

Partnership Version

# HeppSY Cohort Evaluation

Contact Hours and Expectations of  
Applying to HE Aged 18 or 19

October 2020

Authors: Daniel Fletcher and Greg Brown



# Contents

<b>3</b>	<b>Introduction</b>
<b>4</b>	<b>Methodology</b>
4	Matching Process
5	Measures
6	Sample
7	Analytical Method
<b>8</b>	<b>Analysis and Results</b>
8	Regression Analysis
9	Exploratory Analysis
<b>10</b>	<b>Conclusions</b>
10	Limitations
<b>11</b>	<b>Recommendations</b>
11	Recommendations for HeppSY Practice
11	Recommendations for Future Research and Evaluation
<b>12</b>	<b>References</b>

# Introduction

---

Drawing on data from Wave 1 (W1) and Wave 2 (W2) of the CFE annual learner survey, and HeppSY and partner activity data held in the Higher Education Access Tracker (HEAT); the aim of this project was to investigate the association between the number of contact hours participated in and changes in expectation of applying to HE for students that were in Year 10 during the first wave of the survey (Year 11 at W2).



Previous analysis of the HeppSY CFE survey data has shown that Year 11 students have lower expectations of applying to HE than either Year 9 or Year 10 students (HeppSY, 2020), and this relationship has also been observed longitudinally in the wider literature (Anders & Micklewright, 2015). The present analysis was therefore an opportunity to assess whether level of participation in the HeppSY programme was associated with a reduced likelihood of a negative shift in expectations between Year 10 and Year 11.

As such, the primary research question was:

After controlling for prior expectations, what is the association between the number of HeppSY contact hours received by students in Year 10, and their expectation of applying to HE in Year 11?

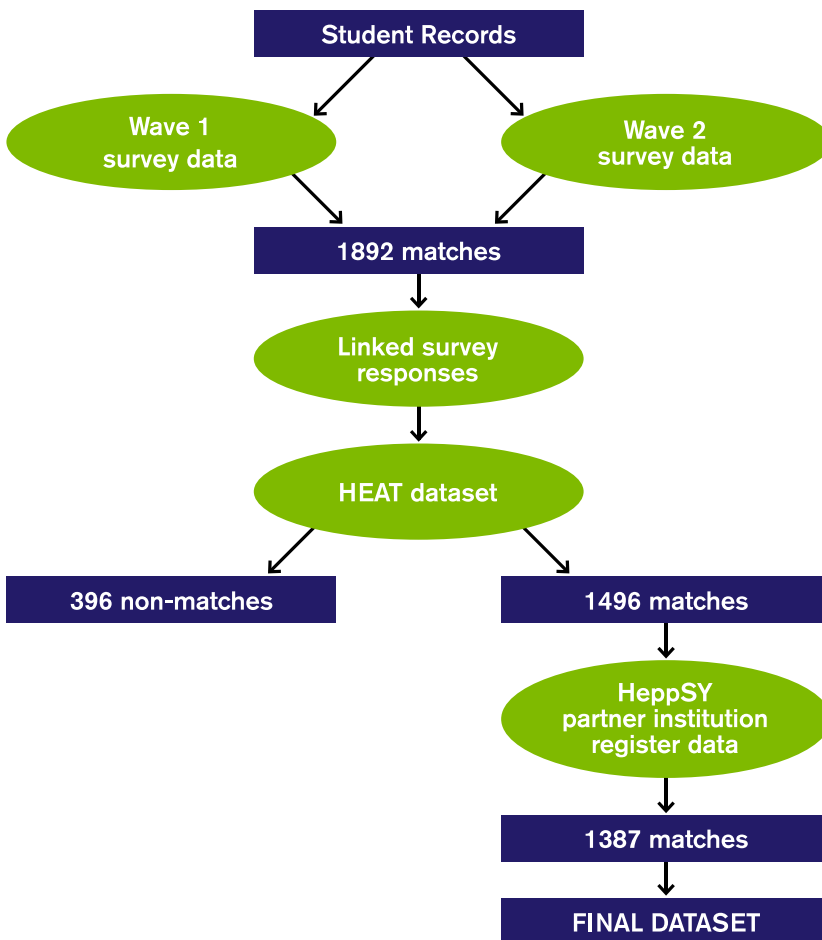
This analysis will be supported by further work utilising the same dataset, but focussing on different questions from the learner survey, to give more detailed insight into the impact of specific HeppSY interventions across the four strands of the programme (HE knowledge, career knowledge, confidence and resilience, attainment).

# Methodology

## Matching Process

The process used for matching students across datasets is described below. Note that the present analysis focused on students who were in Year 10 at W1, and therefore only used a subset of the overall matched dataset.

Individual students were matched across the two waves of survey data using fuzzy matching in Excel Power Query. Matches of 90% or higher were accepted, subject to random spot-checks of the data. This resulted in 1892 matches. Student records were merged so that every case had a W1 and a W2 response for each survey question.



The linked survey responses were then matched to individual student activity records in the HEAT dataset. Activity was filtered by date, so that only engagements that took place in-between the close of the W1 survey and the opening of W2 were included in the matching process. Fuzzy matching of the linked survey and HEAT datasets resulted in 1496 matches.

The 396 linked survey records that could not be matched to the HEAT dataset were either due to data quality issues, or because a student had not participated in any activity between survey waves. As non-matches do not necessarily indicate non-participation, these records were excluded from the dataset rather than being included as zero activity students.

Finally, the 1496 students that had been matched to both survey waves and to the HEAT dataset were matched to HeppSY partner institution register data, to add the HE potential coding that schools/colleges had assigned to students. This resulted in 1387 matches, which formed the final dataset of students.



## Measures

### Likelihood of applying to HE (Survey W1 and W2)

Students were asked, 'how likely are you to apply to higher education at age 18 or 19?' Responses were given on a six-point ordered scale (1 – definitely won't apply, 2 – very unlikely, 3 – fair unlikely, 4 – fairly likely, 5 – very likely, 6 – definitely will apply), with a don't know response option also available.

### Motivation (Survey W1):

Students were asked to rate their level of agreement with the statement, 'I am motivated to do well in my studies.' Responses were given on a 5-point Likert scale (1 – strongly disagree, 2 - slightly disagree, 3 – neither agree nor disagree, 4 – slightly agree, 5 – strongly agree).

### Contact hours (HEAT)

Reflects the cumulative number of hours of HeppSY outreach a student had participated in between the two survey waves.

### HE Potential Coding (School Registers)

HE potential coding is assigned to students by KPOCs at HeppSY institutions, based on whether a student lives in a target postcode, and whether they believe that the student is academically capable of progressing to HE.

Green = student is from a HeppSY target postcode and has the potential to progress to HE.

Red = student is from a HeppSY target postcode but does not have the potential to progress to HE.

Non-UCP = student is not from a HeppSY target postcode.

## Sample

The final sample included 514 students who were in Year 10 at W1. See Table 1 for a breakdown of respondents by gender and HE potential. Students from 20 different schools were included in the sample (school sample sizes ranged from 6 - 62 students).

**Table 1.** *Sample size by gender and HE potential coding.*

Gender	HE Potential			Total
	Green	Red	Non-UCP	
Male	93	30	82	205
Female	159	29	106	294
Prefer not to say / Other	9	1	5	15
Total	264	60	199	514

## Analytical Method

The data were analysed using ordinal logistic regression. Regression analysis refers to a statistical procedure used to estimate the relationship between an outcome variable and one or more predictor variables. Ordinal logistic regression is a specific type of regression analysis used when the outcome variable is on an ordinal scale (points on the scale are in order, but the interval between each point is not assumed to be equal). Ordinal logistic regression is an extension of binary logistic regression, which models the likelihood of a specified outcome occurring from a dichotomous variable, given a set of independent predictor variables (O'Connell, 2006).

In logistic regression models, the effect of each independent predictor is often expressed in terms of odds ratios (ORs). For categorical variables, ORs represent the comparative likelihood that participants in a particular group of an independent predictor (e.g. male respondents) will have the target outcome, compared to participants in a specified reference category (e.g. female respondents). An OR of one indicates that participants from both groups are equally likely to have the target outcome, an OR of less than one indicates that participants in a particular group are less likely to have the target outcome compared to the reference group, and an OR of greater than one indicates that participants in a particular group are more likely to have the target outcome.

For continuous predictor variables, ORs reflect the change in the odds of participants having the target outcome for each additional unit of the predictor (e.g. each additional contact hour).

# Analysis and Results

## Regression Analysis

Ordinal logistic regression was conducted on W2 expectations of applying to HE, with contact hours, HE potential, gender, W1 expectation, W1 motivation to do well in studies, and school entered as predictor variables. The associations between predictor variables and W2 expectations of applying to HE are displayed in Table 2. For each additional contact hour participated in, the odds of a student having a higher W2 score (after controlling for the other variables listed) were 1.05 times higher. Based on this model, a student with six contact hours would be 1.34 times more likely to have a higher W2 score compared to students with zero contact hours. Boys were substantially less likely than girls to have higher W2 expectations of applying to HE, and both Green and Non-UCP students had higher expectations than Red students.

**Table 2.** Model parameter estimates.

Variable	ORs
Contact Hours	1.05
HE Potential: Green <sup>a</sup>	1.50
HE Potential: Non-UCP <sup>a</sup>	1.70
Male <sup>b</sup>	0.60
W1 – Definitely won't apply <sup>c</sup>	0.27
W1 - Very unlikely <sup>c</sup>	0.32
W1 - Fairly unlikely <sup>c</sup>	1.34
W1 - Fairly likely <sup>c</sup>	2.10
W1 - Very likely <sup>c</sup>	4.94
W1 - Definitely will apply <sup>c</sup>	21.88
Motivation (W1)	1.12

<sup>a</sup> Reference category - HE Potential: Red

<sup>b</sup> Reference category – Female

<sup>c</sup> Reference category – 'Don't know' W1 responses

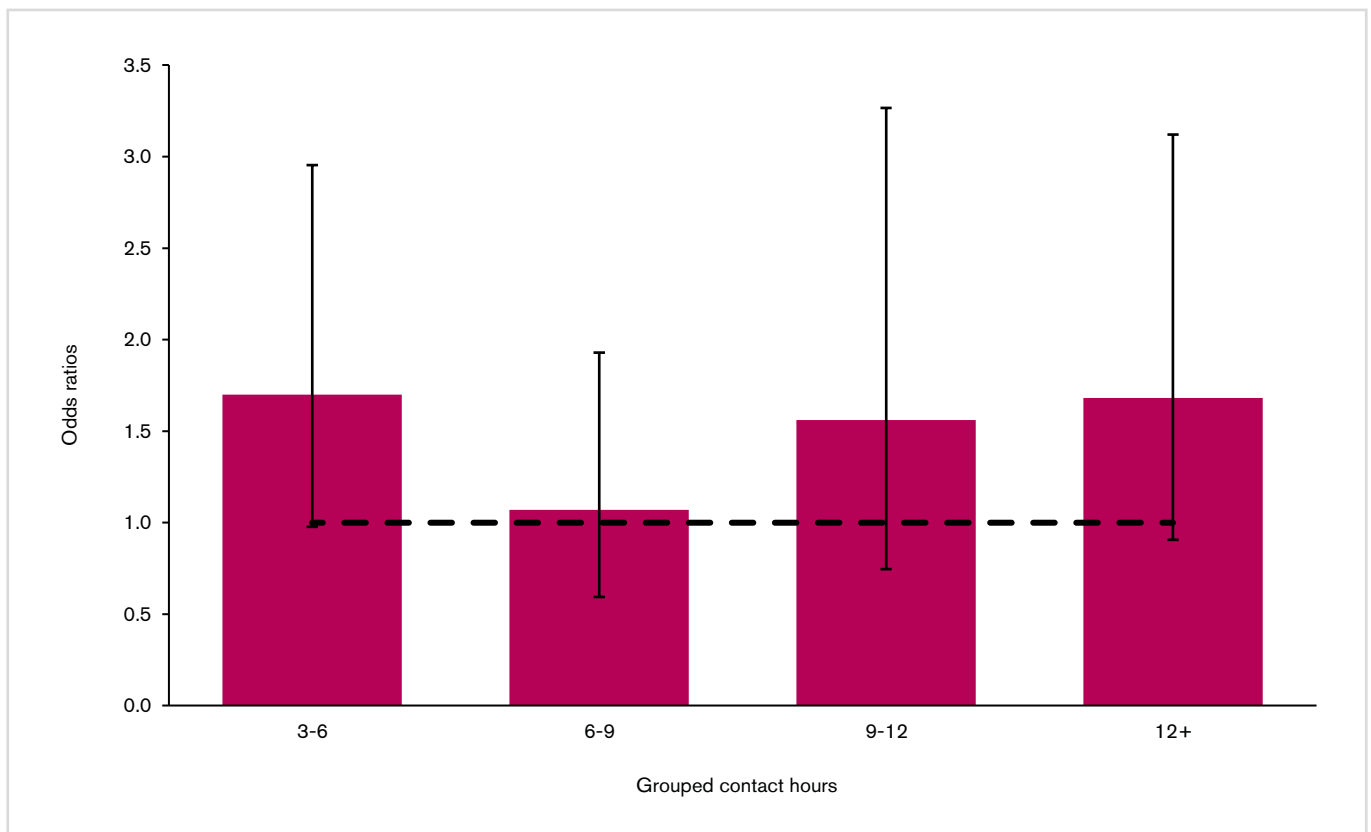


# Exploratory Analysis



Additional analysis was conducted using contact hour groupings to determine whether a certain range was particularly effective (or ineffective). A grouped contact hours variable was created, categorising hours into five groups: 0-3 hours, 3.01-6 hours, 6.01-9 hours, 9.01-12 hours, and 12.01+ hours. These categories were chosen on the basis that they provided equal ranges of contact hours, and acceptable sample sizes. The odds ratios are displayed in Figure 1.

**Figure 1.** Odds ratios by contact hours group.



Reference category = 0-3 contact hours

Error bars reflect 95% confidence intervals

Dashed line reflects even odds

Students with 3-6 contact hours were 1.7 times more likely to have a higher W2 response relative to the reference group (less than three hours). Higher contact hour categories (6-9, 9-12, 12+) were not associated with greater relative ORs than the 3-6-hour category.



# Conclusions

For students in Year 10 during the first wave of the survey, we found a moderate, positive association between contact hours and Wave 2 expectations of applying to HE, after controlling for Wave 1 expectations (in addition to W1 motivation, HE potential, gender, and school).

This suggests that the more a Year 10 student interacts with the HeppSY programme, the greater the odds are that they will maintain or increase their expectations of progression to HE study at age 18 or 19 – a significant focus of the programme as whole.

Further, the exploratory analysis sought to identify a 'saturation point' at which further interaction with the programme offered little to no benefit against the outcome of interest. The exploratory model highlighted that students with 3-6 contact hours were 1.7 times more likely to have higher expectations of applying to HE than students with fewer than three hours, and contact hours greater than six were not associated with a greater relative likelihood. Students who receive in excess of six hours of outreach may therefore accrue no further benefit on the outcome of focus (expectation of HE progression at age 18 or 19), and these incremental interactions are a potential drain on programme resource which could be better used to support engagement of learners with less than three hours of total contact.



## Limitations

An important limitation of this analysis is that the sample was not random. Schools more closely engaged with HeppSY may have been more likely to facilitate student participation in the survey across both waves, and details on how schools select students to complete the survey are not always known. Ideally, all students would be given the opportunity to respond, but it is possible that schools selected specific classes to participate based on convenience or other factors.

Similarly, student response bias is a potential issue. Students who completed the survey across both waves may have differed in important ways from those that had the opportunity to participate but chose not to. For example, perhaps students who felt that HeppSY outreach had helped them to reconsider their HE options were more inclined to complete Wave 2 of the survey than students who had participated in HeppSY outreach but did not find it to be valuable. Issues with sample selection and response bias are highlighted by the disproportionate number of girls in the sample relative to boys.

An additional limitation is that attainment data was not included in the model. Attainment data on individual students from the National Pupil Database (NPD) is not currently available through HEAT for UCP consortia. While the HE potential coding given to students by schools was included, this is not a precise or objective measure, with different schools likely to use varying criteria in establishing the coding. Not including an objective measure of prior educational achievement such as KS2 or KS3 attainment is an important confound that precludes any causal claims being made from the data.

# Recommendations

## Recommendations for HeppSY Practice



Given the findings highlighted above, two recommendations are made for future HeppSY practice. Firstly, alongside the programme aim of having over 20% of target learners engage in at least two interventions per year, these interventions should endeavour to offer at least three hours of total contact time. Secondly, school register data should be used to gradually deprioritise students who have already received over six hours of contact within the academic year, unless there is a compelling case for further specific intervention (e.g. participation in a headline programme, or diversifying the profile of activity that they have undertaken). Minimising delivery of outreach activity beyond six hours per student would reduce a source of programme 'leakage' (Harrison, 2012), where additional intervention appears to be offering minimal benefit on expectations of HE progression. However, it should be noted that the 3-6-hour window was based on exploratory analysis and should be taken as a guide rather than a precise requirement.

## Recommendations for Future Research and Evaluation

The matched analysis approach used here can be extended to different outcome measures drawn from the CFE annual learner survey, such as HE knowledge and career knowledge. This cohort-level analysis could aim to evaluate the association between changes in these self-report measures and participation in specific types of activity, rather than a simple measure of total contact hours. This level of analysis may be possible where multiple year groups are included, therefore providing a larger sample size.

Additionally, access to KS2 and/or KS3 attainment data at a student level would provide an opportunity for HeppSY and other UCP consortia to substantially improve upon the strength of evidence they are able to establish. If these data become available to UCP consortia, then we would revisit our matched analysis to incorporate this important variable and also include it in future work using objective HE outcomes through the HESA track data.

---

# References

---

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

Harrison, N., (2012). The mismeasure of participation: how choosing the 'wrong' statistic helped seal the fate of Aimhigher. *Higher Education Review*, 45(1).

HeppSY, (2020). 2019/2020 HeppSY Partnership Report [online]. Available from: <https://www.heppsy.org/assets/2020/05/HeppSY-Partnership-Report-2019-20-final.pdf>

O'Connell, A.A., 2006. *Logistic regression models for ordinal response variables* (Vol. 146). Sage.



To find out more please contact:

**Mike Garnock-Jones**

Director, Hepp

[M.Garnock-Jones@shu.ac.uk](mailto:M.Garnock-Jones@shu.ac.uk)

**Greg Brown**

Evaluation and Data Manager, HeppSY

[G.Brown@shu.ac.uk](mailto:G.Brown@shu.ac.uk)

**Sharon Woodward-Baker**

HeppSY Programme Manager

[S.Woodward-Baker@shu.ac.uk](mailto:S.Woodward-Baker@shu.ac.uk)

