university students, planning your future, and student finance. Subject specific sessions will be a range of subject tasters delivered by an academic and student ambassador.

Design

This is a pre-test post-test design where the survey items outlined in the main protocol are administered first at baseline prior to the intervention, and again after the intervention. Analysis will establish whether there is a statistically significant difference in the mean scores for pre-summer school responses and post-summer school

responses. HEPs will conduct this themselves locally on their own summer school applicant data, whilst TASO will combine the data from each HEP to carry out aggregate analysis.

Outcome measures

The outcome measures correspond to the survey items covered in the main trial research protocol. The primary outcome measures will include the six (seven for pre-16) questions related to HE, whilst the exploratory outcome measures include the two questions regarding COVID-19 specifically.

Sample selection

The study sample will be made up of applicants to the five HEP's participating in the sub-project evaluation that meet the HEPs' eligibility criteria.

The size of the sample is determined by:

- The number of eligible applicants for each summer school,
- The removal of any students who are already a part of the main trial[19],
- The number of applicants who consented to participate in the research (collected at application stage).

The total sample size by HEP and combined is given in Table 14 below.

Table 14. Actual sample size by cohort

Cohort	Actual sample size
Gloucestershire	56
UEA	22
Pre-16 total	**78**
Suffolk	20
Surrey	224
Leeds	45

[19] Some applicants applied to more than one summer school. Duplicate checking was carried out and any students who were already randomised for the main trial were not offered a place.

Post-16 total	**289**
Combined	**367**

Data collection

The survey data will first be collected when participants fill in a summer school application with the HEP. The time point for this will vary between HEPs depending on when their summer schools are running and therefore when they need to open applications. The same survey data will then be collected after the summer school has been delivered, which will again vary by HEP depending on their timeline.

HEPs will also be collecting student data to aid in matching pre and post intervention survey responses, and to allow for descriptive statistics by demographics, e.g., sex, and family experience of HE.

Finally, students will be asked about any additional outreach activities they have previously taken part in, in addition to the summer school (see Annex A for the full set of questions). This data will support in interpreting the analysis of the primary outcomes.

All individual items of data are listed in table 15 below, along with the collection point and the cohort being asked (pre-16, post-16, or both).

Table 15. Sub-project data

Data item	Collection point	Collector	Cohort
Sex	Baseline (Jan to June 2021)	HEPs	Both
Ethnicity	Baseline (Jan to June 2021)	HEPs	Both
Postcode-level marker of disadvantage	Baseline (Jan to June 2021)	HEPs	Both
Free School Meal (FSM) status	Baseline (Jan to June 2021)	HEPs	Both

Whether anyone in the family has been to university	Baseline (Jan to June 2021)	HEPs	Both
First name	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Last name	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Date of birth	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Postcode	Baseline (Jan to June 2021)	HEPs	Both
Academic year group	Baseline (Jan to June 2021)	HEPs	Both
School name	Baseline (Jan to June 2021)	HEPs	Both
Likelihood of going to HE	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Likelihood of progressing to academic study post-16	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Pre-16
Self-efficacy relating to application success	Baseline (Jan to June 2021)	HEPs	Both

	After event (July to Sept 2021)		
Self-efficacy relating to post-entry success	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Compatibility of HE with social identity	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Perception of financial barriers to HE	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Perception of knowledge barriers to HE	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Effect of COVID-19	Baseline (Jan to June 2021) After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: IAG / general HE information	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: IAG for student finance	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: Campus visit	After event (July to Sept 2021)	HEPs	Both

Capturing additional outreach: Skills and attainment (tutoring)	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: Skills and attainment (practice interviews)	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: Skills and attainment (personal statements)	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: Subject taster/insight	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: Mentoring	After event (July to Sept 2021)	HEPs	Both
Capturing additional outreach: multi-intervention outreach programme	After event (July to Sept 2021)	HEPs	Both

Procedure

Timeframe	Action
July 2021	• Project plan sent to HEPs • Kick-off meeting with HEPs
July 2021 – September 2021	• HEPs administer post-intervention surveys
September – November 2021	• HEPs carry out pre-test post-test analysis of survey items • TASO carry out aggregate analysis
December 2021	• HEPs informally present findings to TASO at collaborative meeting

December 2021 – February 2022	• HEPs carry out report write-up
March 2022	• Final report from HEPs due

Power calculations

Our assumptions are:

- Significance level: 0.05
- Power: 0.8

Sample	Sample size	MDES[20]
Gloucestershire	56	A paired samples t-test with 56 participants would be sensitive to effects of Cohen's d = 0.34 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.34.
UEA	22	A paired samples t-test with 22 participants would be sensitive to effects of Cohen's d = 0.55 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.55.
Pre-16	78	A paired samples t-test with 78 participants would be sensitive to effects of Cohen's d = 0.28 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.28.
Suffolk	20	A paired samples t-test with 20 participants would be sensitive to effects of Cohen's d = 0.58 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.58.
Surrey	224	A paired samples t-test with 224 participants would be sensitive to effects of Cohen's d = 0.17 with 80% power (alpha = .05, one-tailed).

[20] Cohen (1988) suggests that that d = 0.2 be considered a 'small' effect size, 0.5 represents a 'medium' effect size, and 0.8 a 'large' effect size. TASO are expecting the majority of HEPs to be underpowered, however, the goal is to upskill the sector in using this design. Aggregate analysis will be powered.

		This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.17.
Leeds	45	A paired samples t-test with 45 participants would be sensitive to effects of Cohen's d = 0.38 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.38.
Post-16	289	A paired samples t-test with 289 participants would be sensitive to effects of Cohen's d = 0.15 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.15.
Pre- and post-16 combined	367	A paired samples t-test with 367 participants would be sensitive to effects of Cohen's d = 0.13 with 80% power (alpha = .05, one-tailed). This means the study would not be able to reliably detect effects smaller than Cohen's d = 0.13.

Analytical strategy

The following model will be used to estimate the effects of the intervention on each of the primary and exploratory outcome measures:

$$t = \frac{\underline{x}_{diff} - 0}{s_{\underline{x}}}$$

where,

$$s_{\underline{x}} = \frac{s_{diff}}{\sqrt{n}}$$

where,

- $\underline{x}_{diff}$ = the mean difference between the pre and post intervention scores across the sample
- n = sample size (number of observations)
- s_{diff} = sample standard deviation of the differences
- $s_{\underline{x}}$ = estimated standard error of the mean (standard deviation divided by the square root of the sample size)

The calculated *t* value will then be compared to the critical *t* value with df = *n* - 1 from the *t* distribution table for a 95% confidence level. If the calculated *t* value is greater than

the critical *t* value, we can conclude that the pre and post intervention means are significantly different.

For the above analysis to be carried out, the data needs to be normally distributed. If this isn't the case, the data can be transformed using the natural logarithm or the square root, which will likely change the underlying distribution from skewed to normal. If the data is still non-normal, the Wilcoxon matched pairs signed-rank test can be used as follows:

$$W = \sum_{i=1}^{n_r} [sgn(x_{1,i} - y_{1,i}) \cdot R_i]$$

where,

- W = test statistic
- N_r = sample size, excluding pairs in the data (pre and post intervention scores) where $x_1 = y_1$
- sgn = sign function
- $x_{1,i}$, $y_{1,i}$ = corresponding ranked pairs from the pre and post intervention score distributions
- R_i = rank i

Analysis will be conducted on complete cases only, i.e., where participants have responded to both the pre and post intervention surveys. Given that the survey is short there is unlikely to be any missing data for individual items, but rather missing sets of data when participants have not responded to the post-intervention survey at all.

Further descriptive analysis will explore how perceptions of COVID may have impacted responses to the primary outcome questions. For instance, participants may be split into 'more' and 'less' worried about COVID, and their pattern of survey responses analysed. Descriptive analysis will also be conducted on the additional outreach survey responses to aid in interpreting responses to the primary outcome measures.

Risks

Part of evaluation	Risk	Mitigation strategy	Risk owner
Data collection	As the outcome measures are survey-based, they may yield small samples and be subject to differential attrition.	Some HEPs have used their TASO funding to hire an RA to facilitate data collection. Some HEPs are also using their funding to provide incentives for survey completion. HEPs are funded to take part in the project – so there is buy-in.	TASO, HEPs
Power/ analysis	Small sample size at provider-level which will leave the majority underpowered.	Carrying out aggregate analysis of all HEP data.	TASO

Bibliography

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

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 Participaiton Research and Evaluation Partnership.

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.

Annex A: Capturing additional outreach survey items.

Construct	Survey item/s	Response Options
Capturing additional outreach: IAG / general HE information	I have received information, advice and guidance about university life, how to apply to university, or how to make course choices.	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: IAG for student finance	I have received information, advice and guidance about student finance.	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Campus visit	I have visited a university campus	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Skills and attainment (tutoring)	I have taken part in tutoring run by a university to support my grades at school/college.	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Skills and attainment (practice interviews)	I have participated in practice interviews to help with my application to university.	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Skills and attainment (personal statements)	I have received advice and guidance about how to write a personal statement from a university.	Never Once 2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Subject	I have participated in university subject taster sessions (e.g., a short lecture,	Never Once

taster/insight	workshop, or talk from an academic staff member or current student).	2 to 5 times 6 to 10 times 11 to 15 times 16 times or more Not sure
Capturing additional outreach: Mentoring	I have received help from a university student mentor or role model. Note, this may have been face-to-face or online.	Yes No Not sure
Capturing additional outreach: Multi-intervention outreach programme	I have participated in a university outreach programme that involved multiple activities over multiple months.	Yes No Not sure