

Technical appendix 2

An investigation into the relationship between outreach participation and progression to higher education: analysis based on the HEAT aggregate tracking dataset

Eliza Kozman, Deputy Director of Research at the Centre for Transforming Access and Student Outcomes (TASO)

March 2021

Contents

1. Introduction.....	2
2. Description of data.....	3
2.1 Demographic data.....	3
2.2 Outcome data.....	6
2.3 Activity data.....	8
2.4 Limitations of the dataset.....	10
3. Methods.....	11
4. Results.....	12
4.1 Overall HE progression – demographic factors and HEAT intensity measure.....	12
4.2 Overall HE progression – binary measure of participation in activities versus HEAT intensity ..	18
4.3 Progression to top-third HEPs – demographic factors and intensity measures	24
4.4 Progression to top-third HEPs – binary measure of participation in activities versus HEAT intensity.....	29
4.1 Participation in activities.....	35
Annex A: Missingness analysis.....	36
Annex B: Logistic regression tables.....	37
Annex C: Top-third HEPs	49

1. Introduction

- This report is a technical appendix to our summary report: An investigation into the relationship between outreach participation and Key Stage 4 (KS4) attainment/HE progression.¹ Please refer to the summary report for a discussion of the results of this analysis.
- In this report we use regression analysis to explore the extent to which participation in outreach recorded on the HEAT database is associated with:
 - Progression to higher education (HE)
 - Progression to 'top-third' HE providers (i.e. those with the highest entry requirements in terms of entry qualifications)
- It is important to note that this analysis cannot provide *casual* evidence on the efficacy of outreach because:
 - We cannot capture differences between outreach participants in terms of factors such as individual motivation and school/parental support.
 - Individuals who are more interested in HE and have more school/parental support may be more likely to participate in a greater number of activities or different activities.
 - These factors are also strongly correlated with attainment and HE progression.
 - In other words, there is a risk of 'selection bias', where the groups we examine (i.e., those who do and do not take part in outreach) may have been very different to begin with, regardless of those activities.²
 - Therefore, where we find that participation in outreach is associated with progression to attainment/HE progression, it is not possible to attribute this to the activities recorded in HEAT because we cannot rule out that other differences are driving the pattern we observe.
- However, there are a number of factors which are strongly correlated with attainment and HE progression that *are* present in the HEAT dataset. These include prior attainment and proxies for socio-economic background such as Free School Meals (FSM) eligibility, Income Deprivation Affecting Children Index (IDACI) and whether the individual is first in family to attend HE. We include such variables in our analysis in an attempt to take into account some of the measurable differences between individuals who take part in different activities.
- Therefore, although not capable of providing causal evidence, this descriptive analysis is able to provide high level trends which can be used to inform future causal studies. According to the [Office for Students \(OfS\) Standards of Evidence](#), we categorise this work as Type 2 evidence.

¹ See accompanying summary report: <https://taso.org.uk/wp-content/uploads/relationship-outreach-attainment-progression.pdf>

² For more information on selection bias please see [this explanation on the Institute for Work and Health website](#).

1. Description of data

2.1 Demographic data

- HEAT data were matched to Higher Education Statistics Agency (HESA) records to track whether outreach participants recorded on the database progressed into HE.
- HESA data was provided for 208,550 individuals who exist on the HEAT database. 182,899 of these individuals (88%) were present in HEAT activity data.³
- The analysis is restricted to data for participants who were 'HE ready' between 2007/08-2017/18, providing a ten-year period during which participants had the opportunity to progress to HE.⁴ Therefore, we drop 4,524 individuals who do not meet this criteria.
- We also drop 2,061 individuals who went to school in Wales because administrative demographic data was not available for these students (these individuals comprise just over one percent of the individuals in the dataset). Therefore, all the individuals in our analysis went to school in England.
- This process leaves us with 165,448 rows where each row relates to an individual on the HEAT database.
- For this analysis, we use demographic factors to take into account observable differences between individuals. As the dependent variable is HE attendance, we use factors which are predictive of entering HE.⁵ We use a geographic measure of disadvantage as well as FSM-eligibility based on research which suggests this is the most appropriate way of accounting for participant background.⁶ IDACI was selected as the geographic marker as no indicators stand out as considerably better in existing analysis and it applies directly to child poverty. Other proxies for disadvantage are not included as they are likely to be highly correlated with each other and offer little additional information.

³ Where individuals are missing in the HEAT data, it is likely to be because they have been removed from the HEAT database by HEAT members since the HESA data was requested.

⁴ An individual's 'HE ready' year is the academic year student is aged 18 and therefore typically ready to progress to HE.

⁵ For a discussion of how such factors influence the likelihood of an individual entering HE see, for example, Crawford, C., & Greaves, E. (2015). Socio-economic, ethnic and gender differences in HE participation.

⁶ See for example Ilie, S., Sutherland, A. and Vignoles, A. 2017. Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation. *British Educational Research Journal* 43(2): 253-274. doi:10.1002/berj.3260

- A description of the data is given in Table 1. This table shows that there is a substantial amount of data missing for some of the variables.

Table 1. Description of missing data in overall dataset

Variable	Number of rows data missing	Missing rows as percentage of total
HE ready year	0	0%
Gender	28,870	17%
FSM eligibility	92,628	56%
IDACI quintile	6,820	4%
Ethnic group	87,324	53%
School region	0	0%
First generation HE (self-report)	74,587	45%
First engaged in activity post-16	4,104	2%
Average Key Stage 4 point score	98,964	60%

- Further analysis shows that there is a degree of overlap in the missing data.⁷ Figure 1 shows that the most commonly missing data is average Key Stage 4 score (missing for almost 100,000 students, or 60% of the rows in our dataset) and the first column shows that, of these individuals, over 29,000 are also missing FSM eligibility information. The next column shows that around 28,000 individuals (17% of the total) are missing Key Stage 4 scores, FSM eligibility, ethnic group and information on whether they would be first in their family to enter HE.
- Before starting our analysis, we need to consider how best to handle the missing data. In this situation, there are probably underlying reasons why data is missing for some individuals and not others, meaning that there are likely to be systematic differences between these students. In technical terms, we say that this data is ‘missing not at random’.
- Therefore, we must be careful when discarding data to conduct our analysis. With this in mind, we conduct our analysis using two different datasets, as described in Table 2:
 - One dataset is restricted to rows where we only have complete data for individuals.
 - One dataset retains all rows but we include dummy terms for missing data.
- We conduct the same analysis on each dataset and then compare our results to formulate insights about the relationship between outreach activity on our outcomes of interest.

⁷ Using the `naniar` function in R: Tierney, N., Cook, D., McBain, M., & Fay, C. (2019). `naniar`: Data Structures, Summaries, and Visualisations for Missing Data. R package version 0.4. 2.

Figure 1. Analysis of overlap of missing data

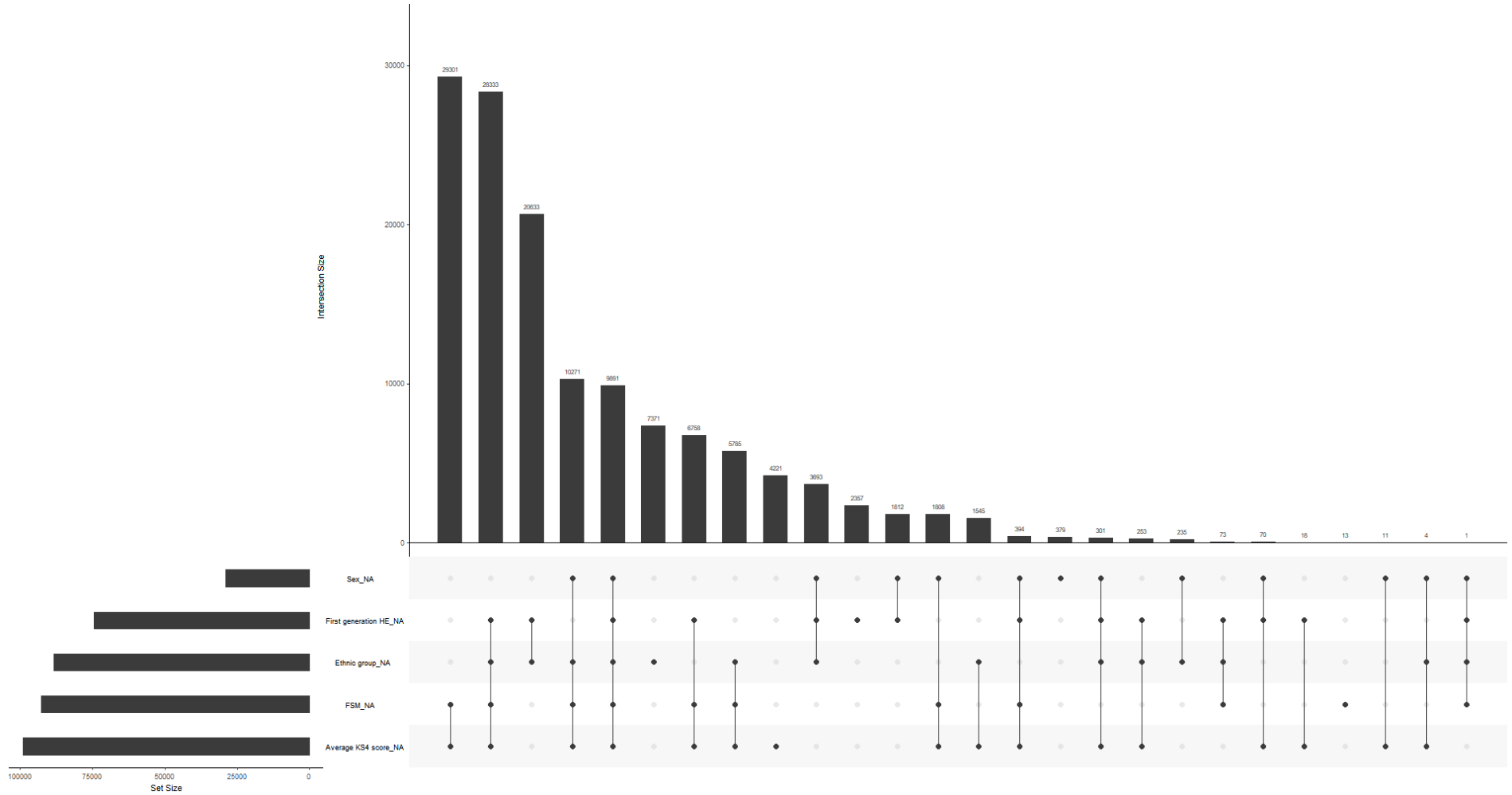


Table 2. Description of datasets used

Dataset	Number of students	Description
Restricted dataset	N= 28,050	<p>In the first instance, we restrict the dataset to rows where we have full data for individuals (the 'restricted dataset'). Because the missing data does not totally overlap, this action results in restricting our dataset to 28,050 individuals. This option gives us rich data but for only a small portion (around 17%) of the unrestricted dataset. Because data is unlikely to be missing at random, this sample is not likely to be representative of the broader population of HEAT users.</p> <p>Note: if we relax the requirement for full data on the 'first generation HE' variable, then we could increase the size of the restricted dataset to N=30, 341 (around 18% of the unrestricted dataset). However, given that the first in family measure is a rich measure of the individual's background, we choose to retain it and sacrifice a small portion of the sample instead.</p>
Unrestricted dataset	N= 165,448	<p>In the second instance, we retain all the rows in the unrestricted dataset. So we can keep as many individuals as possible in our analysis, we re-code every missing value as 'unknown' and treat this as a new category for each variable (e.g. FSM eligibility is then categorised as either 'FSM-eligible', 'not FSM-eligible' or 'unknown').</p> <p>For the missing continuous data, we create additional variables called "KS4 missing" and "IDAC1 missing" which indicate where these variables are missing and we replace the missing values with the mean so we can retain all the rows of data in our modelling.</p> <p>This approach allows us to keep all the rows but results in a lot of data being categorised as 'unknown' which introduces challenges when interpreting the analysis.</p>

2.2 Outcome data

- For this analysis we are interested in:
 - Whether an individual progressed to HE
 - Whether an individual progressed to a top-third HE provider (HEP)
- The mean of each outcome is given in Table 3.
- The fact that progression to a top-third HEP differs between datasets suggests that we were right assume that the missing data is not missing completely at random (i.e. data may be missing for certain types of individual or certain types of HEP).
- Indeed, when we look at the distribution of demographic characteristics also shown in Table 3, we can see that the datasets differ substantially in their composition:
 - The restricted dataset contains no individuals whose HE ready year was 2016-17 or 2017-18, as administrative demographic data was not available for these students.
 - The distribution of school regions differs substantially and this could be partly due to the changing membership of HEAT over the years (as the restricted dataset does not include learners who were HE-ready in 2016-17 or 2017-18) or due to different ways of recording data between HEPs based in different areas of the country.
 - The proportion of learners who first engaged with outreach post-16 is much lower for the restricted dataset, meaning that HEAT is more likely to hold complete demographic data for learners who engage pre-16.
 - The distribution of other demographic variables is comparable between datasets.
- Additional analysis presented in Table 23 in Annex A confirms that students who are missing from the restricted dataset are significantly different from the wider pool: on

average they have higher attainment, they are likely to be older, less likely to be female and more likely to have engaged post-16.

Table 3. Mean outcomes in data and distribution of demographic characteristics

	Restricted dataset	Unrestricted dataset ⁸
Sample	N=28,050	N= 165,448
Progression to HE	53.0%	57.0%
Progression to top-third HEPs (of those who attended HE)	26.1%	43.7%
HE ready year	<1% 2007; <1% 2008; 7% 2009; 11% 2010; 14% 2011; 14% 2012; 17% 2013; 17% 2014; 20% 2016; 0% 2016-17; 0% 2017-18	2% 2007; 3% 2008; 3% 2009; 4% 2010; 5% 2011; 6% 2012; 9% 2013; 9% 2014; 14% 2016; 19% 2017; 26% 2018
Gender	59% female; 41% male or other	58% female; 42% male or other
FSM eligibility	13% FSM; 87% non-FSM	12% FSM; 88% non-FSM
IDACI quintile	22% quintile 1; 27% quintile 2; 21% quintile 3; 17% quintile 4; 13% quintile 5	21% quintile 1; 24% quintile 2; 21% quintile 3; 18% quintile 4; 16% quintile 5
Ethnic group	11% Asian; 5% Black; 4% Mixed 80% White <1% Other	13% Asian; 7% Black; 4% Mixed 75% White 1% Other
School region	20% East Midlands; 2% East of England; 5% London; <1% North East; <1% North West; 67% South East; 2% South West; 2% West Midlands; 2% Yorkshire and the Humber	8% East Midlands; 5% East of England; 12% London; 4% North East; 4% North West; 44% South East; 8% South West; 4% West Midlands; 10% Yorkshire and the Humber
First generation HE (self-report)	75% first generation HE; 25% not first generation HE	73% first generation HE; 27% not first generation HE
First engaged in activity post-16	47% engaged post-16; 53% engaged pre-16	74% engaged post-16; 26% engaged pre-16
Average Key Stage point score	Mean = 42.5; sd=6.7	Mean = 43.1; sd=7.2

⁸ Distributions given for rows where we have data on a particular variable.

2.3 Activity data

- For this analysis it is necessary to include measures of the intensity of activity experienced by individuals. We measure intensity in two ways, as outlined in Table 4.
 - Across both our datasets, between 50% and 60% of participants have engaged in 'intensive' HEAT outreach, defined using a simple categorisation of the HEAT activities. Individuals in the restricted dataset were less likely to have participated in an 'intensive' package of support than those in the unrestricted dataset.
 - On average, participants have taken place in two activities with a standard deviation of around three.

Table 4. Summary of intensity measures

Intensity measure	Description	Restricted data set	Unrestricted dataset (without attainment) and imputed dataset
A binary HEAT intensity marker	<p>An 'intensive' package of activities is defined as:</p> <ul style="list-style-type: none"> - One or more summer schools - One or more HE insight events - One or mentoring interactions - One or more projects - Two or more skills & attainment activities - Two or more campus visits - One or more skills & attainment activities and one or more campus visits - Three or more HE information talks and one or more skills & attainment activities - Three or more HE information talks and one or more visits <p>If a student has taken part in intensive activity this is coded as 1.</p> <p>All other combinations of activity are defined as 'less intensive' and coded as 0.</p>	Mean=0.53	Mean=0.58
Number of activities participated in (activity count)	A count of total activities an individual has participated in.	Mean=2.36 Sd=2.61 Min=1 Max=43	Mean=2.28 Sd=2.93 Min=1 Max=90

- A summary of participation in separate activity types is given in Table 5.
- Here we see that the activity profile for individuals in the restricted dataset is different from that of individuals in the unrestricted dataset. For example, the average number of summer schools attended is lower, and average number of mentoring interactions is higher within the restricted dataset.
- By looking at the range of the counts we can also see that restricting the datasets sometimes excludes students with very high activity counts. Latter analysis finds that these cases are not 'influential' i.e. they do not bias our regression analysis, so we retain them in the data.

Table 5. Summary of activity counts

Activity	Restricted data set	Unrestricted dataset (without attainment) and imputed dataset
Summer schools	Mean=0.14 Sd=0.40 Min=0 Max=5	Mean=0.32 Sd=0.82 Min=0 Max=10
Mentoring	Mean=0.50 Sd=1.89 Min=0 Max=34	Mean=0.20 Sd=1.19 Min=0 Max=60
Projects	Mean=0.12 Sd=0.50 Min=0 Max=9	Mean=0.07 Sd=0.50 Min=0 Max=22
Campus visit	Mean=0.27 Sd=0.59 Min=0 Max=23	Mean=0.27 Sd=0.83 Min=0 Max=23
Subject insight event	Mean=0.31 Sd=0.61 Min=0 Max=6	Mean=0.35 Sd=0.86 Min=0 Max=20
Skills attainment activity	Mean=0.37 Sd=0.79 Min=0 Max=21	Mean=0.55 Sd=1.37 Min=0 Max=56
HE information session	Mean=0.48 Sd=0.96 Min=0 Max=23	Mean=0.37 Sd=1.01 Min=0 Max=34
Exhibition	Mean=0.04 Sd=0.19 Min=0 Max=2	Mean=0.03 Sd=0.21 Min=0 Max=6

- It is important to keep the characteristics of the different datasets in mind throughout the remainder of the analysis. Specifically, it seems that the restricted dataset is focused on a lower-achieving group of students who tend to have engaged pre-16. It is possible that this group is a subset of students who were contacted via whole-school engagement and this is why more complete data exists for these learners (as data is likely to be more complete where provided by schools than individuals).
- Nonetheless, we retain our dual analysis strategy for transparency and robustness.

2.4 Limitations of the dataset

- The first limitation relates to the coverage of the HEAT aggregate tracking dataset. The dataset was compiled by 92 different outreach providers as described in Appendix 1. An up-to-date list of HEAT's member organisations is available on the [HEAT website](#). Although this represents a large proportion of outreach providers, the dataset does not include all organisations providing outreach. There remain several HEIs and two Uni Connect consortia that use their own regional tracking databases that are currently separate from HEAT. Furthermore, there are third sector and private providers of outreach that do not record their data on a central tracker database. Therefore, this dataset can be considered as a sample of outreach participants, rather than a complete dataset of all outreach delivered nationally.
- Furthermore, HEAT's member organisations are free to use the HEAT database to record the student and activity data according to their individual organisation's needs. Thus, even within the membership, there may be gaps in data collection and recording. The extent of the gaps in data are currently unknown, but there are a number of imperatives such as Data Protection regulations and the OfS requirements to evaluate that work to encourage organisations to record their data securely on a tracking system such as HEAT's. The large sample size provides further evidence that the sample of outreach participants we do have represents a large proportion of all outreach work delivered nationally.
- A second limitation relates to the ability of the data to demonstrate 'what works' in terms of outreach influencing HE progression. In spite of rich data collected on activities, the data are observational (i.e. this dataset did not come from a controlled experiment but rather from 'real-life' processes). Measures are taken to control for observed variables known to influence HE progression. However, simply comparing outcomes for students who attended different types of activity cannot isolate the effect of the activity from other unknown factors (such as students' personal motivation or their input from family, friends and teachers). As a result, statistical *associations* are shown between activities and differential participant attainment, providing a strong Type 2 standard of evidence according to the [Office for Students' guidance](#). The research does not claim to show robust causal effects.
- A third limitation relates to the diversity in the packages of activities in which students have participated. Based on Activity Type and Activity Location variables alone, there are over 3,000 different combinations of packages of activity in which students have participated. If we include other variables – such as the sequence of activities, the contact hours and the year groups in which students participated – the number of combinations increases further. This makes isolating the possible effects of one type of activity difficult. Efforts are made to control for the package of activities in which students have participated. However, the very diverse nature of activities renders this task challenging.

2. Methods

- In this report we use regression analysis to analyse our outcomes (progression to HE and progression top-third HEPs) as a function of demographic variables and the level of HEAT activity a student has undertaken.
- All analysis was undertaken in R Statistical Software.⁹
- The output of this type of analysis is a regression table. We are primarily interested in:
 - Whether variables have a ‘statistically significant’ relationship with the outcome in question, meaning that there does appear to be a correlation or relationship which is unlikely to have occurred by chance. Significant results are denoted by asterisk symbols (*) in regression tables.
 - The size of the ‘coefficient’ associated with that variable, which tells us more about the relationship between the variable and the outcome.
- Because our outcomes are binary, the analysis is conducted using binary logistic regression, the results of which are presented in Annex B.¹⁰
- For ease of interpretation, we convert these tables into marginal effects which are presented in the body of this paper.¹¹ Marginal effects describe how a dependent variable (e.g. HE progression) changes when a specific independent variable changes, assuming we control for all other variables.
- In practice, this means that we can interpret the numbers in the ‘marginal effects’ tables as the percentage point change in the outcome which is associated with the variable in question (note: this does not apply to the binary logistic regression tables in the Annex).
- For all our analysis, KS4 attainment is standardised so it has a mean of 0 and a standard deviation of 1. Standardising a variable in this way makes it easier to interpret the outcome of our analysis so that:
 - The constant can be interpreted as the rate of HE progression for individuals with average KS4 attainment (taking into account all other variables).
 - The coefficient for KS4 attainment in our regression tables, can be interpreted as the increase in progression to HE associated with an increase in one standard deviation in KS4 attainment.
- The for the purpose of categorical variables in our analysis our reference categories are those with the lowest rate of HE progression, namely:
 - ‘White’ for ethnicity
 - ‘South East’ for school region
- Where regression table output is interpreted in percentage point terms, figures are rounded to the nearest whole number.
- In line with requirements from the DfE, we follow the HEAT strategy in published and released tabulations designed to prevent the disclosure of personal information about any individual. This strategy involves rounding all numbers to the nearest multiple of 5, rounding numbers less than 2.5 to 0 and suppressing percentages based on fewer than 22.5 individuals. However, the largescale nature of this data means that no numbers have required altering based on this strategy.

⁹ R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

¹⁰ For more information on regression tables, please see: <https://egap.org/resource/10-things-to-know-about-reading-a-regression-table/>

¹¹ We use Fernihough, A. Marginal Effects for Generalized Linear Models: The mfx Package for R. The marginal effects are calculated as a marginal effect at the mean i.e., all other covariates are kept at their mean value.

3. Results

- In the following section we present our results. Each section proceeds as follows:
 - We first present some **descriptive statistics** to see what patterns of progression exist in the raw data. However, we must remember that these statistics only provide a raw breakdown of the HE progression for students who do and do not meet the criteria listed. They do not take into account any other differences between students. For example, they do not take into account different levels of prior attainment for different groups – a factor which is highly predictive of HE progression.
 - Therefore, we then go on to present the marginal effects from our **regression analyses** which can take into account these multiple factors simultaneously and help us more fully understand the patterns we observe.
- Because we are handling a lot of missing data, the results from analysis of the restricted and unrestricted datasets do not always match up perfectly and judgement must be used to triangulate the findings. Therefore:
 - In the tables which summarise the results of the regression analysis, where the findings from the two datasets align, the ‘Agree?’ column is highlighted green; where there is divergence, the cell is shaded pink.
 - Where the results agree, we use the upper and lower bounds of the statistics provided by the analysis of both the restricted and unrestricted datasets.
 - If the results diverge too substantially, we do not report on the analysis as we cannot have confidence in our findings.
- It is important to remember that the sample consists of individuals who have engaged in some activity recorded on HEAT. Therefore, we cannot generalise the results to the whole population of learners, including those who have not engaged in any outreach.

4.1 Overall HE progression – demographic factors and HEAT intensity measure

- First, we examine the data to understand how HE progression varies depending on demographic factors and the HEAT intensity measure.
- The descriptive statistics are given in Table 6:
 - We see that the HEAT intensity measure is associated with higher rates of progression to HE.
 - In-line with existing literature, we also see that females, non-FSM eligible students, those from more advantaged areas (based on IDACI quintile) those who have a family history of HE attendance and BAME students have higher rates of progression to HE than other individuals.
 - We also note that individuals who first engaged with HEAT activity after the age of 16 have higher rates of HE progression than those who first engaged at a younger age.

Table 6. HE progression by demographic characteristics and HEAT intensity measure

Variable/group	Category	Restricted dataset			Unrestricted dataset		
		Mean HE progression	N	Difference in HE progression	Mean HE progression	N	Difference in HE progression ¹²
Intense HEAT activity	Yes	54.8%	14,999	3.6pp	62.6%	96,312	13.5 pp
	No	51.2%	13,279		49.1%	69,136	
Female	Yes	55.3%	16,672	5.3pp	56.9%	79,744	6.4 pp
	No	50.0%	11,606		50.5%	56,865	
FSM	Yes	46.4%	3,651	7.7pp	48.7%	8,638	8.9 pp
	No	54.1%	24,627		57.6%	64,182	
IDACI quintile	1	50.7%	6,364	-	53.5%	33,500	-
	2	49.7%	7,583		53.4%	37,654	
	3	53.1%	6,053		55.6%	32,520	
	4	55.1%	4,652		59.7%	29,187	
	5	61.8%	3,626		66.6%	25,767	
First engaged in activity post-16	Yes	63.2%	13,314	19.1pp	62.6%	118,357	19.4 pp
	No	44.1%	14,964		43.2%	42,987	
First in family	Yes	50.3%	21,108	11.3pp	53.5%	66,108	4.2 pp
	No	61.5%	7,170		57.7%	24,753	
Ethnicity	Asian	76.4%	3,113	-	67.6%	9,949	-
	Black	72.8%	1,303		65.5%	5,404	
	Mixed	57.4%	1,074		53.1%	3,354	
	Other	71.4%	227		61.5%	1,065	
	White	48.4%	22,561		47.6%	58,352	

¹² Differences presented as percentage points (pp)

- We next build a regression model including all demographic covariates (listed previously in Table 1) and our measures of intensity. The resulting regression tables are given in Table 24 and Table 25 in Annex B.
- For ease of interpretation, here we present the marginal effects (as described on page 11). Table 7 relates to the restricted dataset and Table 8 relates to the unrestricted dataset.

Table 7. Overall HE progression - marginal effects - function of demographic variables and intensity measures - restricted dataset

	Entered HE			
	(1)	(2)	(3)	(4)
Gender - female	0.045*** (0.006)	0.048*** (0.006)	-0.001 (0.007)	-0.003 (0.007)
FSM - eligible	-0.112*** (0.010)	-0.108*** (0.010)	-0.044*** (0.011)	-0.046*** (0.011)
Engaged post-16 - yes	0.067*** (0.007)	0.064*** (0.007)	0.113*** (0.008)	0.123*** (0.008)
IDACI quintile	0.047*** (0.003)	0.044*** (0.003)	0.022*** (0.003)	0.024*** (0.003)
First in family - yes		-0.088*** (0.007)	-0.076*** (0.008)	-0.079*** (0.008)
Av. KS4 score			0.300*** (0.005)	0.297*** (0.005)
Intense HEAT activity - yes				0.062*** (0.008)
Observations	28,278	28,278	28,278	28,278

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for ethnicity, HE ready year and school region

Table 8. Overall HE progression - marginal effects - function of demographic variables and intensity measure - unrestricted dataset

	Entered HE			
	(1)	(2)	(3)	(4)
Gender - female	0.060*** (0.003)	0.061*** (0.003)	0.041*** (0.003)	0.035*** (0.003)
Gender - unknown	0.119*** (0.004)	0.120*** (0.004)	0.114*** (0.004)	0.097*** (0.004)
FSM - eligible	-0.092*** (0.006)	-0.090*** (0.006)	-0.023*** (0.007)	-0.029*** (0.007)
FSM - unknown	-0.044*** (0.003)	-0.051*** (0.004)	0.102*** (0.009)	0.084*** (0.009)
Engaged post-16 - yes	0.117*** (0.003)	0.122*** (0.003)	0.126*** (0.003)	0.141*** (0.003)
Engaged post-16 - unknown	-0.084*** (0.009)	-0.080*** (0.009)	-0.095*** (0.009)	-0.084*** (0.009)
IDACI quintile	0.043*** (0.001)	0.043*** (0.001)	0.033*** (0.001)	0.035*** (0.001)
IDACI - unknown	-0.039*** (0.007)	-0.034*** (0.007)	-0.089*** (0.007)	-0.071*** (0.007)
First in family - yes		-0.052*** (0.004)	-0.042*** (0.004)	-0.055*** (0.004)
First in family - unknown		-0.093*** (0.005)	-0.098*** (0.005)	-0.100*** (0.005)
Av. KS4 score			0.192*** (0.002)	0.189*** (0.002)
KS4 unknown			-0.126*** (0.007)	-0.119*** (0.007)
Intense HEAT activity - yes				0.132*** (0.003)
Observations	165,448	165,448	165,448	165,448

Note:

*p<0.5; **p<0.01; ***p<0.001

Also controlling for ethnicity, HE ready year and school region

- An interpretation is given in Table 9 Table 1below.

Table 9. HE progression by demographic characteristics and HEAT intensity measure – summary of findings from regression analysis

Variable	Restricted dataset	Unrestricted dataset	Agree?
Gender	Progression for female students is 5pp higher, but this is not the case when we control for attainment, suggesting that grades drive this pattern in the data.	Progression for female students is 6pp higher and even after taking into account attainment, progression for female students remains 4pp higher. Progression is also 10pp higher for individuals whose gender is unknown.	Somewhat
FSM	Progression for FSM-eligible students is 10-11pp lower but this figure reduces to around 5pp when controlling for attainment and participation in intense HEAT activity.	Progression for FSM-eligible students is 9pp lower but this figure reduces to around 2-3pp when controlling for attainment and participation in intense HEAT activity. Progression is lower for individuals whose FSM status is unknown but this flips to a positive relationship when we control for attainment.	Yes
First engaged in activity post-16	Progression for those who first engaged post-16 is 6pp higher and this increases to 12pp when controlling for attainment and participation in intense HEAT activity.	Progression for those who first engaged post-16 is 14pp higher. Progression is 8pp lower for individuals whose date of first engagement with outreach is unknown.	Yes
IDACI	For every increase of one IDACI quintile, HE progression is 4pp higher. When we control for attainment, this reduces to around 2pp.	For every increase of one IDACI quintile, HE progression is 4pp higher. When we control for attainment, this reduces to around 3pp. Progression is 7pp lower for individuals for whom IDACI quintile is unknown.	Yes
First in family	Progression for individuals who would be first in their family to attend HE is 8pp lower even after controlling for attainment.	Progression for individuals who would be first in their family to attend HE is 6pp lower even after controlling for attainment. Progression is 10pp lower for individuals where we do not know if they would be first in family to attend HE.	Yes
Av. KS4 score	For every increase of one standard deviation in KS4 attainment, HE progression is 30pp higher.	For every increase of one standard deviation in KS4 attainment, HE progression is 19pp higher. Progression is 12pp lower for individuals for whom KS4 data is missing.	Yes
Intense HEAT activity	Individuals who take part in an intensive package of HEAT activities are 6pp more likely to progress to HE than those who do not. This is after taking account of all demographic variables and prior attainment.	Individuals who take part in an intensive package of HEAT activities are 13pp more likely to progress to HE than those who do not. This is after taking account of all demographic variables and prior attainment.	Somewhat

- In summary, the analysis finds, when controlling for other variables:
 - Female students are 5-6pp more likely to progress to HE, but part of this pattern is driven by patterns of attainment because females are likely to have better KS4 grades.
 - FSM-eligible students are around 10pp less likely to progress to HE, but this gap reduces to 2-5pp when controlling for attainment, suggesting part of the reason behind this pattern is that FSM-eligible students have lower KS4 grades.
 - Students who first engaged post-16 are 12-14pp more likely to progress to HE, even when we control for attainment. It is possible this finding reflects the different profile of students who engage pre- and post-16. Those who engage when they are older may do so under their own initiative and be more likely to be 'on track' for HE; therefore, higher rates of progression might be expected for this group.
 - For every increase of one IDACI quintile, HE progression is 4pp higher. When we control for attainment, this gap reduces to 2-3pp, suggesting grades play a part in driving this pattern.
 - Progression for individuals who would be first in their family to attend HE is 6-8pp lower. Controlling for prior attainment does not eliminate this gap, suggesting that grades cannot account for the lower progression observed for this group.
 - There is some discrepancy between the results from the two different datasets which is probably due to the underlying differences between the students included, but both datasets suggest that attainment plays a large role in driving HE progression: for every increase of one standard deviation in KS4 attainment, HE progression is 19-30pp higher, depending on which dataset you look at.
 - The HEAT intensity marker is associated with higher HE progression after taking account of all demographic variables and prior attainment. Using the restricted data, individuals who take part in an intensive package of HEAT activities are 6pp more likely to progress to HE than those who do not; in the unrestricted data, the figure is 13pp. Again, this probably stems from differences between the two groups; students in the restricted dataset have a profile which means they may be less likely to be interested in HE and this could be reflected in the figures we see here.

4.2 Overall HE progression – binary measure of participation in activities versus HEAT intensity

- Next we consider the role that participating in activities might have on HE progression.
- We collapse each activity count to either a 1 (student has taken part at least one count of this activity) or 0 (student has not taken part in this type of activity).
- We also include the HEAT intensity marker in our models.
- Therefore, in the following analysis:
 - **We are comparing students who have taken part in any activity versus no activity (for example, any mentoring session versus none)**, rather than trying to take into account the number of times a student has participated in an activity. This approach means we are making fewer assumptions about the consistency with which activity count data is recorded in the HEAT database.
 - **The comparison is with students who have participated in a similarly intense package of activities.**
- A description of the data is given in Table 10. These raw numbers suggest that, when we compare students who have taken part in a similarly intensive package of activities:
 - Summer schools and projects are associated with an increase in HE progression.
 - Mentoring, subject insight events and HE insight events are associated with a decrease in HE progression.
 - The picture is less clear for the other activities.

Table 10. HE progression by participation in activities (binary)

Activity	Intense HEAT activity	Restricted dataset				Difference in HE progression	Unrestricted dataset				
		Did not attend activity		Attended activity			Did not attend activity		Attended activity		Difference in HE progression
		Mean HE progression	N	Mean HE progression	N		Mean HE progression	N	Mean HE progression	N	
Any summer school	Yes	53.5%	11,487	59.1%	3,512	5.6 pp	55.5%	65,069	77.5%	31,243	22 pp
	No	51.2%	13,279	-	-		49.1%	69,136	-	-	
Any mentoring	Yes	58.3%	10,959	45.2%	4,040	-13.1 pp	64.1%	85,913	50.2%	10,399	-13.9 pp
	No	51.2%	13,279	-	-		49.1%	69,136	-	-	
Any project	Yes	50.9%	13,204	83.0%	1,795	32.1 pp	61.9%	90,107	73.0%	6,205	11.1 pp
	No	51.2%	13,279	-	-		49.1%	69,136	-	-	
Any campus visit	Yes	54.7%	11,782	55.1%	3,217	0.4 pp	64.8%	82,424	49.9%	13,888	-14.9 pp
	No	50.7%	9,971	52.9%	3,308	2.2pp	50.0%	52,152	46.5%	16,984	-3.5 pp
Any subject insight	Yes	60.5%	7,948	48.3%	7,051	-12.3 pp	69.4%	56,122	53.1%	40,190	-16.3 pp
	No	51.2%	13,279	-	-		49.1%	69,136	-	-	
Any skills and attainment	Yes	55.8%	11,112	51.8%	3,887	-4.1 pp	63.0%	69,643	61.8%	26,669	-1.2 pp
	No	50.0%	9,066	53.9%	4,213	4 pp	46.7%	43,261	53.2%	25,875	6.4 pp
Any HE information	Yes	56.2%	11,047	50.7%	3,952	-5.6 pp	65.2%	82,120	47.8%	14,192	-17.4 pp
	No	52.6%	8,069	49.1%	5,210	-3.5 pp	51.4%	46,276	44.5%	22,860	-6.9 pp
Any exhibition	Yes	54.7%	14,748	59.0%	251	4.3 pp	62.6%	94,802	62.6%	1,510	0.0 pp
	No	51.6%	12,539	44.6%	740	-7 pp	49.1%	65,363	49.9%	3,773	0.8 pp

Note: any student who took part in a summer school, mentoring, project of skills and attainment activity was involved in intensive HEAT activity (as defined in Table 4. Summary of intensity measures Table 4). Therefore, some cells are left blank on purpose. Differences are coded green for positive and blue for negative values.

- We next build a regression model including all demographic covariates (listed previously in Table 1) and our binary activity measures. The resulting regressions table are given in Table 26 and Table 27 in Annex B.
- For ease of interpretation, here we present the marginal effects (as described on page 11). Table 11 relates to the restricted dataset and Table 12 to the unrestricted dataset.

Table 11. Overall HE progression - marginal effects - function of binary activity marker - restricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.297*** (0.005)	0.297*** (0.005)	0.297*** (0.005)	0.297*** (0.005)	0.297*** (0.005)	0.297*** (0.005)	0.298*** (0.005)	0.297*** (0.005)	0.295*** (0.005)
Intense HEAT activity - yes	0.053*** (0.008)	0.067*** (0.008)	0.062*** (0.008)	0.062*** (0.008)	0.066*** (0.009)	0.062*** (0.008)	0.065*** (0.008)	0.061*** (0.008)	0.054*** (0.011)
Any summer school - yes	0.039*** (0.011)								0.052*** (0.012)
Any mentoring - yes		-0.019 (0.011)							-0.025* (0.011)
Any project -yes			0.007 (0.023)						0.032 (0.024)
Any campus visit - yes				0.019* (0.008)					0.024** (0.008)
Any subject insight - yes					-0.009 (0.010)				0.011 (0.011)
Any skills attainment - yes						-0.001 (0.008)			0.009 (0.008)
Any HE information - yes							0.024** (0.008)		0.032*** (0.008)
Any exhibition - yes								-0.025 (0.020)	-0.012 (0.020)
Observations	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278
Note:	*p<0.5; **p<0.01; ***p<0.001 Also controlling for a vector of all demographic covariates								

Table 12. Overall HE progression - marginal effects - function of binary activity marker - unrestricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.185*** (0.002)	0.188*** (0.002)	0.188*** (0.002)	0.189*** (0.002)	0.186*** (0.002)	0.189*** (0.002)	0.188*** (0.002)	0.189*** (0.002)	0.182*** (0.002)
KS4 score missing	-0.110*** (0.007)	-0.121*** (0.007)	-0.119*** (0.007)	-0.118*** (0.007)	-0.112*** (0.007)	-0.119*** (0.007)	-0.117*** (0.007)	-0.119*** (0.007)	-0.109*** (0.007)
Intense HEAT activity - yes	0.091*** (0.003)	0.135*** (0.003)	0.127*** (0.003)	0.132*** (0.003)	0.174*** (0.003)	0.132*** (0.003)	0.125*** (0.003)	0.134*** (0.003)	0.103*** (0.004)
Any summer school - yes	0.148*** (0.004)								0.137*** (0.004)
Any mentoring - yes		-0.027*** (0.006)							-0.018** (0.006)
Any project -yes			0.086*** (0.007)						0.099*** (0.007)
Any campus visit - yes				-0.005 (0.003)					0.001 (0.004)
Any subject insight - yes					-0.094*** (0.004)				-0.037*** (0.005)
Any skills attainment - yes						0.002 (0.003)			0.010** (0.003)
Any HE information - yes							-0.040*** (0.003)		-0.028*** (0.004)
Any exhibition - yes								0.044*** (0.007)	0.037*** (0.007)
Activity count									
Observations	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448

Note: *p<0.5; **p<0.01; ***p<0.001

Also controlling for a vector of all demographic covariates

- An interpretation is given in Table 13 Table 1below. Please note that attainment (and other demographic variables) are included in all models so all of the findings relating to the activities are not due to measurable differences between the groups on these factors.

Table 13. HE progression by activity - summary of findings from regression analysis

Variable	Restricted dataset	Unrestricted dataset	Agree?
Av. KS4 score	For every increase of one standard deviation in KS4 attainment, HE progression is 30pp higher.	For every increase of one standard deviation in KS4 attainment, HE progression is 19pp higher. Progression is 11pp lower for individuals for whom KS4 data is missing.	Yes
Intense HEAT activity	The HEAT intensity marker is associated with an increase in HE progression of 5pp. In other words, individuals who take part in an intensive package of HEAT activities are 5pp more likely to progress to HE than those who do not. This is after taking account of all demographic variables, prior attainment and participation in different types of activity (using binary measures as described above).	The HEAT intensity marker is associated with an increase in HE progression of 10pp. In other words, individuals who take part in an intensive package of HEAT activities are 10pp more likely to progress to HE than those who do not. This is after taking account of all demographic variables, prior attainment and participation in different types of activity (using binary measures as described above).	Somewhat
For all activities, the comparison is with students who have participated in a similarly intense package of activities			
Any summer school	Summer school attendance is associated with the biggest increase in HE progression; HE progression is 5pp higher for those who attended any summer school versus those who did not.	Summer school attendance is associated with the biggest increase in HE progression; HE progression is 14pp higher for those who attended any summer school versus those who did not.	Somewhat
Any mentoring	Mentoring is associated with a decrease in HE progression; HE progression is 3pp lower for those who attended any mentoring versus those who did not.	Mentoring is associated with a decrease in HE progression; HE progression is 2pp lower for those who attended any mentoring versus those who did not.	Yes
Any project	No significant relationship	Projects are associated with an increase in HE progression; HE progression is 10pp higher for those who attended any project versus those who did not.	No
Any campus visit	Campus visits are associated with an increase in HE progression; HE progression is 2pp higher for those who attended any campus versus those who did not.	No significant relationship.	No
Any subject insight	No significant relationship.	Subject insight events are associated with a decrease in HE progression; HE progression is 4pp lower for those who attended any subject insight event versus those who did not.	No
Any skills and attainment	No significant relationship.	Skills and attainment activities are associated with an increase in HE progression; HE progression is 1pp higher for those who attended any activity versus those who did not.	No
Any HE information	HE information sessions are associated with an increase in HE progression; HE progression is 3pp higher for those who attended any session those who did not.	HE information sessions are associated with a decrease in HE progression; HE progression is 3pp lower for those who attended any session those who did not.	No
Any exhibition	No significant relationship.	For each exhibition attended, HE progression is 4pp higher.	No

- This analysis finds, when comparing students who have participated in a similarly intense package of activities:
 - For every increase of one standard deviation in KS4 attainment, HE progression is 19-30pp higher.
 - The HEAT intensity marker is associated with an increase in HE progression of 5-10pp. In other words, individuals who take part in an intensive package of HEAT activities are 5-10pp more likely to progress to HE than those who do not. This is after taking account of all demographic variables, prior attainment and participation in different types of activity (using binary measures as described above).
 - The range of estimates for the value of the coefficient associated with attainment and the HEAT intensity marker is likely a produce of underlying differences between the students in the two groups, as discussed previously.
 - Summer school attendance is associated with the biggest increase in HE progression; HE progression is 5-14pp higher for those who attended any summer school versus those who did not. The size of the coefficient is larger in the unrestricted dataset; this could be because, as these are more likely to be post-16 students, attendance at a summer school for this group may be more effective or it may be a more meaningful measure of their existing likelihood to apply to HE.
 - Mentoring is associated with a decrease in HE progression; HE progression is 2-3pp lower for those who attended any mentoring versus those who did not.
 - Results relating to the other activities are mixed or suggest no significant relationship and so we do not report them.

- For completeness, we also build models which contain activity counts (i.e. counts of how many times a student has taken part in each activity) – see Table 28 and Table 29 in Annex B.

- This approach is arguably less useful than the one described above as we do not take into account any measure of the intensity of the different activities. Moreover, collapsing the counts to binary markers avoids any issues with how different activities might have been coded (for example, some HEPs record each mentoring encounter as an activity whereas others would only count a whole term). The results presented in the annex are mixed but broadly align with the findings discussed above.

4.3 Progression to top-third HEPs – demographic factors and intensity measures

- The analysis presented in Section 4.1 is now replicated for progression to top-third HE providers (see Annex C for a list of these providers).
- The descriptive statistics are given in Table 14 and align with the results when considering overall HE progression:
 - We see that the HEAT intensity measure is associated with higher rates of progression to top-third providers.
 - Females, non-FSM eligible students, those from more advantaged areas (based on IDACI quintile) those who have a family history of HE attendance and BAME students have higher rates of progression to top-third HEPs than other individuals.
 - We also note that individuals who first engaged with HEAT activity after the age of 16 have higher rates of top-third progression than those who first engaged at a younger age.

Table 14. Top-third progression by demographic characteristics and HEAT intensity measure

Variable/group	Category	Restricted			Unrestricted		
		Mean HE progression	N	Difference in HE progression	Mean HE progression	N	Difference in HE progression
Intense HEAT activity	Yes	54.8%	14,999	3.6pp	62.6%	96,312	13.5 pp
	No	51.2%	13,279		49.1%	69,136	
Female	Yes	55.3%	16,672	5.3pp	56.9%	79,744	6.4 pp
	No	50.0%	11,606		50.5%	56,865	
FSM	Yes	46.4%	3,651	7.7pp	48.7%	8,638	8.9 pp
	No	54.1%	24,627		57.6%	64,182	
IDACI quintile	1	50.7%	6,364	-	53.5%	33,500	-
	2	49.7%	7,583		53.4%	37,654	
	3	53.1%	6,053		55.6%	32,520	
	4	55.1%	4,652		59.7%	29,187	
	5	61.8%	3,626		66.6%	25,767	
First engaged in activity post-16	Yes	63.2%	13,314	19.1pp	62.6%	118,357	19.4 pp
	No	44.1%	14,964		43.2%	42,987	
First in family	Yes	50.3%	21,108	11.3pp	53.5%	66,108	4.2 pp
	No	61.5%	7,170		57.7%	24,753	
Ethnicity	Asian	76.4%	3,113	-	67.6%	9,949	-
	Black	72.8%	1,303		65.5%	5,404	
	Mixed	57.4%	1,074		53.1%	3,354	
	Other	71.4%	227		61.5%	1,065	
	White	48.4%	22,561		47.6%	58,352	

- We next build a regression model including all demographic covariates (listed previously in Table 1) and our measures of intensity. The resulting regression tables are given in Table 30 and Table 31 in Annex B.
- For ease of interpretation, here we present the marginal effects (as described on page 11). Table 15 relates to the restricted dataset and Table 16 relates to the unrestricted dataset.

Table 15. Top-third progression - marginal effects - function of demographic variables and intensity measures - restricted dataset

	Entered top-third provider			
	(1)	(2)	(3)	(4)
Gender - female	-0.020** (0.007)	-0.018* (0.007)	-0.060*** (0.007)	-0.060*** (0.007)
FSM - eligible	-0.019 (0.012)	-0.016 (0.012)	0.014 (0.012)	0.013 (0.012)
Engaged post-16 - yes	-0.031*** (0.009)	-0.032*** (0.009)	0.012 (0.009)	0.013 (0.009)
IDACI quintile	0.038*** (0.003)	0.036*** (0.003)	0.013*** (0.003)	0.013*** (0.003)
First in family - yes		-0.053*** (0.008)	-0.033*** (0.008)	-0.034*** (0.008)
Av. KS4 score			0.261*** (0.005)	0.261*** (0.005)
Intense HEAT activity - yes				0.005 (0.008)
Observations	15,018	15,018	15,018	15,018

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for ethnicity, HE ready year and school region

Table 16. Top-third progression - marginal effects - function of demographic variables and intensity measure - unrestricted dataset

	Entered top-third provider			
	(1)	(2)	(3)	(4)
Gender - female	-0.018*** (0.004)	-0.017*** (0.004)	-0.045*** (0.004)	-0.052*** (0.004)
Gender - unknown	0.101*** (0.006)	0.097*** (0.006)	0.108*** (0.006)	0.082*** (0.006)
FSM - eligible	-0.062*** (0.009)	-0.059*** (0.009)	0.021 (0.011)	0.012 (0.011)
FSM - unknown	0.105*** (0.004)	0.092*** (0.005)	0.151*** (0.013)	0.128*** (0.013)
Engaged post-16 - yes	0.041*** (0.005)	0.047*** (0.005)	0.053*** (0.005)	0.063*** (0.005)
Engaged post-16 - unknown	-0.079*** (0.015)	-0.069*** (0.015)	-0.065*** (0.016)	-0.036* (0.016)
IDACI quintile	0.067*** (0.001)	0.066*** (0.001)	0.054*** (0.001)	0.056*** (0.001)
IDACI - unknown	0.038*** (0.009)	0.044*** (0.009)	-0.015 (0.009)	0.022* (0.010)
First in family - yes		-0.051*** (0.005)	-0.044*** (0.006)	-0.067*** (0.006)
First in family - unknown		-0.098*** (0.006)	-0.097*** (0.006)	-0.107*** (0.006)
Av. KS4 score			0.302*** (0.003)	0.293*** (0.003)
KS4 unknown			0.239*** (0.012)	0.244*** (0.012)
Intense HEAT activity - yes				0.152*** (0.004)
Observations	94,285	94,285	94,285	94,285

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for ethnicity, HE ready year and school region

- An interpretation is given in Table 17 below.

Table 17. Top-third progression by demographic characteristics and HEAT intensity measure – summary of findings from regression analysis

Variable	Restricted dataset	Unrestricted dataset	Agree?
Gender	Top-third progression for female students is 2pp lower, and this increases to 5pp when we control for attainment.	Top-third progression for female students is 2pp lower, and this gap increases to 5pp when we control for attainment and participation in intense HEAT activity. Top-third progression is also 8pp higher for individuals whose gender is unknown.	Yes
FSM	When we control for IDACI, there is no significant association between FSM eligibility and top-third progression.	Top-third progression for FSM-eligible students is 6pp lower, even when controlling for other measures of disadvantage (IDACI and first in family). When we control for attainment, there is no significant association. Top-third progression is 13pp higher for individuals whose FSM status is unknown.	Somewhat
First engaged in activity post-16	In models which do not include attainment, top-third progression for those who first engaged post-16 is 3pp lower but when we control for attainment there is no significant association.	Top-third Progression for those who first engaged post-16 5pp higher. Top-third progression is 4pp lower for individuals whose date of first engagement with outreach is unknown.	No
IDACI	For every increase of one in IDACI quintile, top-third progression is 3pp higher. When we control for attainment, this reduces to around 1pp.	For every increase of one in IDACI quintile, top-third progression is 6pp higher. When we control for attainment, this reduces to around 5pp. Top-third progression is 2pp lower for individuals for whom IDACI quintile is unknown.	Yes
First in family	Top-third progression for individuals who would be first in their family to attend HE is 5pp lower but this reduces to 3pp when we control for attainment.	Top-third progression for individuals who would be first in their family to attend HE is 5pp lower. Top-third progression is 10pp lower for individuals where we do not know if they would be first in family to attend HE.	Yes
Av. KS4 score	For every increase of one standard deviation in KS4 attainment, top-third progression is 26pp higher.	For every increase of one standard deviation in KS4 attainment, top-third progression is 25pp higher. Progression is 24pp higher for individuals for whom KS4 data is missing.	Yes
Intense HEAT activity	There is no significant association between the HEAT intensity marker and top-third progression.	The HEAT intensity marker is associated with an increase in top-third progression of 15pp.	No

- In summary, the analysis finds, when controlling for other variables:
 - Although female students are more likely to progress to top-third HEPs in the raw data, when we control for other measurable factors, they are actually 5pp less likely to progress to these institutions.
 - Although FSM-eligible students are less likely to progress to top-third HEPs in the raw data, this pattern appears to disappear when we control for other measures of disadvantage and attainment; suggesting FSM-eligibility does not meaningfully predict this outcome over and above the other factors controlled for.
 - For every increase of one IDACI quintile, top-third progression is 3-6pp higher. This figure decreases when we control for attainment, suggesting that part of the reason for the lower rates of progression is lower attainment.
 - Top-third progression for individuals who would be first in their family to attend HE is 3-5pp lower.
 - For every increase of one standard deviation in KS4 attainment, top-third progression is 25-26pp higher, depending on which dataset you look at.
 - The HEAT intensity marker is only associated with higher top-third progression when we consider the unrestricted database. This finding may reflect the different age and profile of students contained in the two datasets.

4.4 Progression to top-third HEPs – binary measure of participation in activities versus HEAT intensity

- The analysis presented in Section 4.2 is now replicated for progression to top-third HE providers (see Annex C for a list of these providers).
- The descriptive statistics are given in Table 18 and align with the results when considering overall HE progression:
 - Summer schools are associated with an increase in top-third progression.
 - Mentoring, campus visits, subject insight events and HE information events are associated with a decrease in top-third progression.
 - The picture is less clear for the other activities.

Table 18. Top-third progression by participation in activities (binary)

Activity	Intense HEAT activity	Restricted dataset					Unrestricted dataset				
		Did not attend activity		Attended activity		Difference in HE progression	Did not attend activity		Attended activity		Difference in top-third progression
		Mean HE progression	N	Mean HE progression	N		Mean top-third progression	N	Mean top-third progression	N	
Any summer school	Yes	26.4%	6,142	34.4%	2,074	8 pp	44.2%	36,095	62.2%	24,216	18 pp
	No	23.6%	6,802	-	-	-	30.0%	33,974	-	-	-
Any mentoring	Yes	32.2%	6,390	15.2%	1,826	-17 pp	53.6%	55,086	28.9%	5,225	-24.6 pp
	No	23.6%	6,802	-	-	-	30.0%	33,974	-	-	-
Any project	Yes	25.9%	6,726	39.7%	1,490	13.8 pp	51.6%	55,782	49.2%	4,529	-2.4 pp
	No	23.6%	6,802	-	-	-	30.0%	33,974	-	-	-
Any campus visit	Yes	28.7%	6,443	27.4%	1,773	-1.3 pp	53.8%	53,386	33.1%	6,925	-20.7 pp
	No	23.8%	5,052	23.1%	1,750	-0.6 pp	31.0%	26,070	26.8%	7,904	-4.2 pp
Any subject insight	Yes	31.4%	4,812	24.2%	3,404	-7.2 pp	55.4%	38,962	44.1%	21,349	-11.3 pp
	No	23.6%	6,802	-	-	-	30.0%	33,974	-	-	-
Any skills and attainment	Yes	30.5%	6,204	22.1%	2,012	-8.4 pp	52.9%	43,842	47.5%	16,469	-5.3 pp
	No	22.3%	4,530	26.3%	2,272	4 pp	29.8%	20,218	30.3%	13,756	0.5 pp
Any HE information	Yes	32.5%	6,213	15.9%	2,003	-16.6 pp	55.6%	53,534	18.6%	6,777	-37 pp
	No	25.7%	4,245	20.1%	2,557	-5.6 pp	32.6%	23,799	23.9%	10,175	-8.7 pp
Any exhibition	Yes	28.6%	8,068	19.6%	148	-9 pp	51.5%	59,365	45.1%	946	-6.4 pp
	No	23.7%	6,472	21.8%	330	-1.9 pp	29.1%	32,091	45.6%	1,883	16.5 pp

Note: any student who took part in a summer school, mentoring, project of skills and attainment activity was involved in intensive HEAT activity (as defined in Table 4. **Summary of intensity measures Table 4**). Therefore, some cells are left blank on purpose. Differences are coded green for positive and blue for negative values.

- We next build a regression model including all demographic covariates (listed previously in Table 1) and our binary activity measures. The resulting regressions table are given in Table 32 and Table 33 in Annex B.
- For ease of interpretation, here we present the marginal effects (as described on page 11). Table 19 relates to the restricted dataset and Table 20 relates to the unrestricted dataset.

Table 19. Top-third progression - marginal effects - function of binary activity marker - restricted dataset

	Entered top-third provider								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.260*** (0.005)	0.259*** (0.005)	0.261*** (0.005)	0.261*** (0.005)	0.261*** (0.005)	0.261*** (0.005)	0.260*** (0.005)	0.261*** (0.005)	0.259*** (0.005)
Intense HEAT activity - yes	-0.005 (0.008)	0.014 (0.008)	0.006 (0.008)	0.005 (0.008)	0.011 (0.009)	0.004 (0.008)	0.002 (0.008)	0.005 (0.008)	0.009 (0.011)
Any summer school - yes	0.036** (0.011)								0.028* (0.013)
Any mentoring - yes		-0.048*** (0.011)							-0.045*** (0.011)
Any project -yes			-0.010 (0.015)						-0.012 (0.017)
Any campus visit - yes				0.006 (0.008)					0.007 (0.009)
Any subject insight - yes					-0.015 (0.010)				-0.013 (0.011)
Any skills attainment - yes						-0.017* (0.007)			-0.015 (0.008)
Any HE information - yes							-0.025** (0.008)		-0.022** (0.008)
Any exhibition - yes								0.032 (0.023)	0.021 (0.022)
Observations	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018
Note:	*p<0.5; **p<0.01; ***p<0.001 Also controlling for a vector of all demographic covariates								

Table 20. Top-third progression - marginal effects - function of binary activity marker - unrestricted dataset

	Entered top-third provider								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.287*** (0.004)	0.292*** (0.003)	0.293*** (0.003)	0.293*** (0.003)	0.292*** (0.003)	0.293*** (0.003)	0.290*** (0.003)	0.294*** (0.003)	0.283*** (0.004)
KS4 score missing	0.255*** (0.012)	0.241*** (0.012)	0.244*** (0.012)	0.247*** (0.012)	0.250*** (0.012)	0.243*** (0.012)	0.249*** (0.012)	0.242*** (0.012)	0.259*** (0.012)
Intense HEAT activity - yes	0.111*** (0.004)	0.156*** (0.004)	0.151*** (0.004)	0.143*** (0.004)	0.164*** (0.004)	0.151*** (0.004)	0.139*** (0.004)	0.157*** (0.004)	0.097*** (0.006)
Any summer school - yes	0.122*** (0.005)								0.116*** (0.006)
Any mentoring - yes		-0.069*** (0.008)							-0.044*** (0.009)
Any project -yes			0.007 (0.008)						0.034*** (0.009)
Any campus visit - yes				-0.072*** (0.005)					-0.065*** (0.005)
Any subject insight - yes					-0.037*** (0.005)				0.002 (0.006)
Any skills attainment - yes						-0.020*** (0.004)			-0.007 (0.005)
Any HE information - yes							-0.082*** (0.005)		-0.070*** (0.006)
Any exhibition - yes								0.112*** (0.011)	0.089*** (0.011)
Activity count									
Observations	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for a vector of all demographic covariates

- An interpretation is given in Table 21 below. Please note that attainment (and other demographic variables) are included in all models so all of the findings relating to the activities are not due to measurable differences between the groups on these factors.

Table 21. . Top-third progression by activity - summary of findings from regression analysis

Variable	Restricted dataset	Unrestricted dataset	Agree?
Av. KS4 score	For every increase of one standard deviation in KS4 attainment, top-third progression is 26pp higher.	For every increase of one standard deviation in KS4 attainment, top-third progression is 29pp higher.	Yes
Intense HEAT activity	Not consistent evidence of a significant relationship.	The HEAT intensity marker is associated with an increase in top-third progression of 10pp. In other words, individuals who take part in an intensive package of HEAT activities are 10pp more likely to progress to a top-third HEP than those who do not. This is after taking account of all demographic variables, prior attainment and participation in different types of activity (using binary measures as described above).	No
For all activities, the comparison is with students who have participated in a similarly intense package of activities			
Any summer school	Summer school attendance is associated with an increase in top-third progression; top-third progression is 3pp higher for those who attended <u>any</u> summer school versus those who did not.	Summer school attendance is associated with an increase in top-third progression; top-third progression is 12pp higher for those who attended <u>any</u> summer school versus those who did not.	Somewhat
Any mentoring	Mentoring is associated with a decrease in top-third progression; HE progression is 5pp lower for those who attended <u>any</u> mentoring versus those who did not.	Mentoring is associated with a decrease in top-third progression; top-third progression is 5pp lower for those who attended <u>any</u> mentoring versus those who did not.	Yes
Any project	No significant relationship	Top-third progression is 3pp higher for those who attended <u>any</u> project versus those who did not.	No
Any campus visit	No significant relationship	Top-third progression is -7pp lower for those who attended <u>any</u> campus visit versus those who did not.	No
Any subject insight	No significant relationship	No significant relationship when we control for participation in other activities.	Somewhat
Any skills and attainment	No significant relationship when we control for participation in other activities.	No significant relationship when we control for participation in other activities.	Yes
Any HE information	Top-third progression is 2pp lower for those who attended <u>any</u> HE information session versus those who did not.	Top-third progression is 7pp lower for those who attended <u>any</u> HE information session versus those who did not.	Yes

- This analysis finds, when comparing students who have participated in a similarly intense package of activities:
 - For every increase of one standard deviation in KS4 attainment, top-third progression is 26-29pp higher.
 - There is some discrepancy between the results from the two different datasets which is probably due to the underlying differences between the students included: only the unrestricted dataset suggests the HEAT intensity marker is associated with an increase in top-third progression (of 10pp). This is after taking account of all demographic variables, prior attainment and participation in different types of activity (using binary measures as described above).
 - Summer school attendance is associated with an increase in top-third progression; Top-third progression is 3-12pp higher for those who attended any summer school versus those who did not. The size of the coefficient is larger in the unrestricted dataset; this could be because, as these are more likely to be post-16 students, attendance at a summer school for this group may be more effective or it may be a more meaningful measure of their existing likelihood to apply to HE and/or a top-third HEP.
 - Mentoring is associated with a decrease in top-third progression; top-third progression is 5pp lower for those who attended any mentoring versus those who did not.
 - HE information sessions are associated with a decrease in top-third progression; top-third progression is 2-7pp lower for those who attended any sessions versus those who did not.
 - Results relating to the other activities are mixed or suggest no significant relationship and so we do not report them.

- For completeness, again we build models which contain activity counts (i.e. counts of how many times a student has taken part in each activity) – see Table 34 and Table 35 in Annex B.

- As stated previously, this approach is arguably less useful than the one described above as we do not take into account any measure of the intensive of the different activities. Moreover, collapsing the counts to binary markers avoids any issues with how different activities might have been coded (for example, some HEPs record each mentoring encounter as an activity whereas others would only count a whole term). The results presented in the annex are mixed but broadly align with the findings discussed above.

4.1 Participation in activities

- To contextualise our findings, we examine whether there are any significant associations between participating in activities and demographic factors.
- The table below models binary participation in activities (0 or 1) as a function of demographic variables. The table presents marginal effects from a binary logistic regression using the unrestricted dataset.
- This analysis shows:
 - KS4 attainment is positively associated with summer schools and HE information sessions but negatively associated with mentoring, implying that higher attaining students may access summer schools while mentoring may be targeted at those with lower grades.
 - FSM eligibility is positively associated with mentoring and HE information sessions but negatively associated with summer schools, implying that summer schools may attract more advantaged participants.
 - These findings may help explain the patterns of HE and top-third progression we see associated with these activities elsewhere in this paper.

Table 22. Participation in activities - marginal effects - function of demographic variables

	Participation in activities - marginal effects - function of demographic variables		
	Participated at least once		
	Summer school (1)	Mentoring (2)	HE Information (3)
Av. KS4 score	0.060*** (0.001)	-0.004*** (0.0004)	0.008*** (0.0004)
Female - yes	0.030*** (0.002)	0.004*** (0.001)	0.003*** (0.001)
Female- unknown	0.152*** (0.004)	0.007*** (0.002)	0.008*** (0.001)
FSM - yes	-0.008* (0.004)	0.005*** (0.002)	0.014*** (0.002)
FSM - unknown	0.086*** (0.002)	-0.010*** (0.001)	-0.010*** (0.001)
First in family - yes	0.066*** (0.003)	0.016*** (0.001)	0.018*** (0.001)
First in family - unknown	-0.089*** (0.003)	-0.028*** (0.002)	0.002** (0.001)
Engaged post-16 - yes	-0.015*** (0.002)	-0.032*** (0.001)	0.004*** (0.001)
Engaged post-16 - unknown	-0.054*** (0.004)	-0.032*** (0.001)	-0.015*** (0.002)
Observations	165,448	165,448	165,448

Note:

*p<0.5; **p<0.01; ***p<0.001

Also controlling for ethnicity, HE ready year and school region

Annex A: Missingness analysis

- The following table reports on a binary logistic regression which examines missingness in our data.
- The outcome is 1 if the student is missing in the restricted dataset (because of missing data) and 0 otherwise.

Table 23. Missing data - logistic regression - function of demographic variables

	Missing - not present in restricted
Av. KS4 score	0.078*** (0.007)
HE ready year	-2.385*** (0.043)
Female - yes	-0.263*** (0.069)
FSM - yes	0.392*** (0.132)
IDACI	-0.023 (0.027)
First generation - yes	0.006 (0.083)
Post-16 - yes	0.518*** (0.071)
Constant	3.598*** (0.431)
Observations	32,234

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for ethnicity and region

Annex B: Logistic regression tables

Overall HE progression - demographic

Table 24. Overall HE progression - logistic regression - function of demographic variables and intensity measures - restricted dataset

	Entered HE			
	(1)	(2)	(3)	(4)
Gender - female	0.183*** (0.026)	0.193*** (0.026)	-0.004 (0.029)	-0.011 (0.029)
FSM - eligible	-0.451*** (0.040)	-0.435*** (0.040)	-0.178*** (0.045)	-0.185*** (0.045)
Engaged post-16 - yes	0.269*** (0.029)	0.258*** (0.030)	0.456*** (0.033)	0.497*** (0.033)
IDACI quintile	0.191*** (0.010)	0.175*** (0.010)	0.087*** (0.011)	0.097*** (0.011)
First in family - yes		-0.360*** (0.030)	-0.308*** (0.033)	-0.322*** (0.033)
Av. KS4 score			1.209*** (0.019)	1.198*** (0.019)
Intense HEAT activity - yes				0.250*** (0.031)
Constant	1.110*** (0.082)	1.417*** (0.086)	1.983*** (0.097)	1.939*** (0.097)
Observations	28,278	28,278	28,278	28,278

Note:

*p<0.5; **p<0.01; ***p<0.001

Also controlling for ethnicity, HE ready year and school region

Table 25. Overall HE progression - function of demographic variables and intensity measure - unrestricted dataset

	Entered HE			
	(1)	(2)	(3)	(4)
Gender - female	0.245*** (0.012)	0.249*** (0.012)	0.169*** (0.012)	0.146*** (0.012)
Gender - unknown	0.507*** (0.017)	0.512*** (0.017)	0.490*** (0.018)	0.414*** (0.018)
FSM - eligible	-0.372*** (0.025)	-0.365*** (0.025)	-0.096*** (0.028)	-0.118*** (0.028)
FSM - unknown	-0.180*** (0.014)	-0.211*** (0.015)	0.422*** (0.036)	0.347*** (0.036)
Engaged post-16 - yes	0.476*** (0.013)	0.495*** (0.013)	0.514*** (0.014)	0.575*** (0.014)
Engaged post-16 - unknown	-0.339*** (0.035)	-0.324*** (0.035)	-0.384*** (0.037)	-0.339*** (0.037)
IDACI quintile	0.177*** (0.004)	0.174*** (0.004)	0.138*** (0.004)	0.143*** (0.004)
IDACI - unknown	-0.157*** (0.026)	-0.140*** (0.026)	-0.361*** (0.028)	-0.289*** (0.028)
First in family - yes		-0.212*** (0.016)	-0.172*** (0.017)	-0.228*** (0.017)
First in family - unknown		-0.380*** (0.019)	-0.405*** (0.019)	-0.411*** (0.020)
Av. KS4 score			0.790*** (0.008)	0.780*** (0.008)
KS4 unknown			-0.529*** (0.031)	-0.497*** (0.031)
Intense HEAT activity - yes				0.545*** (0.012)
Constant	0.768*** (0.037)	0.854*** (0.040)	1.176*** (0.043)	1.045*** (0.043)
Observations	165,448	165,448	165,448	165,448

Note:

*p<0.5; **p<0.01; ***p<0.001

Also controlling for ethnicity, HE ready year and school region

Overall HE progression – binary activity

Table 26. Overall HE progression - logistic regression - function of binary activity marker - restricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.196***	1.196***	1.197***	1.197***	1.197***	1.198***	1.198***	1.197***	1.189***
	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)
Intense HEAT activity - yes	0.215***	0.271***	0.248***	0.250***	0.267***	0.250***	0.261***	0.247***	0.216***
	(0.032)	(0.033)	(0.031)	(0.031)	(0.036)	(0.031)	(0.031)	(0.031)	(0.046)
Any summer school - yes	0.158***								0.211***
	(0.047)								(0.051)
Any mentoring - yes		-0.074							-0.099*
		(0.044)							(0.045)
Any project -yes			0.028						0.129
			(0.092)						(0.099)
Any campus visit - yes				0.078*					0.096**
				(0.033)					(0.034)
Any subject insight - yes					-0.037				0.043
					(0.041)				(0.045)
Any skills attainment - yes						-0.002			0.038
						(0.031)			(0.033)
Any HE information - yes							0.099**		0.128***
							(0.032)		(0.033)
Any exhibition - yes								-0.101	-0.048
								(0.079)	(0.080)
Activity count	1.917***	1.949***	1.939***	1.929***	1.937***	1.939***	1.881***	1.972***	1.851***
	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.099)	(0.101)	(0.104)
Observations	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278
Note:	*p<0.5; **p<0.01; ***p<0.001								
	Also controlling for a vector of all demographic covariates								

Table 27. Overall HE progression - logistic regression - function of binary activity marker - unrestricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.764*** (0.008)	0.778*** (0.008)	0.776*** (0.008)	0.780*** (0.008)	0.771*** (0.008)	0.780*** (0.008)	0.775*** (0.008)	0.780*** (0.008)	0.753*** (0.008)
KS4 score missing	-0.462*** (0.031)	-0.507*** (0.031)	-0.496*** (0.031)	-0.496*** (0.031)	-0.469*** (0.031)	-0.497*** (0.031)	-0.489*** (0.031)	-0.500*** (0.031)	-0.456*** (0.031)
Intense HEAT activity - yes	0.374*** (0.012)	0.557*** (0.012)	0.523*** (0.012)	0.543*** (0.012)	0.719*** (0.014)	0.545*** (0.012)	0.514*** (0.012)	0.554*** (0.012)	0.425*** (0.017)
Any summer school - yes	0.647*** (0.018)								0.597*** (0.021)
Any mentoring - yes		-0.110*** (0.024)							-0.074** (0.024)
Any project -yes			0.372*** (0.032)						0.432*** (0.034)
Any campus visit - yes				-0.019 (0.014)					0.003 (0.015)
Any subject insight - yes					-0.384*** (0.015)				-0.153*** (0.018)
Any skills attainment - yes						0.010 (0.013)			0.043** (0.014)
Any HE information - yes							-0.166*** (0.014)		-0.115*** (0.015)
Any exhibition - yes								0.186*** (0.030)	0.155*** (0.031)
Activity count	0.938*** (0.044)	1.058*** (0.044)	1.052*** (0.044)	1.049*** (0.044)	0.996*** (0.044)	1.041*** (0.044)	1.125*** (0.044)	1.035*** (0.044)	0.973*** (0.045)
Observations	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for a vector of all demographic covariates

Overall HE progression – activity count

Table 28. Overall HE progression - logistic regression - function of activity counts - restricted dataset

Overall HE progression - logistic regression - function of activity counts - restricted dataset									
	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.204*** (0.019)	1.209*** (0.019)	1.206*** (0.020)	1.208*** (0.019)	1.207*** (0.019)	1.208*** (0.019)	1.210*** (0.019)	1.208*** (0.019)	1.194*** (0.020)
Summer school (count)	0.218*** (0.036)								0.236*** (0.037)
Mentoring (count)		0.011 (0.007)							0.002 (0.007)
Project (count)			0.062 (0.043)						0.107* (0.044)
Campus visit (count)				0.074** (0.024)					0.082*** (0.024)
Subject insight (count)					0.069** (0.024)				0.080*** (0.024)
Skills attainment (count)						0.033 (0.019)			0.043* (0.019)
HE information (count)							0.067*** (0.015)		0.070*** (0.015)
Exhibition (count)								-0.118 (0.075)	-0.051 (0.075)
Constant	1.938*** (0.097)	1.977*** (0.097)	1.984*** (0.097)	1.973*** (0.097)	1.975*** (0.097)	1.980*** (0.097)	1.935*** (0.098)	2.022*** (0.100)	1.877*** (0.102)
Observations	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278	28,278

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for a vector of all demographic covariates

Table 29. Overall HE progression - function of activity counts - unrestricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	0.762*** (0.008)	0.786*** (0.008)	0.780*** (0.008)	0.785*** (0.008)	0.785*** (0.008)	0.784*** (0.008)	0.782*** (0.008)	0.785*** (0.008)	0.753*** (0.008)
KS4 score missing	-0.431*** (0.031)	-0.503*** (0.031)	-0.510*** (0.031)	-0.511*** (0.031)	-0.514*** (0.031)	-0.516*** (0.031)	-0.509*** (0.031)	-0.514*** (0.031)	-0.421*** (0.031)
Summer school (count)	0.514*** (0.010)								0.518*** (0.010)
Mentoring (count)		0.025*** (0.005)							0.018*** (0.005)
Project (count)			0.202*** (0.014)						0.216*** (0.014)
Campus visit (count)				-0.024*** (0.007)					-0.023*** (0.007)
Subject insight (count)					0.011 (0.006)				0.034*** (0.007)
Skills attainment (count)						0.039*** (0.004)			0.046*** (0.004)
HE information (count)							-0.072*** (0.006)		-0.053*** (0.006)
Exhibition (count)								0.025 (0.024)	0.069** (0.024)
Constant	0.965*** (0.044)	1.145*** (0.043)	1.158*** (0.043)	1.155*** (0.043)	1.157*** (0.043)	1.161*** (0.043)	1.183*** (0.043)	1.154*** (0.043)	0.991*** (0.044)
Observations	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448	165,448

Note: *p<0.5; **p<0.01; ***p<0.001

Also controlling for a vector of all demographic covariates

Top-third progression - demographic

Table 30. Top-third progression - logistic regression - function of demographic variables and intensity measures - restricted dataset

Top-third progression - logistic regression - function of demographic variables and intensity measures - restricted dataset				
	Entered top-third provider			
	(1)	(2)	(3)	(4)
Gender - female	-0.107** (0.039)	-0.097* (0.039)	-0.382*** (0.044)	-0.383*** (0.044)
FSM - eligible	-0.101 (0.064)	-0.085 (0.065)	0.087 (0.073)	0.086 (0.073)
Engaged post-16 - yes	-0.163*** (0.049)	-0.171*** (0.049)	0.081 (0.056)	0.085 (0.057)
IDACI quintile	0.204*** (0.016)	0.190*** (0.016)	0.085*** (0.018)	0.086*** (0.018)
First in family - yes		-0.277*** (0.043)	-0.211*** (0.048)	-0.214*** (0.048)
Av. KS4 score			1.699*** (0.037)	1.698*** (0.037)
Intense HEAT activity - yes				0.030 (0.049)
Constant	-2.044*** (0.120)	-1.814*** (0.125)	-2.455*** (0.142)	-2.461*** (0.142)
Observations	15,018	15,018	15,018	15,018
<i>Note:</i>	*p<0.5; **p<0.01; ***p<0.001 Also controlling for ethnicity, HE ready year and school region			

Table 31. Top-third progression - logistic regression - function of demographic variables and intensity measure - unrestricted dataset

	Entered top-third provider			
	(1)	(2)	(3)	(4)
Gender - female	-0.074*** (0.017)	-0.069*** (0.017)	-0.189*** (0.018)	-0.221*** (0.018)
Gender - unknown	0.410*** (0.022)	0.391*** (0.023)	0.445*** (0.024)	0.339*** (0.025)
FSM - eligible	-0.260*** (0.040)	-0.245*** (0.040)	0.088 (0.047)	0.050 (0.047)
FSM - unknown	0.431*** (0.019)	0.377*** (0.019)	0.646*** (0.056)	0.545*** (0.056)
Engaged post-16 - yes	0.169*** (0.020)	0.195*** (0.021)	0.227*** (0.022)	0.270*** (0.022)
Engaged post-16 - unknown	-0.334*** (0.066)	-0.292*** (0.066)	-0.284*** (0.071)	-0.154* (0.071)
IDACI quintile	0.272*** (0.006)	0.271*** (0.006)	0.228*** (0.006)	0.235*** (0.006)
IDACI - unknown	0.153*** (0.037)	0.177*** (0.037)	-0.065 (0.040)	0.092* (0.040)
First in family - yes		-0.210*** (0.023)	-0.187*** (0.024)	-0.284*** (0.024)
First in family - unknown		-0.404*** (0.026)	-0.410*** (0.027)	-0.455*** (0.027)
Av. KS4 score			1.271*** (0.015)	1.236*** (0.015)
KS4 unknown			1.052*** (0.054)	1.074*** (0.055)
Intense HEAT activity - yes				0.658*** (0.017)
Constant	-2.140*** (0.050)	-2.084*** (0.054)	-3.172*** (0.061)	-3.400*** (0.062)
Observations	94,285	94,285	94,285	94,285

Note:

*p<0.5; **p<0.01; ***p<0.001

Also controlling for ethnicity, HE ready year and school region

Top-third progression – binary activity

Table 32. Top-third progression - logistic regression - function of binary activity marker - restricted dataset

Top-third progression - logistic regression - function of binary activity marker - restricted dataset									
	Entered top-third provider								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.694*** (0.037)	1.689*** (0.037)	1.699*** (0.037)	1.699*** (0.037)	1.698*** (0.037)	1.700*** (0.037)	1.695*** (0.037)	1.700*** (0.037)	1.687*** (0.037)
Intense HEAT activity - yes	-0.033 (0.053)	0.090 (0.051)	0.038 (0.051)	0.031 (0.049)	0.071 (0.056)	0.023 (0.049)	0.010 (0.050)	0.034 (0.049)	0.060 (0.074)
Any summer school - yes	0.222** (0.066)								0.176* (0.079)
Any mentoring - yes		-0.341*** (0.083)							-0.314*** (0.085)
Any project -yes			-0.065 (0.099)						-0.082 (0.119)
Any campus visit - yes				0.037 (0.052)					0.043 (0.055)
Any subject insight - yes					-0.098 (0.065)				-0.083 (0.076)
Any skills attainment - yes						-0.116* (0.050)			-0.098 (0.054)
Any HE information - yes							-0.168** (0.054)		-0.146** (0.056)
Any exhibition - yes							0.195 (0.133)	0.130 (0.135)	
Activity count	-2.449*** (0.142)	-2.395*** (0.143)	-2.459*** (0.142)	-2.465*** (0.142)	-2.455*** (0.142)	-2.434*** (0.143)	-2.346*** (0.147)	-2.512*** (0.146)	-2.301*** (0.154)
Observations	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018
<i>Note:</i>	*p<0.5; **p<0.01; ***p<0.001 Also controlling for a vector of all demographic covariates								

Table 33. Top-third progression - logistic regression - function of binary activity marker - unrestricted dataset

Top-third progression - logistic regression - function of binary activity marker - unrestricted dataset									
	Entered top-third provider								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.207*** (0.015)	1.230*** (0.015)	1.236*** (0.015)	1.234*** (0.015)	1.229*** (0.015)	1.233*** (0.015)	1.220*** (0.015)	1.237*** (0.015)	1.190*** (0.015)
KS4 score missing	1.125*** (0.055)	1.059*** (0.055)	1.075*** (0.055)	1.089*** (0.055)	1.102*** (0.055)	1.068*** (0.055)	1.101*** (0.055)	1.065*** (0.055)	1.147*** (0.055)
Intense HEAT activity - yes	0.478*** (0.019)	0.680*** (0.017)	0.656*** (0.017)	0.621*** (0.017)	0.715*** (0.019)	0.654*** (0.017)	0.600*** (0.018)	0.681*** (0.017)	0.417*** (0.025)
Any summer school - yes	0.504*** (0.021)								0.479*** (0.026)
Any mentoring - yes		-0.300*** (0.037)							-0.188*** (0.038)
Any project -yes			0.030 (0.035)						0.141*** (0.039)
Any campus visit - yes				-0.314*** (0.023)					-0.282*** (0.023)
Any subject insight - yes					-0.158*** (0.021)				0.009 (0.026)
Any skills attainment - yes						-0.084*** (0.019)			-0.030 (0.021)
Any HE information - yes							-0.357*** (0.024)		-0.300*** (0.025)
Any exhibition - yes								0.456*** (0.044)	0.364*** (0.045)
Activity count	-3.506*** (0.063)	-3.369*** (0.062)	-3.400*** (0.062)	-3.324*** (0.062)	-3.423*** (0.062)	-3.372*** (0.062)	-3.229*** (0.063)	-3.418*** (0.062)	-3.278*** (0.065)
Observations	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for a vector of all demographic covariates

Top-third progression – activity count

Table 34. Top-third progression - function of activity counts - restricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.696*** (0.037)	1.693*** (0.037)	1.700*** (0.037)	1.699*** (0.037)	1.700*** (0.037)	1.701*** (0.037)	1.694*** (0.037)	1.701*** (0.037)	1.690*** (0.037)
Summer school (count)	0.113* (0.053)								0.115* (0.055)
Mentoring (count)		-0.060*** (0.017)							-0.054** (0.018)
Project (count)			-0.019 (0.046)						-0.025 (0.048)
Campus visit (count)				0.014 (0.035)					0.017 (0.035)
Subject insight (count)					-0.046 (0.041)				-0.044 (0.041)
Skills attainment (count)						-0.069* (0.030)			-0.057 (0.031)
HE information (count)							-0.105*** (0.029)		-0.092** (0.030)
Exhibition (count)								0.174 (0.125)	0.131 (0.126)
Constant	-2.458*** (0.142)	-2.417*** (0.142)	-2.453*** (0.142)	-2.456*** (0.142)	-2.445*** (0.142)	-2.445*** (0.142)	-2.366*** (0.144)	-2.501*** (0.146)	-2.363*** (0.149)
Observations	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018	15,018

Note: *p<0.5; **p<0.01; ***p<0.001
Also controlling for a vector of all demographic covariates

Table 35. Top-third progression - function of activity counts - unrestricted dataset

	Entered HE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. KS4 score	1.249*** (0.015)	1.284*** (0.015)	1.284*** (0.015)	1.283*** (0.015)	1.286*** (0.015)	1.285*** (0.015)	1.264*** (0.015)	1.286*** (0.015)	1.233*** (0.015)
KS4 missing	1.135*** (0.055)	1.040*** (0.054)	1.049*** (0.054)	1.054*** (0.054)	1.036*** (0.054)	1.041*** (0.054)	1.079*** (0.055)	1.041*** (0.054)	1.156*** (0.055)
Summer school (count)	0.393*** (0.010)								0.389*** (0.010)
Mentoring (count)		-0.017* (0.008)							-0.006 (0.009)
Project (count)			0.064*** (0.013)						0.092*** (0.014)
Campus visit (count)				-0.219*** (0.013)					-0.163*** (0.013)
Subject insight (count)					0.040*** (0.009)				0.083*** (0.010)
Skills attainment (count)						-0.034*** (0.006)			-0.004 (0.006)
HE information (count)							-0.298*** (0.013)		-0.225*** (0.013)
Exhibition (count)								0.114** (0.036)	0.144*** (0.036)
Constant	-3.304*** (0.062)	-3.076*** (0.061)	-3.083*** (0.061)	-3.052*** (0.061)	-3.075*** (0.061)	-3.081*** (0.061)	-2.938*** (0.061)	-3.082*** (0.061)	-3.162*** (0.062)
Observations	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285	94,285

Note: *p<0.5; **p<0.01; ***p<0.001

Also controlling for a vector of all demographic covariates

Annex C: Top-third HEPs ¹³

Cardiff University
Courtauld Institute of Art
Glasgow School of Art
Guildhall School of Music and Drama
Heriot-Watt University
Imperial College of Science, Technology and Medicine
King's College London
London School of Economics and Political Science
Loughborough University
Queen Mary University of London
Queens University Belfast
Royal Academy of Music
Royal College of Music
Royal Conservatoire of Scotland
Royal Holloway and Bedford New College
Royal Northern College of Music
St George's Hospital Medical School
Stranmillis University College
The Royal Central School of Speech and Drama
The Royal Veterinary College
The School of Oriental and African Studies
The University of Aberdeen
The University of Bath
The University of Birmingham
The University of Bristol

¹³ From Department for Education Destinations reporting.

The University of Cambridge
The University of Dundee
The University of East Anglia
The University of Edinburgh
The University of Exeter
The University of Glasgow
The University of Lancaster
The University of Leeds
The University of Leicester
The University of Liverpool
The University of Manchester
The University of Oxford
The University of Reading
The University of Sheffield
The University of Southampton
The University of St Andrews
The University of Strathclyde
The University of Surrey
The University of Sussex
The University of Warwick
The University of York
University College London
University of Abertay Dundee
University of Durham
University of Newcastle-upon-Tyne
University of Nottingham