Research protocol

TASO Summer Schools Evaluation - Phase 2

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.		
1.2 [original]	06/04/2022	NA
1.3	04/10/2022	Adding definitions of compliance to section 12.13.
1.4	01/02/2023	Adding intervention descriptions as Appendix III.
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. The QA rating below was conducted by the BIT Principal Research Advisor who QAd the protocol.

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	Sample size for the primary outcome per arm after expected attrition is between 500 and 1500 participants. While the existing evidence is very limited, even the best case scenario for the MDES (5.8pp) seems high. The literature suggests that the maximum effect of summer schools of this type on progression to HE is likely to be 2.3pp (see section 11).	3
Outcome measure	The primary outcome measure is a direct measurement of the desired outcome (HE attendance in 2023/24). It is unambiguously defined and the data is expected to be of high quality. The secondary outcome is equally a direct measurement of the desired outcome (HE attendance at the summer school institution).	5
Attrition	Post-randomisation attrition is expected to be <20% for the primary outcome, which will be collected from a government administrative dataset (HESA) and has been factored into power calculations. Non-compliance could be more of an issue, though there is a CACE analysis specified to measure this.	4
Validity	No threats to validity identified.	5

Overall		4.4

BIT protocol ref: 2022018

Contents

1. Summary

Background

This project is a collaboration between the Centre for Transforming Access and Student Outcomes in Higher Education (TASO), five Higher Education Providers (HEPs) and the Behavioural Insights Team (BIT). In summer 2022, a series of summer schools 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. Five 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 2023/24 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 Ordinary Least Squares (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), five Higher Education Providers (HEPs) and the Behavioural Insights Team (BIT). In summer 2022, 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. This is the second phase of a research project about the impact of university summer schools. Phase 1 is evaluating a collection of summer schools that were delivered in summer 2021. These summer schools were delivered online due to COVID-19 restrictions. The protocol for phase 1 can be found here. The reports for phase 1 will be published in May and a link to these will be included in an updated version of this TP. The roles and responsibilities for the phase 2 evaluation are as follows.

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, data collection, 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;
- designing and implementing the IPE.

The five HEPs are responsible for:

[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.

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

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

Organisation	Name	Role and responsibilities
BIT	Dr. Patrick Taylor	Evaluation Manager
	Ruth Persian	Evaluation QA
	Dilhan Perera	Evaluation Supervisor
	Pujen Shrestha	Data Analyst
TASO	Dr Helen Lawson	Research Programme Manager. IPE Lead and overall responsibility for the project.
	Sarah Chappell	Senior Research Officer. RCT Lead and 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 Kent	Amy Burt	Project lead at the University of Kent. Responsible for implementing randomisation and data collection there.
	RA TBC	Supporting data collection.
Nottingham Trent University (NTU)	Laura Hope	Project lead at NTU. Responsible for implementing randomisation and data collection there.
	RA TBC	Supporting data collection.
University of Gloucestershire	Liz Gray	Project lead at the University of Gloucestershire. Responsible for implementing randomisation and data collection there.

	Hannah Kent	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 Leicester	Hannah Grosvenor	Project lead at the University of Leicester. 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).

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 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 2-4 days, with students staying in

[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.

campus accommodation. Students also 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 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 stay in campus accommodation for 2-4 days and have tours of the university campus. Students will also experience subject tasters, unless the summer school they have applied to is focused on one subject stream in particular, 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 and course.

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 2022 and November 2024[3] (including final reporting). Fig. 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 on TASO providing final outcome data with BIT by the end of July 2024.

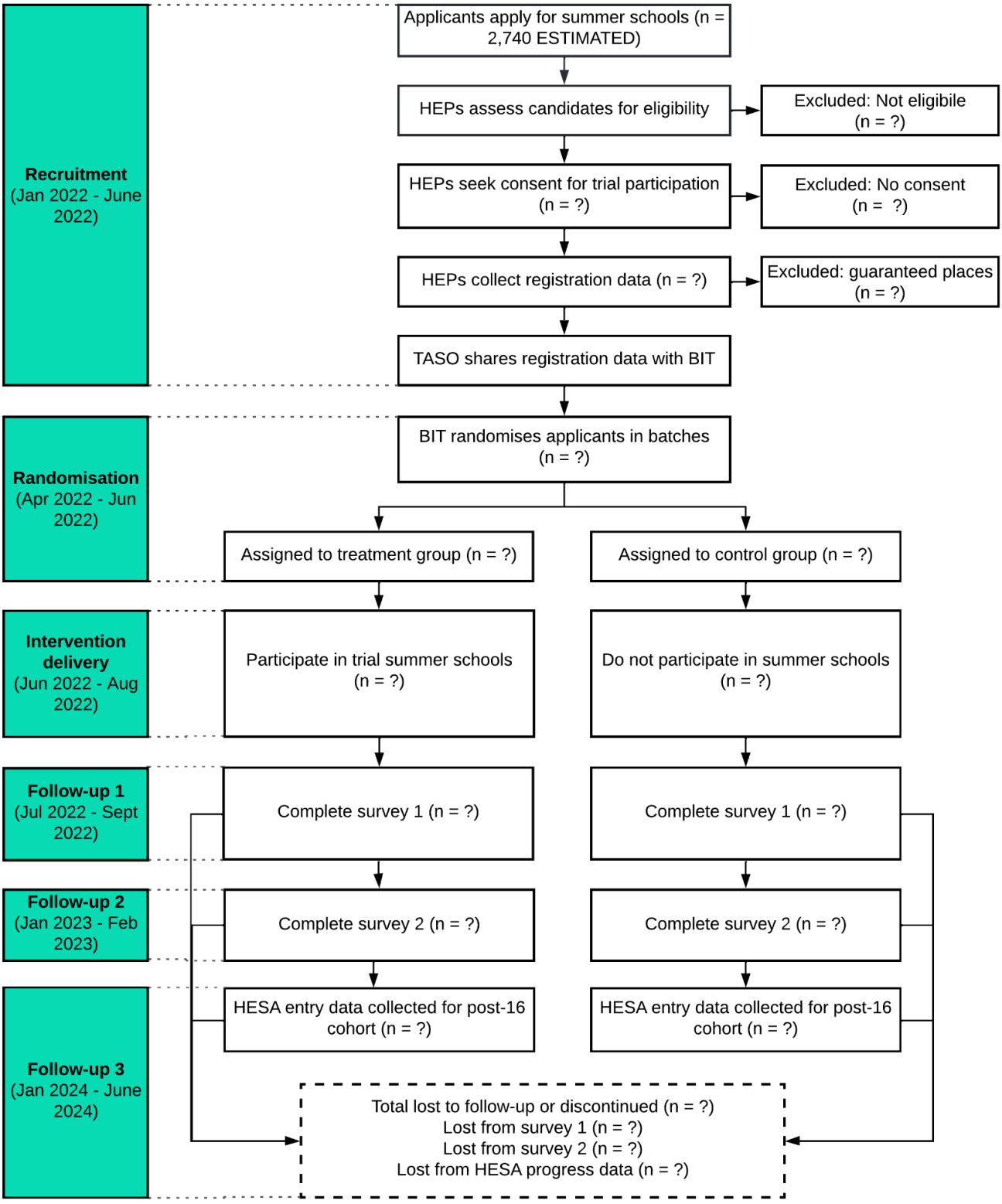

Figure 1: Trial flow diagram

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

These definitions are used here to help clarify the intervention's theory, but also to determine some important analytic choices. The primary outcome 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.

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 2023/24 according to the HESA dataset? **Binary:** yes/no	NA	After endpoint (June 2024)	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 in 2023/24? Binary: yes/no	NA	After endpoint (June 2024)	Post-16 only
EXPLORATORY 1 (PROXIMAL): Application to HE	Survey 2: Have you applied to university? Binary: yes/no	NA	After endpoint (January 2023)	Post-16 only
EXPLORATORY 2 (PROXIMAL): Likelihood of going to HE	Survey 1: How likely are you to apply to university? Likert: 7-point "Extremely likely to extremely unlikely"	NA	Baseline After endpoint (Sept 2022)	Both
EXPLORATORY 3 (PROXIMAL): Likelihood of progressing to academic study post-16[4]	Survey 1: How likely is it that you will study at school or a sixth form after you've finished Year 11? Likert: 5-point "Extremely likely to extremely unlikely"	NA	Baseline After endpoint (Sept 2022)	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”	Mean average	Baseline After endpoint (Sept 2022)	Both
EXPLORATORY 5 (MEDIATOR): Compatibility of HE with social identity	Survey 1: How much do you agree with the following: "University is for people like me"? Likert scale: 5-point ‘‘strongly agree to strongly disagree’’	NA	Baseline After endpoint (Sept 2022)	Both
EXPLORATORY 6 (MEDIATOR): Perception of	Survey 1: 1. How confident are you that you could afford to go to university?	Mean average	Baseline After endpoint	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.

practical barriers to HE	2. How confident are you that you know how to apply to university? Likert: 5-point "Extremely confident" to "Not confident at all"		(Sept 2022)	

Table 2: Outcome measures

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[5]:

- identify as coming from a black or minority ethnic background;
- identify as Gypsy, Roma, or Traveller;
- live in an area of deprivation (as defined by the most deprived quintile (Q1) of the Index of Multiple Deprivation (IMD) and/or the participation of local area in higher education (POLAR) classification);
- be care-experienced;
- be estranged (students who have lost contact with their parents and/or are studying without the support of their parents);
- be a young carer;
- have a disability;
- be the first in their family to attend HE;
- 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;
- have a low household income (£25,000 per annum or below);
- have had their studies disrupted by circumstances in their personal, social or domestic lives (for example, through trauma, medical or mental health issues); and/or
- be a refugee or asylum seeker.

The University of Kent has the additional criterion that all applicants must attend a school that partners with the university. This means that only students who attend a

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

partner school will be deemed eligible for the summer school. Partner schools have a longstanding relationship with the university based on their location and high percentage of students that meet widening participation criteria.

The sample 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 of post-16 education.

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

Cohort	Estimated sample size	Size of treatment group
Pre-16	550	200
Post-16	2,190	625
Combined	2,740	825

Table 3: Estimated sample size by cohort

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. Some HEPs have quotas that they want 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.

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 want 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 stratum 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, it is part of BIT's analysis policy that 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 stratum/block is the combination of summer school x quota (for example, Leeds Dentistry x female).

In the Phase 1 trial, when these rules were applied, no quotas could be forced through stratification, so randomisation was stratified only at the HEP level.

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.[6] Balance checks will be conducted on all of the control variables used in the primary analysis.

[6] Both post-16 and pre-16 participants will read an information sheet about the research and have the opportunity to opt out of participation.

8.3. Randomisation procedure[7]

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

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

These spreadsheets will be securely sent to BIT in batches. 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.
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.

[7] This section describes the randomisation procedure for the trial. Eligible applicants who do not consent to participation in the trial will also be randomly assigned to either participate in the summer school that they apply to or not. This will affect the number of places available in the treatment group for the trial. BIT will carry out this randomisation before randomising consenting applicants into the treatment or control group for the trial. HEPs will only share the study IDs of non-consenters with TASO (who will share these with BIT) for the purposes of this randomisation procedure.

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. 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 and where they are eligible for a guaranteed place.
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. If a sex or socioeconomic status (SES) quota cannot be fulfilled, 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
8. Cost evaluation interviews, conducted by BIT

All individual items of data to be collected are listed in Table 4 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 is 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.

Data item	Purpose	Collection point	Source	Collector	Sample
Student data					
HEAT ID	Matching datasets shared by TASO and checking for duplicates	Baseline	HEAT database	TASO	Both
Sex	Meeting treatment group quota (stratification) Balance checks Control variable Assessing external validity	Baseline After endpoint	Self-report registration form	HEPs HESA	Both

Ethnicity	Balance checks Control variable Assessing external validity	Baseline After endpoint	Self-report registration form	HEPs HESA	Both
Postcode-level marker of disadvantage (IMD, POLAR and IDACI)	Meeting treatment group quota (stratification) Control variable Assessing external validity	Baseline	HEAT database	TASO	Both
Free School Meal (FSM) status	Meeting treatment group quota (stratification) Balance checks Control variable Assessing external validity	Baseline After endpoint	Self-report registration form	HEPs NPD (TBC)	Both
Whether anyone in the family has been to university	Balance checks Control variable Assessing external validity	Baseline	Self-report registration form	HEPs	Both
Disability status	Assessing external validity	Baseline	Self-report registration form	HEPs	Both
Experience of children's social care	Meeting treatment group guaranteed places Assessing external validity	Baseline	Self-report registration form	HEPs	Both
Whether from an underrepresented group (Young carer, estranged, Gypsy, Roma, Traveller	Meeting treatment group guaranteed places Assessing external validity	Baseline	Self-report registration form	HEPs	Both

communities, refugees, children of military families)					
First name	Uniqueness check Accessing HESA and NPD data	Baseline	Self-report registration form	HEPs	Both
Last name	Uniqueness check Accessing HESA and NPD data	Baseline	Self-report registration form	HEPs	Both
Date of birth	Uniqueness check Accessing NPD data	Baseline	Self-report registration form	HEPs	Both
Postcode	Uniqueness check Accessing HESA and NPD data	Baseline	Self-report registration form	HEPs	Both
Academic year group	Balance checks Control variable Accessing HESA and NPD data Assessing external validity	Baseline	Self-report registration form	HEPs	Both
School name	To identify School ID	Baseline	HEP admin data	HEPs	Both
School location	To identify School ID	Baseline	HEP admin data	HEPs	
School ID (URN)	Uniqueness check Control variable Accessing HESA and NPD data	Endpoint	HEP admin data	HEPs	Both
Summer school applied to	Subgroup analysis	Baseline	HEP admin data	HEPs	Both

	Control variable				
Pre or post-16 programme	Subgroup analysis	Baseline	HEP admin data	HEPs	Both
Summer school attended	Estimating effects of intervention Compliance check	Endpoint	HEP admin data	HEPs	Both (treatment group only)
Attainment at Key Stage 2 Maths and English	Control variable Assessing external validity	Baseline After endpoint	Self-report registration form NPD	HEPs BIT	Both
Attainment at Key Stage 4 (Attainment 8 score)	Control variable Assessing external validity	Baseline After endpoint	Self-report registration form NPD	HEPs BIT	Post-16 only
Progression to HE	Primary outcome	After endpoint	HESA	TASO	Post-16 only
Progression to host university	Secondary outcome	After endpoint	HESA	TASO	Post-16 only
Application to HE	Exploratory outcome	After endpoint	Self-report survey 2	HEPs	Post-16 only
Likelihood of going to HE	Exploratory outcome	Baseline After endpoint	Self-report survey 1	HEPs	Both
Likelihood of progressing to academic study post-16	Exploratory outcome	Baseline After endpoint	Self-report survey 1	HEPs	Pre-16 only
Desirability of HE	Exploratory outcome	Baseline After endpoint	Self-report survey 1	HEPs	Both
Self-efficacy relating to HE	Exploratory outcome	Baseline	Self-report survey 1	HEPs	Both

		After endpoint			
Compatibility of HE with social identity	Exploratory outcome	Baseline After endpoint	Self-report survey 1	HEPs	Both
Perception of practical barriers to HE	Exploratory outcome	Baseline After endpoint	Self-report survey 1	HEPs	Both
Summer school attendance data	Complier Average Causal Effect (CACE) analysis	After endpoint	HEP admin data	HEPs	Both
Participation in other outreach activities during the trial period	Contextualising estimated effects of intervention	After endpoint	HEAT	HEPs	Both
Cost data					
Intervention cost estimates	Estimating cost per participant	Endpoint	Interview	BIT	NA
HEP staff data					
First name	Project management	Baseline	TASO admin data	TASO	NA
Last name	Project management	Baseline	TASO admin data	TASO	NA
Work email address	Project management	Baseline	TASO admin data	TASO	NA
Work telephone number	Project management	Baseline	TASO admin data	TASO	NA

Table 4: Trial data

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 5. (See the outcomes section above for the full questions and methods of aggregation).

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 a 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."

Table 5: Validity and reliability of survey items

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

- A data sharing specification, including details on the variables and their required coding, will be given to all providers to ensure consistent and reliable data collection across all HEPs.
- 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.

Timeframe	Action
January 2022 - June 2022	• Complete trial protocol • Set up data sharing processes and agreements • Recruit participants and assign to treatment or control group • Collect baseline registration data
June 2022 - March 2023	• Deliver summer schools • Collect outcome data through survey 1 (knowledge and attitudes) • Analyse data and complete interim report 1

January 2023 - April 2023	• Collect outcome data through survey 2 (HE applications) • Analyse data and complete interim report 2
June 2023 - December 2023	• Collect NPD data (TBC)
January 2024 - November 2024	• Collect HESA outcome data • Analyse data and complete final report

Table 6: Trial timeline

11. Power calculations

11.1. Summary of findings

The power calculations were conducted for the primary outcome. If neither of the two summer schools with post-16 students drop out, and 76% of the control group progress to HE (our best guess), then we estimate that the MDES ranges from 5.8pp to 6.5pp.

If no summer schools drop out, and 57% of the control group progress to HE (a more conservative guess), then we estimate that the MDES ranges from 7.0pp to 7.9pp.

If Leicester drops out, then we estimate that the MDES ranges from 7.6pp to 11.0pp.

If Leeds drops out, then we estimate that the MDES ranges from 10.3pp to 12.9pp.

While the existing evidence is very limited, even the best case scenario (an MDES of 5.8pp) seems high. The one quasi-experimental study that most closely matches this trial in terms of intervention and sample, estimates that the maximum effect of summer schools of this type on progression to HE is likely to be 2.3pp (Hoare & Mann, 2011, p.79). Interim findings from phase 1 of this present trial, that evaluated the proximal impact of online summer schools, found a null effect on self-reported likelihood of progressing to HE (Taylor & Shrestha, 2022).

11.2. Introduction

Power calculations have been conducted for the primary outcome only (i.e. progression to HE). This means that only the 13 post-16 summer schools (run by Leeds and Leicester) 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

Power calculations are based on this information (see Appendix IV) and a series of assumptions based on the data from phase 1 (see Table 7 below). The total post-16 applicant pool is estimated to be 2,190. There are approximately 625 places available to participate in the post-16 summer schools, including those places that are guaranteed to applicants meeting the relevant criteria.

11.3. Baseline progression to HE

To estimate the minimum effect size that is detectable with a sample of 2,190, we need to estimate 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.

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. did not 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.

11.4. Summary of assumptions

Table 7 summarises the assumptions used to estimate the MDES under a range of scenarios.

Assumptions	
Power	0.8
Significance level	0.05
Total N of estimated applicants	2,190
Total N of estimated places	625
Primary outcome	Progression to HE (a binary outcome)
Primary outcome baseline	Based on the literature we estimate that the rate of progression to HE in the control group will be between 76% and 39%. The estimate of 76% is closer to what we expect in this trial (see above for more info). We have included additional calculations based on a 57% baseline (the midpoint between 76% and 39%).
Quotas/stratification	We are assuming that no quotas can be met due to the stratification rules (as observed in the first trial). (The simulations used to calculate the MDESs for the first trial - that did take quotas into account - produced results that were very similar to the non-simulation approach anyway).
Unequal allocation	The size of the treatment group is capped by the places available to applicants at the summer schools, so

	randomisation will be conducted using an unequal allocation of participants to treatment and control arms.
HEP drop-outs	As observed in phase 1, whole HEPs can drop out between the current stage and randomisation. When this occurs, both the number of applicants and the number of places available for the treatment group are reduced. In phase 2 there are only two post-16 HEPs (Leeds and Leicester), so we have included two HEP drop-out scenarios in our estimates: one where Leicester drops out and one where Leeds drops out.[8]
Other sources of data loss	There are three other types of data loss that we expect to occur between the estimated figures provided by the HEPs and the analytic sample. **1. Opt-outs** - applicants who ask not to be part of the data collection and analysis. (We expect this figure to be lower than in phase 1, which required opt-in consent). **2. Overestimation of applications** - in phase 1, on average, HEPs overestimated the number of applications they would receive. **3. Missing data** - outcome data that is missing from the final dataset. (We expect this number to be fairly low because the outcome data will come from a government administrative dataset. Matching should also be high-quality as we have HEAT IDs which identify individuals in the HESA dataset, as well as a large number of identifying covariates). In phase 1, the first two types of attrition above resulted in a loss of 37.7% of participants. In our calculations we estimated a range of attrition across these three categories including 0%, 20%, and 40%.
Duplicate applicants	In phase 1, duplicated applicants occurred at a rate of 3.0%, so we have assumed the same level here.
Guaranteed places	5.5% of applicants in phase 1 were guaranteed places on the summer schools so we have assumed the same level here.

[8] If Leicester withdraws we would lose 1,010 estimated applicants and 180 summer school places, if Leeds withdraws we would lose 1,180 estimated applicants and 445 summer school places.

	The proportion of guaranteed places affects both the overall N and the N available to the treatment arm.

Table 7: Assumptions for power calculations

11.5. Findings

Based on these assumptions, our estimates indicate that the MDES may range from 5.8pp to 12.9pp. 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 57% to ~70%. While the existing evidence is very limited, even the best case scenario (an MDES of 5.8pp) seems high. The one quasi-experimental study that most closely matches this trial in terms of intervention and sample, estimates that the maximum effect of summer schools of this type on progression to HE is likely to be 2.3pp (Hoare & Mann, 2011, p.79). The inclusion of covariates in our analysis may increase the precision of the estimate effect, and thus reduce the MDES, but it is unlikely to reduce it to 2.3pp. Interim findings from phase 1 of this present trial, that evaluated the proximal impact of online summer schools, found a null effect on self-reported likelihood of progressing to HE (Taylor & Shrestha, 2022).

Table 8 shows the detailed findings from the power calculations. To show how the numbers in this table were calculated, a worked example is given for row 4 in the table in Appendix V.

Alpha							**0.05%**			
Power							**80%**			
Baseline		**HEP drop-outs**		**Other attrition**						
76%	**57%**	**Leeds**	**Leicester**	**0%**	**20%**	**40%**	**Sample size**	**n of treated group***	**n of control group**	**MDES pp**
✓				✓			2008	509	1499	5.8
✓					✓		1606	532	1074	6.0
✓						✓	1204	555	649	6.5

✓			✓		✓		866	395	471	7.6
✓			✓			✓	649	407	242	9.0
✓		✓			✓		741	137	604	10.3
✓		✓				✓	556	148	408	10.4
	✓			✓			2008	509	1499	7.0
	✓				✓		1606	532	1074	7.2
	✓					✓	1204	555	649	7.9
	✓		✓		✓		866	395	471	9.3
	✓		✓			✓	649	407	242	11.0
	✓	✓			✓		741	137	604	12.7
	✓	✓				✓	556	148	408	12.9

Table 8: Power calculations

Notes: pp = percentage points; MDES = minimum detectable effect size. Sample size estimate for the 0% attrition group includes data loss due to duplicates and guaranteed places, this is the case for all scenarios. *For each given 'Baseline' and 'HEP drop-outs scenario' as the attrition rate increases so does 'n of treated group'. This is an effect of the artificial cap on the size of the treatment group (due to the limit on the places available at each summer school) and the applicants with guaranteed places. As attrition increases the absolute number of applicants with guaranteed places decreases. Thus, fewer of the available places are allocated to applicants with a guaranteed place leaving more space for eligible participants to be allocated into the treatment group.

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).[9]

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,

[9] 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.

- Y_i is a binary indicator of whether the individual goes on to study at the HEP that delivers the summer school applied to[10] (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
- 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

[10] 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'.

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 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 5, using 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 6, using 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 will select a summer school 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.

12.10. Programme differentiation

It is possible that trial participants (in both the treatment and control groups) will participate in other activities during the intervention period that could have an effect on the outcomes targeted by the intervention. Knowing whether or not this is the case and how this differs by treatment assignment adds helpful context to the estimated treatment effects. To add this context, TASO will collect data from HEAT to establish whether they have participated in other outreach activities during the trial period (see Appendix I). BIT will conduct the following analysis on this survey data:

1. report descriptive statistics to show the number and percentage of participants that have participated in additional outreach activities during the intervention period (broken down by activity type - see Appendix I for full list of activity types - and assignment); and

2. conduct balance checks (using a normalised differences approach) to see whether there are substantial differences on this point between the treatment and control groups.

We will not re-estimate treatment effects based on the outcome of this analysis, because conditioning on post-treatment variables can introduce bias (Montgomery et al. 2018).

12.11. 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).

12.12. 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). Table 9 shows when we use this procedure. The procedure is applied separately for primary and secondary outcomes, but does not apply to exploratory outcomes.

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+				

Table 9: When to correct for multiple comparisons

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.

12.13. Compliance

12.13.1. Introduction

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. Table 10 shows the definitions of compliance given by the providers. There is variation in these estimates because there is variation in what constitutes a summer school in each case.

Because these compliance definitions vary substantially, we will run two CACE analyses. The first analysis will use the compliance definitions given by the providers. The second analysis will define compliance in the treatment group as 100% attendance. The second analysis will act as a robustness check to the first, and will help us to understand the effect of dosage on the primary outcome.

Summer school	Definition of compliance with treatment	Length of summer school
Gloucestershire	≥ 3 days attendance	4 days on campus (3 nights)
Kent pre-16	≥ 3 days attendance	4 days on campus (3 nights)
Kent post-16	≥ 3 days attendance	4 days on campus (3 nights)
NTU	≥ 2 days attendance	2 days on campus (1 night)
Leeds Biosciences	≥ 2 days attendance online, OR ≥ 1 day attendance on campus	3 days online with 1 on-campus day
Leeds Dentistry	≥ 1 day attendance	2 days on campus (1 night)
Leeds Medicine	≥ 1 day attendance	2 days on campus (1 night)
Leeds Psychology	≥ 1 day attendance at on-campus day, OR ≥ 2 days attendance online	2 days online with 1 on-campus day
Leeds Social Sciences	≥ 2 days attendance	2 days on campus (1 night)

Leicester Arts	≥ 4 days attendance	4 days on campus (3 nights)
Leicester Business	≥ 4 days attendance	4 days on campus (3 nights)
Leicester Law	≥ 4 days attendance	4 days on campus (3 nights)
Leicester Medicine	≥ 4 days attendance	4 days on campus (3 nights)

Table 10: Compliance definitions by summer school

12.13.2. CACE specifications

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

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 estimations 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).

12.14. Missing data

All analysis described above will be conducted on complete cases only. Where covariates are missing, we will run two robustness checks on all analyses. First, we will create a new variable to indicate missingness and use this to re-estimate the effects. Second, we will re-run all analyses without covariates and report the unadjusted estimates.

Further missing data analysis will 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.

12.15. 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 and the cost per additional participant progressing to HE. 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 design and data collection will be carried out by BIT. The questionnaire will be conducted through structured interviews with a sample of five HEPs (two 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 NTU Schools of Business, Law and Social Sciences Research Ethics Committee (REC), reference 2021/378. See separate document for details.

15. Risks

Part of evaluation	Risk	Mitigation strategy	Risk owner
Participant recruitment	Leeds ethics committee is conducting an additional ethics review. This is delaying the opening of applications to Leeds summer schools which could reduce applicant numbers.	TASO and Leeds to prioritise the completion of the Leeds ethics review. TASO to maintain regular communication with HEPs to address any issues caused by delays such as these.	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
Partner commitment	HEPs have internal struggles related to resourcing and applicant recruitment which may cause them to pull out of the research, or for the project to be terminated.	TASO has supported HEPs as much as possible to ensure they can commit to the project. HEPs will notify TASO at an early stage if applicant numbers are looking lower than expected so that efforts can be put in to aid recruitment.	TASO, HEPs

Table 11: Risk analysis

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

Montgomery, J.M., Nyhan, B. and Torres, M., 2018. How conditioning on post-treatment variables can ruin your experiment and what to do about it. American Journal of Political Science, 62(3), pp.760-775.

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

Taylor, P., & Shrestha, P. (2022). Analysis report 1: TASO Summer Schools Evaluation. The Behavioural Insights Team.

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 2023**)

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

Gloucestershire

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

Kent pre-16 and post-16 summer schools

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

NTU

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

Leeds

The Leeds summer schools vary by specific subject summer school as outlined below. Activities delivered across the summer schools include:

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

Leeds Biosciences

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

Leeds Dentistry

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

Leeds Medicine

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

Leeds Psychology

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

Leeds Social Sciences

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

Leicester

The Leicester summer schools vary by specific subject summer school as outlined below. All take place on campus over four days, with three nights spent in university accommodation.

Leicester Arts

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

Leicester Business

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

Leicester Law

Students on the Law stream have sessions on a crime scene, why study law, a campus tour and on homicide and interviewing clients. Further workshops include interviewing

and advising, plea in mitigation, presenting your plea in mitigation, impact of imprisonment and Q&A. A total of 9 subject specific sessions.

Leicester Medicine

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

Leicester STEM

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

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

HEP	Pre or post-16	No. of places	Est. applicant numbers	Delivery mode	Subject specific?	Guaranteed places	Eligibility criteria
Kent	Pre-16	60	200	Residential	No	Care-leavers / care-experienced	Attend partner school/college, one WP criteria - first-gen, BAME, POLAR 4 Q1, IMD Q1, disability
NTU	Pre-16	80	200	Residential	No	No	One of the following - IMD Q1, POLAR4 Q1, FSM, disability, care-experienced, young carer
Gloucestershire	Pre-16	60	150	Residential	No	No	One of the following - first-gen, FSM, IMD Q1, POLAR4 Q1, disability or BAME
TOTAL PRE-16:		**200**	**550**				
Leeds Dentistry	Post-16	40	140	Residential	NA	Yes (but characteristics not yet specified)	One of the following - low participation neighbourhood, FSM, low income (£25,000 per annum or less), care-experienced or studies disrupted (studies disrupted by circumstances in their personal, social or domestic lives).
Leeds Healthcare	Post-16	40	100	Residential	NA		
Leeds Medicine	Post-16	100	400	Residential	NA		
Leeds Psychology	Post-16	60	100	Residential	NA		
Leeds Social Sciences	Post-16	40	115	Residential	NA		
Leeds Languages	Post-16	65	150	Residential	NA		
Leeds Biosciences	Post-16	50	100	Offering online and face-to-face	NA		

Leicester Medicine	Post-16	30	530	Residential	NA	Care-leavers and low participation areas - approx. 10	One of the following - care-experienced, young carer, disability, estranged, FSM, GRT, refugee or asylum seeker, first-gen, POLAR4 Q1
Leicester Law	Post-16	45	200	Residential	NA	Care-leavers	
Leicester Arts	Post-16	30	130	Residential	NA	Care-leavers	
Leicester STEM	Post-16	45	200	Residential	NA	Care-leavers	
Leicester Business	Post-16	30	80	Residential	NA	Care-leavers	
Kent	Post-16	40	80	Residential	NA	Care-experienced	Attend partner school/college, one WP criteria - first-gen, BAME, POLAR 4 Q1, IMD Q1, disability
TOTAL POST-16:		**615**	**2325**				

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

Note: The figures in this table are slightly different to those in the body of the report. This is because they are based on updated estimates from the HEPs, after power calculations were conducted.

21. Appendix V: Power calculations worked example

For extra clarity, below we present a worked example of a likely scenario, demonstrating how the different types of data loss we have identified affect both the overall sample and the places available for the treatment group. This example refers to row 4 in Table 8.

1. We consider the total number of estimated applicants is **N(total) = 2,190.**
2. We account for HEP level drop-outs.
    a. This calculation assumes that Leicester will drop out. This results in the total number of estimated applicants dropping to **N(total) = 1,180**, as we lose n = 1,010 applicants. Leicester leaving the trial also reduces the number of places available to N(treatment) = 445, as we lose n = 180 places.
3. We account for data loss from opt-outs, overestimation of applicants, and missing data.
    a. In phase 1, we observed drop-outs of the first two types at 37.7%, so we estimated a range including 0%, 20%, and 40%. In this example we use the 20% estimate. We therefore drop n = 236 pupils leaving **N(total) = 944**.
4. We account for the expected number of duplicates.
    a. Based on the duplicate rate from phase 1, after opt-outs, overestimation of applicants, and missing data attrition are considered, we drop 3% (n = 28) of total N, leaving **N(total) = 916**.
5. We account for the expected number of guaranteed places.
    a. Based on the guaranteed places rate from phase 1, after duplicates are considered, we drop 5.5% (n = 50) of the total N, leaving **N(total) = 866**. These 50 guaranteed places are also dropped from the places available, N(treatment) = 395.
6. Therefore, 395 pupils would be assigned to the treatment group and 471 pupils would be assigned to the control group.
7. The MDES for this example is 7.6 pp.