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*]	24/03/2021	NA
Pre-registration		This design has been pre-registered on the Open Science Framework (OSF) registry.

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

Transforming Access and Student Outcomes in Higher Education

Contents

Transforming Access and Student Outcomes in Higher Education

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.
University of Surrey	Katherine Sela	Project lead at the University of Surrey. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
University College London (UCL)	Shireen Quraishi	Project lead at UCL. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
University of Leeds	Liz Hurley	Project lead at the University of Leeds. Responsible for implementing randomisation and data collection there.

	RA (TBC)	Supporting data collection.
University of Suffolk	Marianna Stella	Project lead at the University of Suffolk. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
University of Gloucestershire	Fiona Curry	Project lead at the University of Gloucestershire. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
University of Kent	Marta Almeida	Project lead at the University of Kent. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
Nottingham Trent University (NTU)	Peter Cassidy	Project lead at NTU. Responsible for implementing randomisation and data collection there.
	RA (TBC)	Supporting data collection.
University of East Anglia (UEA)	Rosie Hannant	Project lead at the UEA. Responsible for implementing randomisation and data collection there.
	RA (TBC)	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).

[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.

4. Intervention

4.1. Introduction

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. Each summer school will have its own specific characteristics, but all have the same broad aims and involve similar activities. At the time of writing the protocol, logic models and detailed activity plans have not been developed. This is partly because business-as-usual models are being redesigned for online delivery due to Covid-19. These plans are due to be complete by 30th April 2021, at which point the protocol will be updated and republished. 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.

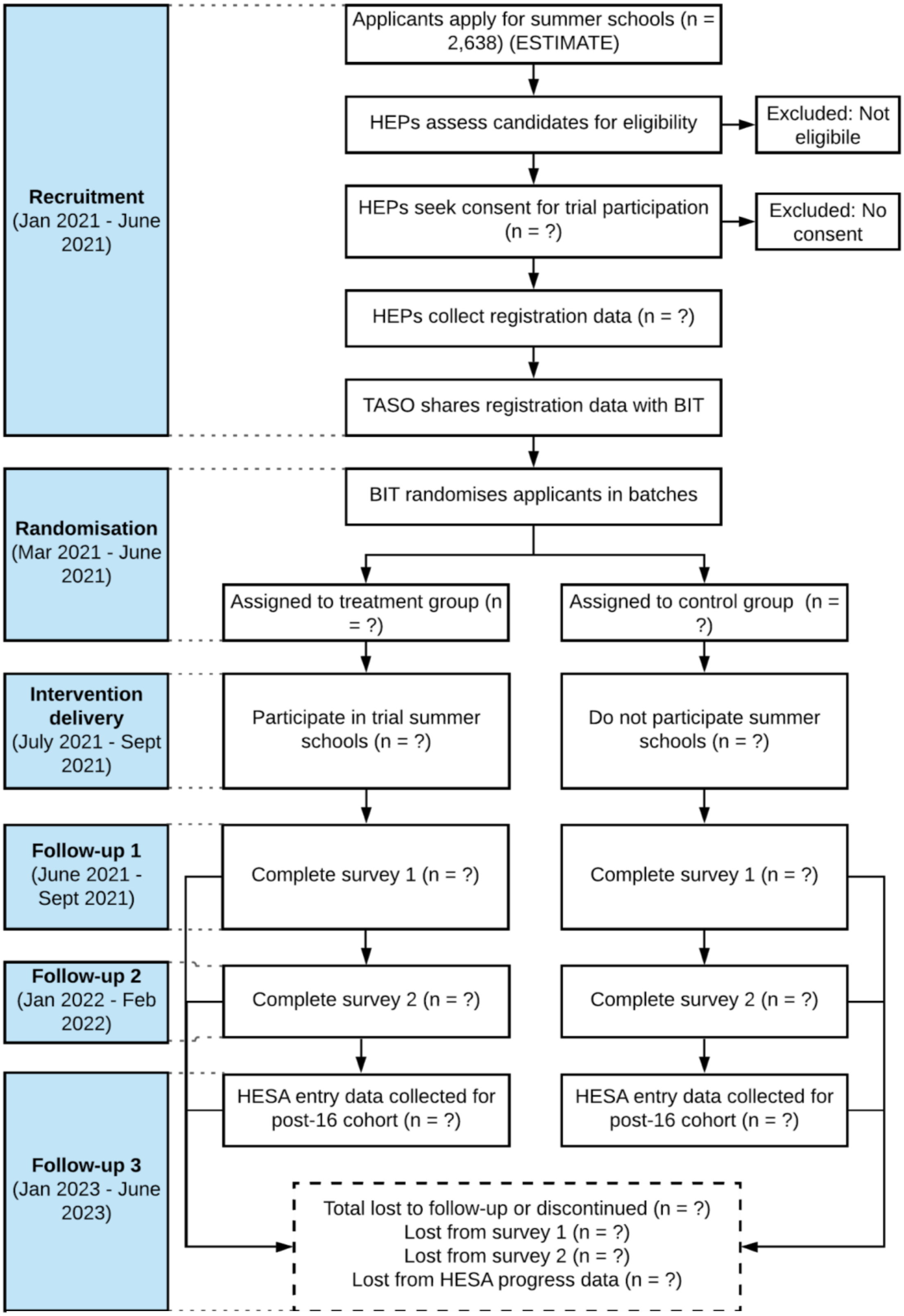
5. Design

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.

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

Figure 1. Study flow diagram



Notes: ‘Registration data’ includes baseline survey data.

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?	NA	After endpoint (June 2023)	Post-16 only

	Binary: yes/no			
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?	NA	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.

	Likert: 5-point "Extremely confident" to “Not confident at all”			
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

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.

[5] All figures rounded to nearest 10.

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 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]

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	Is condition (i) met? (>20)	Condition (ii) How many	Is condition (ii) met?

[6] All figures rounded to the nearest 10.
[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.

	average block size?[9]		blocks contain n ≤ 4?	
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	✔

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.

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

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.
    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 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

[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.

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	Baseline After	HEPs NPD	Both

	Control variable Assessing external validity	endpoint	(TBC)	
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	Baseline	HEPs	Both

	data Assessing external validity			
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 endpoint	HEPs NPD (TBC)	Both
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. At the time of writing the protocol, the summer schools are undergoing a redesign process to move delivery online. The providers are unable to define minimal compliance at this point, so this will be added to the protocol by 30th April 2021. This will be prior to intervention launch and before any outcome data is collected.

We do not know the true minimal dosage needed to generate a treatment effect, so the cut-off chosen for compliance will be based on the providers' best estimate. 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:

$$T_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \eta_i \text{ (1)}$$

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

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

17. 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.

19. 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**)

21. 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.

23. Appendix III: Intervention descriptions by HEP

Pre-16 providers

University of Gloucestershire

A unique opportunity for Year 10 students to experience university life and try new subjects that are not available at school. These subjects include Natural and Social Sciences, Arts and Performing Arts, Health and Social Care, Business, Computing and Technology, Media, Education, Humanities and Sport. Successful applicants will have the chance to work with current university students and learn about student life, finance, and study skills.

University of Kent

Year 10 students are introduced to a range of different academic subjects available at university and experience the student lifestyle. Each student will be appointed a mentor, who will be a current student at the University of Kent. The mentors will ensure that students feel at home whilst attending the Summer School and offer an additional point of contact to find out more information about what it is like to be a university student.

Nottingham Trent University (NTU)

Year 9 students take part in a 3-day summer school to experience the life of a university student. Activities include ice breakers, introduction to university study and research skills, subject tasters and student life social activities. Students are required to complete a self-reflective journal as part of the summer school.

University of East Anglia (UEA)

The content of the summer school focuses on supporting Year 9 and 10 students to develop “social and academic capital”, which centres around improving students’ knowledge and decision-making capacity within the HE landscape. The programmes include a mix of academic tasters, HE information workshops, and social activities with current students. The Year 9 and year 10 summer schools end with a finale to which parents and carers are invited, with the aim of informing and inspiring them about the potential of their young person’s future.

Post-16 providers

University of Leeds

This provider runs subject specific summer schools as well a more general HE focused summer school. 'Linguastars' is aimed at students interested in Arts, Humanities and Cultures Courses, and 'Social Sciences' is aimed at students interested in studying subjects such as Psychology and Sociology. The Reach for Excellence Summer School provides a wide range of activities including team building, preparation for and transition to HE, subject specific support, study skills, enterprise, personal statement support and social activities. It aims to give an all-round experience of preparation for Year 13 and the transition to HE.

University of Suffolk

This summer school is made up of a range of academic and social activities. Students can select four different academic taster sessions based on their interests. These interactive, activity led sessions give students an insight into what studying that subject at degree level is like. Participants learn first-hand about the academic relationship between lecturers and students and experience activities used by academics in their teaching to successfully engage the diversity of students the provider supports. These carefully planned and accessible activities are designed to capture the interest of students and allow them to recognise that university life is available to them. As part of this approach, students also collaborate in a group academic project to assist their development of research and presentation skills. The overall aim is to provide a social, collaborative learning space where those who attend can gain useful and transferable skills to take forward into their futures.

University of Surrey

Students will participate in lectures, labs and seminars in their chosen subject area as well as tasters in related fields they might not have previously considered. Subject options include Engineering, Midwifery, Medicine, Computer Science and Languages. There are also sessions on UCAS application, interview techniques, student finance, study skills, and university support. In the evenings, students will take part in a variety of social activities to develop a sense of belonging and confidence including games, quizzes and take-away nights (though these will take a different form online).

University College London (UCL)

The summer school is designed to give participants a taste of studying in London at UCL and studying a particular subject. Students apply for a particular subject strand and work on a project throughout the week, culminating in a presentation. Each subject strand contains a combination of lectures, seminars, practical sessions and visits designed to give participants an insight into studying that subject at university in an engaging, interactive way. Participants are split into small groups each led by a current student who acts as a mentor throughout the week. The small group size helps them to bond, build their confidence and get to know their student leader. Throughout the week information, advice and guidance sessions are included focusing on topics such as student funding and application support

24. 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	65	30	0	1. Sex	1. 50/50	65% female

Psychology						
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**			