



HeppSY Outreach Interventions' Long-term Outcome

A Regression Analysis of Contact Hours' Contribution to HE Progression

August 2023

Author: Daniel Fletcher
Produced by the HeppSY Data & Evaluation Team

Contents

3 Introduction

5 Method

5 Sample

5 Data Sources

5 Variables

7 Results

7 Descriptive Statistics

10 Regression Analysis

20 Summary

20 Limitations

22 References

Introduction

The Higher Education Progression Partnership South Yorkshire (HeppSY) is part of the Uni Connect Programme (UCP), funded by the Office for Students. The main focus of Uni Connect is to provide targeted higher education (HE) outreach to young people in Years 9 to 13 living in particular geographic areas. From August 2021 this broadened out to include the targeting of adult learners (learners aged 19 and above). HeppSY is working in partnership with Sheffield Hallam University, The University of Sheffield and South Yorkshire schools and colleges.

Previous analyses matching together multiple waves of the annual student survey (designed and hosted by CFE Research) with HeppSY activity data found a positive association between level of engagement in HeppSY outreach and increases in students' self-reported HE knowledge – particularly for secondary school students (HeppSY 2020b) and students with low baseline HE knowledge (HeppSY, 2022). Further analysis also indicated that greater engagement in HeppSY activity was associated with higher expectations of applying to HE among secondary school students, after controlling for prior expectations (HeppSY, 2020a) (though evidence for this is mixed; HeppSY, 2022).

However, to date, no HeppSY impact evaluation has been conducted using progression to HE as an outcome variable. By bringing together multiple sources of student data, the aim of the present report was to examine the relationship between engagement in HeppSY activity and progression to HE. Specifically, this analysis used a dataset made available by Higher Education Access Tracker (HEAT), which matched students who had participated in HeppSY activity to the Higher Education Statistics Agency (HESA) database - containing a record of students who have registered with a HE provider. This HEAT-HESA matched dataset therefore included, at an individual student level, measures of HeppSY programme engagement and an indicator of HE access.

It was hypothesised that level of engagement in HeppSY outreach (quantified by number of contact hours a student had participated in) would be positively associated with the odds of progression to HE aged 18. Since an association between contact hours and HE progression could arise if students who were already more likely to access HE were selected by schools and colleges (or self-selected) to participate in more outreach, analysis was also conducted on a subset of the original sample who had completed the CFE baseline¹ survey in 2017-18. This allowed the relationship between HeppSY engagement and odds of progressing to HE to be examined after controlling for students' pre-intervention motivation and expectations of applying to HE. Additionally, as previous research has indicated the relationship between outreach engagement and intention to apply to HE and HE access

.....

1 Baseline survey responses were collected throughout academic year 2017-18, and some students may therefore have participated in HeppSY activity prior to completing the survey (see Limitations section).

may be non-linear (HeppSY, 2020a; Aimhigher West Midlands, 2020), with the incremental benefit of additional engagement diminishing at higher levels of overall engagement, we examined possible non-linear effects of HeppSY engagement.

Method

Sample

9328 students were in the HEAT-HESA track dataset. A subset of students in the HEAT-HESA dataset were also matched to a baseline survey response ($n = 2590$). See results section for details of sample characteristics and exclusions.

Data Sources

HEAT-HESA Track Dataset

The HEAT-HESA track dataset consisted of students who were assigned to the HeppSY HEAT database between academic year 2017-18 and 2019-20, and who were aged 18 or older prior to the beginning of academic year 2020-21². The dataset included demographic information about students (e.g., sex, UCP status, IMD Quintile), the HeppSY outreach they had participated in (e.g., activity type, contact hours), and whether they had been found in the HESA database (and therefore progressed to HE).

CFE Baseline Survey

The student survey data used here were from the CFE baseline learner survey. Survey responses were collected from students in Years 9 to 13 in HeppSY partner schools and colleges throughout academic year 2017-18. HeppSY co-ordinated with key points of contact in partner institutions, who administered the survey to students on HeppSY's behalf.

Variables

HEAT-HESA Measures:

UCP Status: Variable reflecting whether or not a student's home postcode was within the Uni Connect geographical target region.

.....

- 2 Note that the HESA student population database is based on registered students at HE providers, meaning students who defer their place will not appear in the dataset until the academic year they commence their course.

Index of Multiple Deprivation (IMD) Quintile: Derived from multiple measures (e.g., income, employment, crime rates) used to classify geographical regions in England into the most and least deprived areas. Quintile 1 represents the 20% most deprived areas and Quintile 5 represents the 20% least deprived areas.

HE Access: Binary measure indicating whether or not a student had accessed HE, according to the HESA database, by the beginning of academic year 2020-21. Students found in the HESA database were assumed to have progressed to HE, whereas students who were not found were assumed to have not (yet) progressed. Data were coded so that 0 = not progressed, 1 = progressed.

Contact Hours: Measure derived from the HEAT database, reflecting the total number of hours of HeppSY outreach a student had participated in between the start of the Uni Connect programme in 2017-18 and the end of the academic year 2019-20.

Survey Measures:

Expectations of Applying to Higher Education: Students were asked, "How likely are you to apply to higher education aged 18 or 19?", with responses provided on a six-point ordered scale (1 = definitely won't apply; 6 = definitely will apply), with a "don't know" option also available.

Motivation: Students were asked, "I am motivated to do well in my studies", on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree), with a "don't know" response option also available.

Know Someone at HE: Students were asked to indicate whether they "know anyone that has gone on to HE". Responses coded as a binary variable (know someone that has gone to HE; do not know anyone that has gone to HE).

Results

Descriptive Statistics

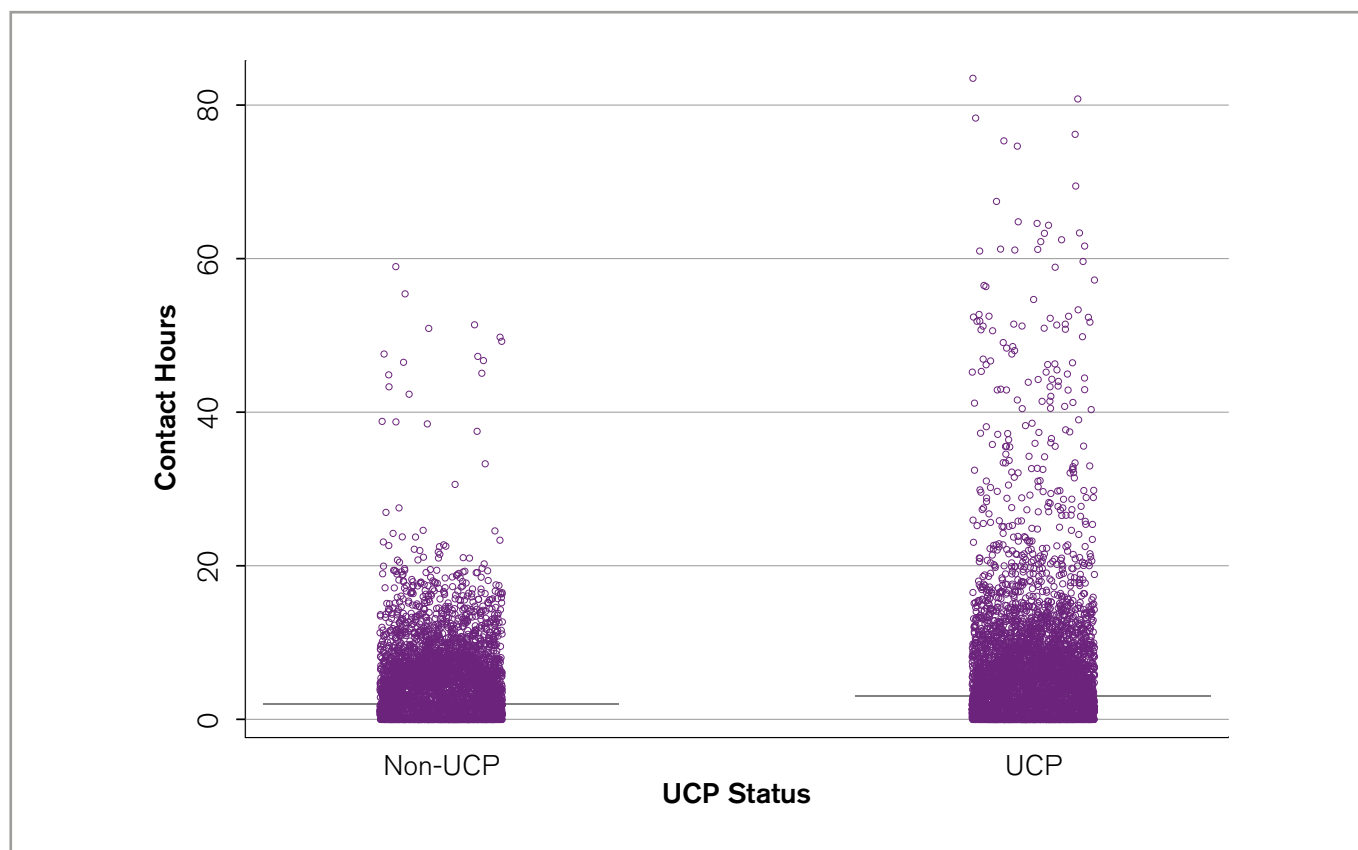
As the primary focus of the Uni Connect programme prior to August 2021 was on learners between the ages of 13-19, students aged 20 or over when first participating in HeppSY activity were excluded from the analysis (though note that including these students did not alter the significance of the results). After exclusions, there were 8,991 students in the HEAT-HESA dataset ([see Table 1](#)).

Table 1. *Sample Characteristics*

| UCP Status | Male | Female | Unknown | Total |
|------------|------|--------|---------|-------|
| UCP | 2041 | 2638 | 71 | 4750 |
| Non-UCP | 1831 | 2326 | 84 | 4241 |
| Total | 3872 | 4964 | 155 | 8991 |

Contact hours per student ranged from 0-83 with a median of 2.4 (number of activities ranged from 0-45, with a median of 2.0). Median contact hours were higher for UCP students (3.0) than Non-UCP students (2.0) ([Figure 1](#)).

Figure 1. Student contact hours by UCP status



Note. Dots represent individual data points. Horizontal lines reflect median contact hours by group.

HE Access

Overall, 40% of students in this sample had accessed HE by academic year 2020-21. This figure seems particularly high, given that 81% of students in this sample were from a POLAR4 Quintile 1 or 2 postcode. Indeed, 38% of the POLAR4 Quintile 1 students in the present sample had accessed HE, compared to an access rate of only 24% nationally for 18-year-olds living in POLAR4 Quintile 1 regions (UCAS, 2022). One possible reason for this discrepancy is that over 50% of the current sample were already aged 17 or older when first participating in HeppSY activity. Since HeppSY only aims to target post-16 learners who are studying for Level 3 qualifications (i.e., A-Levels or equivalent), post-16 students in the present sample would be disproportionately likely to access HE relative to the average for this age group. The HE access rate for POLAR4 Quintile 1 students who were aged under 16³ when they first participated in HeppSY activity ($n = 715$) was more in line with expectations, at 27%.

.....

³ Under 16 was chosen as a comparison rather than under 17 since 16-year-old students may be in either Year 11 or Year 12.

As shown in Figure 2, at a descriptive level, HE access rates in this sample appeared to increase linearly as a function of the number of contact hours that students participated in.

Figure 2. Percentage of students accessing higher education by number of HeppSY contact hours participated in

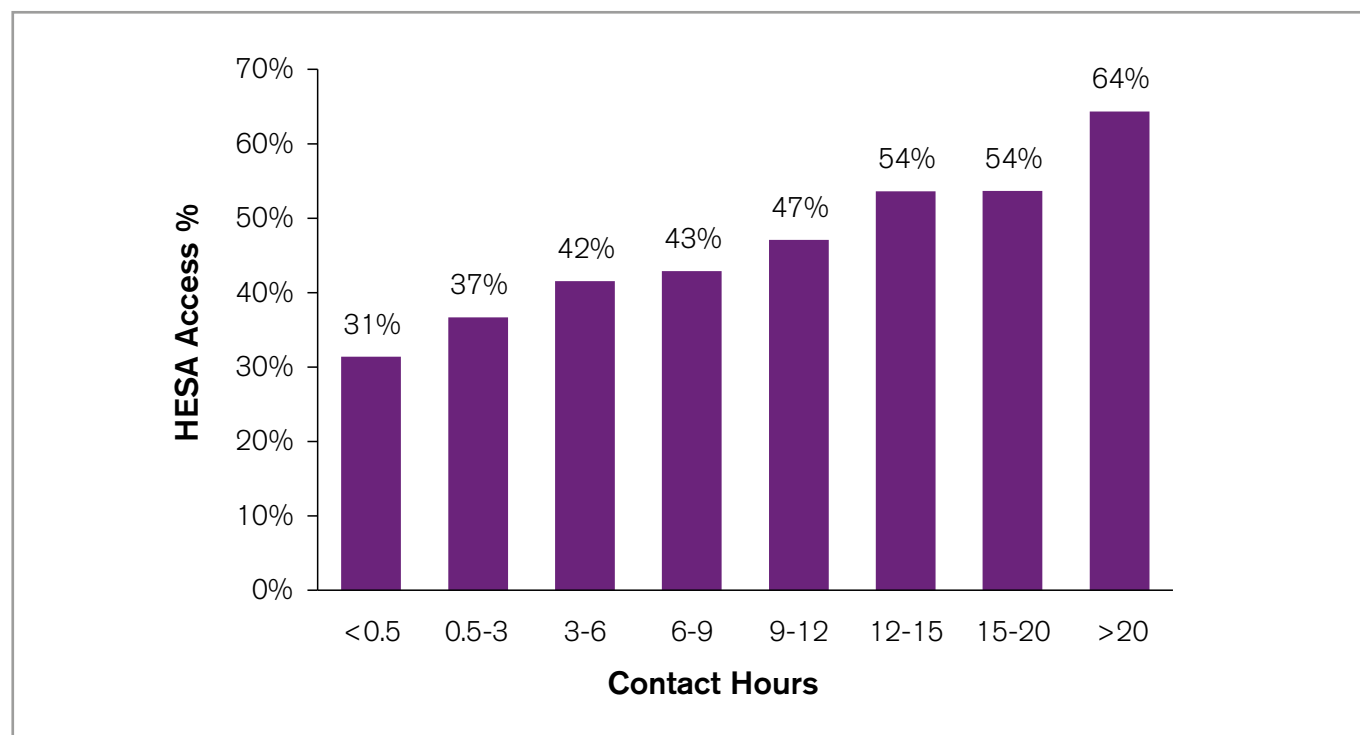
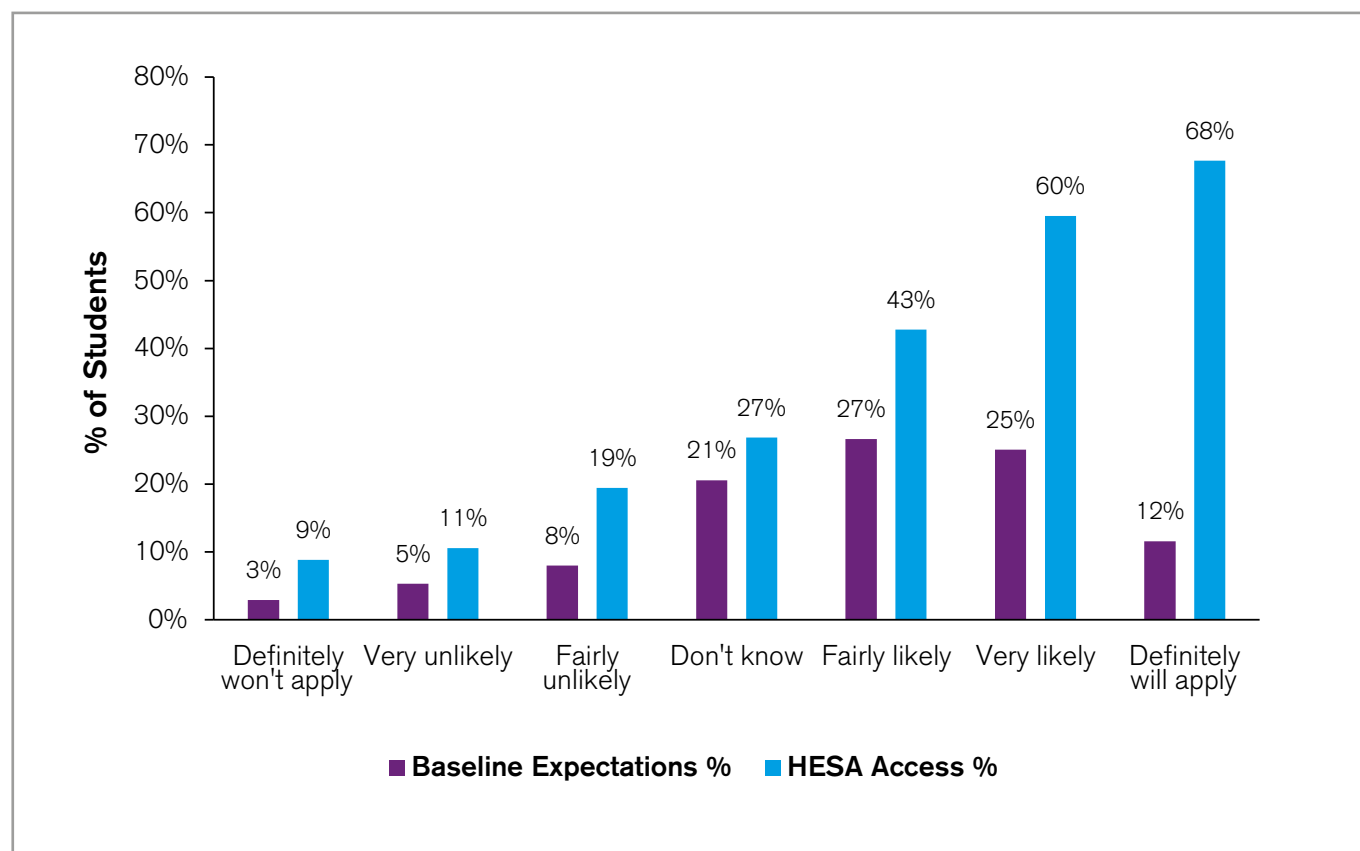


Figure 3 demonstrates the relatively high pre-intervention expectations of applying to HE (64% of students fairly likely, very likely, or definitely likely to apply), based on the subset of students matched to a baseline survey response. Strength of intention to apply to HE at baseline was important to eventual HE access, with 43% of students who reported that they were fairly likely to apply accessing HE, compared to 60% and 68% of students who reported they were very likely, or definitely would apply, respectively.

Figure 3. HE progression rates by baseline expectations of applying to HE



Note. Baseline expectations refers to the percentage of total students providing each response to the expectations of applying to HE question in the baseline survey. HESA access refers to the percentage of students who had accessed HE according to the HESA database, split by their expectations of applying to HE.
n = 2321.

Regression Analysis

Multilevel binary logistic regression analyses were conducted to determine the association between contact hours and odds of HE progression, after controlling for other theoretically meaningful variables. Analysis was first conducted on the total sample, and then the subset of students who could be matched to the baseline survey.

Total Sample

As data were structured hierarchically, with students nested within schools, an intercept-only multilevel binary logistic model utilising the full sample (8974⁴ students across 57 schools/colleges) was fitted to evaluate the requirement for a multilevel model. The model indicated that school/college accounted for 14% of the variance in HE progression outcomes (ICC = .14, 95% CI [.09, .21]). Given the substantial variance accounted for by school/college, the remainder of the analyses proceeded with a multilevel approach.

Student demographic characteristics (sex, UCP status, ethnicity, IMD quintile) were first added to the model, which significantly improved model fit relative to an intercept-only model, $\chi^2(23) = 175.00, p < .001$. Male students were less likely than female students to progress to HE, and several ethnic groups (Arab; Chinese; Asian/Asian British – Pakistani; Asian/Asian British – Indian; Black or Black British – African, Other Mixed Background) had significantly greater odds of HE progression than White students, and none had significantly lower odds. Students in a higher IMD quintile were more likely to progress to HE, though UCP status was not significantly associated with HE progression.

The contact hours variable was subsequently added to the model, which had a significant positive association with odds of HE progression. Although descriptive analysis ([see Figure 2](#)) appeared to indicate a linear relationship between contact hours and HE access (when students with over 20 contact hours were collapsed into a single group), given previous research indicating a diminishing incremental benefit of activity at higher levels of engagement (HeppSY, 2020a; Aimhigher West Midlands, 2020), a quadratic polynomial term for contact hours was added to the model. This quadratic term had a significant, negative coefficient, indicating a reduction in the incremental effect of each additional contact hour on odds of HE progression as the total number of contact hours increased ([Figure 4](#)). See Table 2 for a summary of model coefficients for the final model.

Table 2. Model coefficients for multilevel binary logistic regression on HE progression outcomes

| Variable | Log Odds | SE | p |
|----------------------------|----------|------|-------|
| Sex (ref = female) | | | |
| Male | -0.19 | 0.05 | <.001 |
| Unknown | -0.04 | 0.22 | .84 |
| UCP (ref = Non-UCP) | -0.10 | 0.05 | .07 |

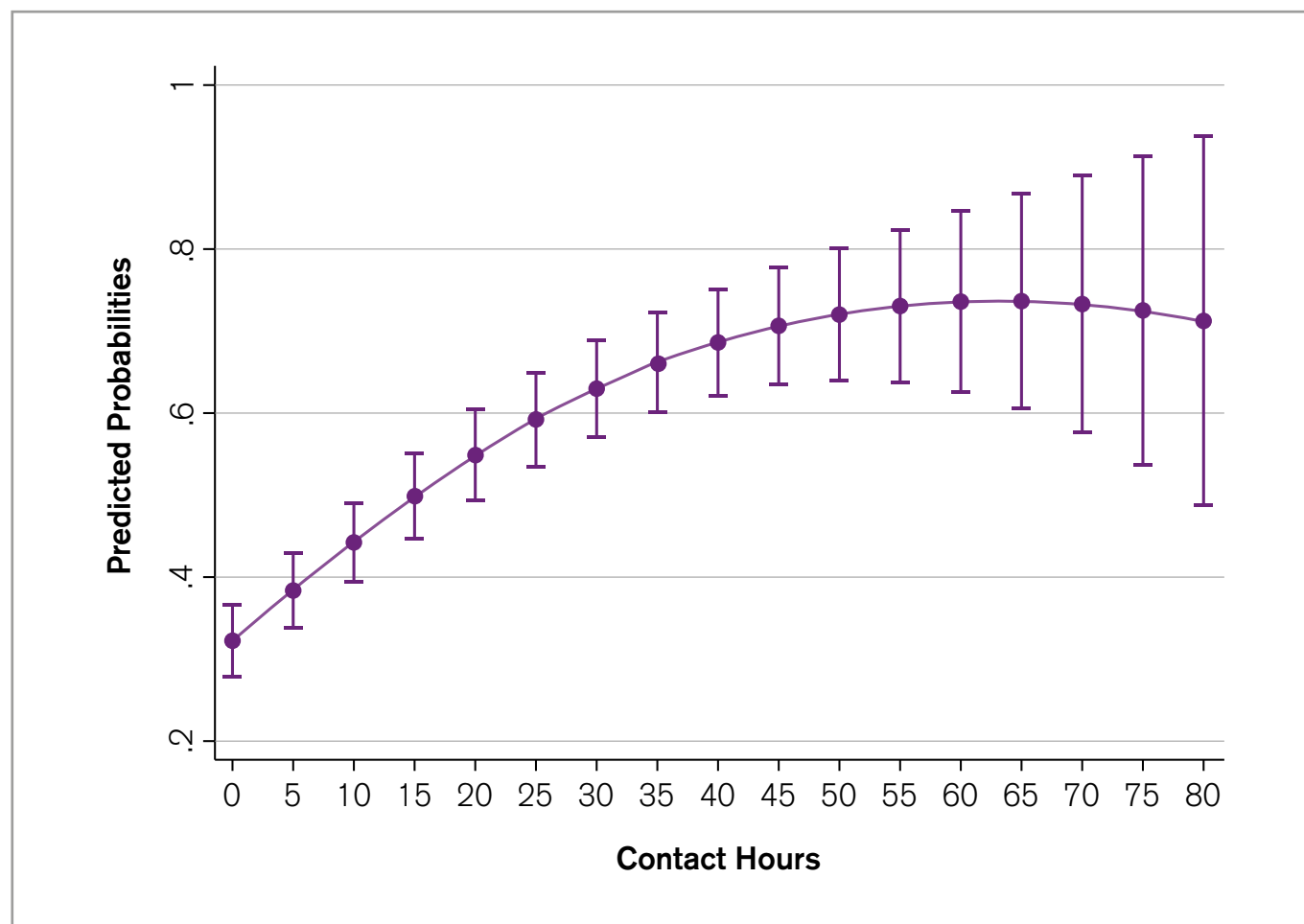
.....

4 17 students with missing IMD Quintile data were excluded, which explains the discrepancy from the total sample of 8991.

| Variable | Log Odds | SE | p |
|---------------------------------|----------|--------|-------|
| IMD Quintile (ref = 1) | | | |
| 2 | 0.18 | 0.06 | .003 |
| 3 | 0.40 | 0.07 | <.001 |
| 4 | 0.47 | 0.08 | <.001 |
| 5 | 0.82 | 0.11 | <.001 |
| Ethnicity (ref = White) | | | |
| Arab | 0.85 | 0.28 | .002 |
| Asian - Bangladeshi | 0.92 | 0.49 | .06 |
| Asian - Indian | 0.88 | 0.30 | .004 |
| Asian – Pakistani | 0.47 | 0.11 | <.001 |
| Black – African | 0.84 | 0.16 | <.001 |
| Black – Caribbean | 0.59 | 0.42 | .17 |
| Chinese | 1.12 | 0.51 | .03 |
| Gypsy/Traveller | -0.96 | 0.63 | .13 |
| Mixed – White & Asian | 0.44 | 0.26 | .09 |
| Mixed – White & Black African | 0.42 | 0.39 | .25 |
| Mixed – White & Black Caribbean | 0.17 | 0.22 | .45 |
| Other Asian Background | 0.43 | 0.25 | .08 |
| Other Black Background | -0.43 | 0.42 | .31 |
| Other Ethnic Background | 0.45 | 0.25 | .07 |
| Other Mixed Background | 0.71 | 0.31 | .02 |
| Not known | -0.09 | 0.15 | .47 |
| Contact Hours | 0.05 | 0.01 | <.001 |
| Contact Hours-Quadratic | -0.0005 | 0.0002 | .003 |

Note. SE = standard error of log odds. Ref = reference category. Contact hours was mean-centred.

Figure 4. Predicted probability of HE access by number of contact hours received



Matched to Survey Data

A separate regression was conducted using the subset of students who were in the HESA track dataset and matched to the CFE baseline survey (see [Table 3](#) for sample characteristics). 2590 students were matched to a baseline survey response, though 288 of these students were excluded from the analysis due to missing data on at least one of the following variables: expectations of applying to HE, motivation to do well in studies, know someone at HE - leaving a total of 2302 students. Around half of these exclusions ($n = 147$) were due to survey logic, where Year 13/Level 3 Year 2 students who indicated that they had already applied to HE were not subsequently asked how likely they were to apply. This means that in the following analysis, Year 13/Level 3 Year 2 students were only included if they had not already applied to HE when surveyed (this is appropriate, given that the programme is unlikely to materially impact HE access for students who had already applied at baseline).

Table 3. Sample characteristics for students matched to baseline survey responses

| Year Group | UCP | | | Non-UCP | | | Total |
|--------------------|------|--------|---------|---------|--------|---------|-------|
| | Male | Female | Unknown | Male | Female | Unknown | |
| Year 11 | 226 | 356 | 7 | 77 | 110 | 3 | 779 |
| Year 12 | 132 | 189 | 0 | 91 | 156 | 0 | 568 |
| Year 13 | 19 | 34 | 0 | 4 | 8 | 2 | 67 |
| Level 3: Y1 | 146 | 211 | 0 | 172 | 239 | 7 | 775 |
| Level 3: Y2 | 32 | 36 | 0 | 10 | 15 | 1 | 94 |
| Level 2 | 5 | 3 | 0 | 5 | 6 | 0 | 19 |

Demographic characteristics (sex, UCP status, IMD quintile) and year group were first added to the model (the ethnicity variable was not included in this analysis due to small group sizes), which significantly improved model fit relative to an intercept-only model, $\chi^2(12) = 91.36, p < .001$. In this model, the odds of male students progressing to HE did not significantly differ to female students, and UCP status was also not significantly associated with odds of HE progression. Relative to students who were in Year 11 at baseline, students in Year 12 were significantly more likely to progress to HE, though odds of HE access did not significantly differ between Year 11 and any other year group. Higher IMD quintiles were again associated with greater odds of accessing HE.

Next, survey response variables (likelihood of applying to HE, motivation to do well in studies, know someone at HE) were added, which significantly improved model fit, $\chi^2(12) = 298.38, p < .001$. Relative to students who responded "don't know" to the likelihood of applying to HE question in the baseline survey, students who reported that they definitely would not apply or were very unlikely to apply were significantly less likely to progress to HE, and students who reported that they were fairly likely, very likely, or definitely would apply were significantly more likely to progress to HE. For the study motivation question, there were no significant differences in the odds of HE progression for students who responded "don't know" relative to any other response. Students who knew someone at HE were also not significantly more likely to progress to HE than those who did not know someone at HE.

The contact hours variable was then added, which again was associated with significantly greater odds of HE progression. The interaction between year group and contact hours was also then added to examine whether the effect of contact hours differed based on which year group students were in when the HeppSY outreach programme started. Including the interaction significantly improved model fit, $(5) = 12.61, p = .03$, with contact hours more positively associated with odds of HE progression among students in Year 11 at baseline relative to those in Year 12 or Level 3 Year 1. Finally, a quadratic polynomial term for contact hours was added, which again had a significant negative coefficient. See [Table 4](#) for final model coefficients.

Table 4. Model coefficients for multilevel binary logistic regression on HE progression outcomes, using students matched to the baseline survey

| Variable | Log Odds | SE | p |
|--|----------|------|-------|
| Sex (ref = female) | | | |
| Male | 0.14 | 0.10 | .10 |
| Unknown | -1.41 | 0.74 | .06 |
| UCP (ref = Non-UCP) | -0.16 | 0.12 | .17 |
| IMD Quintile (ref = 1) | | | |
| 2 | 0.09 | 0.13 | .49 |
| 3 | 0.50 | 0.15 | .001 |
| 4 | 0.35 | 0.19 | .06 |
| 5 | 0.94 | 0.26 | <.001 |
| Baseline Year Group (ref = Year 11) | | | |
| Year 12 | 0.59 | 0.16 | <.001 |
| Year 13 | 0.22 | 0.76 | .77 |
| Level 3: Year 1 | 0.05 | 0.17 | .78 |
| Level 3: Year 2 | 0.07 | 0.40 | .87 |
| Level 2 | 0.91 | 1.21 | .46 |
| Likelihood of applying (ref = don't know) | | | |
| Definitely won't apply | -1.28 | 0.47 | .01 |
| Very unlikely | -1.14 | 0.35 | .001 |
| Fairly unlikely | -0.37 | 0.25 | .14 |
| Fairly likely | 0.74 | 0.18 | <.001 |
| Very likely | 1.33 | 0.18 | <.001 |
| Definitely will apply | 1.66 | 0.18 | <.001 |
| Study Motivation (ref = don't know) | | | |
| Strongly disagree | -0.03 | 0.65 | .96 |

| Variable | Log Odds | SE | <i>p</i> |
|--|----------|--------|----------|
| Disagree | 0.24 | 0.77 | .76 |
| Neither agree nor disagree | -0.61 | 0.58 | .29 |
| Agree | -0.25 | 0.53 | .64 |
| Strongly agree | -0.14 | 0.53 | .79 |
| Know Someone at HE (ref = yes) | -0.10 | 0.14 | .50 |
| Contact Hours | 0.08 | 0.01 | <.001 |
| Contact Hours-Quadratic | -0.001 | 0.0003 | .02 |
| Contact Hours x Baseline Year Group (ref = Year 11) | | | |
| Year 12 | -0.03 | 0.02 | .04 |
| Year 13 | -0.02 | 0.09 | .81 |
| Level 3: Year 1 | -0.03 | 0.02 | .05 |
| Level 3: Year 2 | -0.04 | 0.04 | .34 |
| Level 2 | 0.26 | 0.20 | .18 |

Note. SE = standard error of log odds. Ref = reference category. Contact hours was mean-centred.

Contact Hours in Academic Year 2017-18

There is a potential confound in the data presented so far, since HeppSY only continued to target students for outreach after Year 11 if they progressed to a Level 3 course. This means that among students in Year 11 at baseline, those that progressed to a Level 3 course would have been in the target population for HeppSY outreach across three academic years (Year 11 and two post-16 years), while those that did not progress would only have been in the target population for one academic year. To illustrate the issue this causes, imagine a hypothetical situation in which all students received four contact hours for each academic year they were in the HeppSY target population, and that the programme had no actual effect on HE access. Students in Year 11 at baseline would all receive four contact hours in 2017-18, regardless of whether they ultimately progressed to a Level 3 course. Students that progressed to a Level 3 course would then receive an additional four hours in each subsequent post-16 year, making a total of 12 contact hours across the three-year period. In contrast, students that did not progress would receive zero hours in subsequent academic years, making a total of four contact hours across the three-year period. Since students that progress to a Level 3 course

straight from Year 11 are clearly more likely to access HE by age 18 or 19 than students that do not progress, analysis may reveal a spurious association between contact hours and HE access – driven by a form of survivorship bias.

To examine whether an association between contact hours and HE access could be established while accounting for this confound, a separate analysis was conducted using contact hours delivered in academic year 2017-18 only as the predictor variable. Since all students in the sample were eligible to receive HeppSY outreach in academic year 2017-18, a positive association between 2017-18 contact hours and HE access could not be caused by differing length of HeppSY outreach eligibility depending on whether students in Year 11 at baseline progressed to a Level 3 course or not. Note that an effect of 2017-18 contact hours on eventual HE access among students who were in Year 11 at baseline, and therefore only eligible to access HE several years later, is plausible since HeppSY activity may influence the likelihood that students progress to a Level 3 course and therefore remain on the pathway to HE. Indeed, students' attitudes towards HE may be particularly malleable when they are approaching key transition points, such as the end of secondary school (The Sutton Trust, 2008).

Using 2017-18 contact hours as a predictor variable, the association between contact hours and odds of HE progression was attenuated but still significant (Table 5), suggesting that survivorship bias could not entirely account for observed relationship.

Table 5. Model coefficients for multilevel binary logistic regression on HE progression outcomes, based on academic year 2017-18 contact hours

| Variable | Log Odds | SE | p |
|--|----------|------|------|
| Sex (ref = female) | | | |
| Male | 0.11 | 0.10 | .27 |
| Unknown | -1.54 | 0.69 | .03 |
| UCP (ref = Non-UCP) | -0.12 | 0.11 | .31 |
| IMD Quintile (ref = 1) | | | |
| 2 | 0.09 | 0.13 | .46 |
| 3 | 0.51 | 0.15 | .001 |
| 4 | 0.37 | 0.18 | .05 |
| 5 | 0.86 | 0.26 | .001 |
| Baseline Year Group (ref = Year 11) | | | |
| Year 12 | 0.48 | 0.15 | .002 |

| Variable | Log Odds | SE | p |
|--|----------|------|-------|
| Year 13 | -0.15 | 0.30 | .62 |
| Level 3: Year 1 | -0.02 | 0.15 | .92 |
| Level 3: Year 2 | -0.02 | 0.29 | .95 |
| Level 2 | | | |
| Likelihood of applying (ref = don't know) | | | |
| Definitely won't apply | -1.30 | 0.47 | .01 |
| Very unlikely | -1.17 | 0.34 | .001 |
| Fairly unlikely | -0.38 | 0.25 | .12 |
| Fairly likely | 0.74 | 0.18 | <.001 |
| Very likely | 1.33 | 0.18 | <.001 |
| Definitely will apply | 1.66 | 0.18 | <.001 |
| Study Motivation (ref = don't know) | | | |
| Strongly disagree | 0.02 | 0.64 | .97 |
| Disagree | 0.32 | 0.76 | .67 |
| Neither agree nor disagree | -0.58 | 0.58 | .31 |
| Agree | -0.18 | 0.53 | .73 |
| Strongly agree | -0.08 | 0.53 | .87 |
| Know Someone at HE (ref = yes) | -0.07 | 0.14 | .64 |
| Contact Hours (2017-18) | 0.02 | 0.01 | .01 |

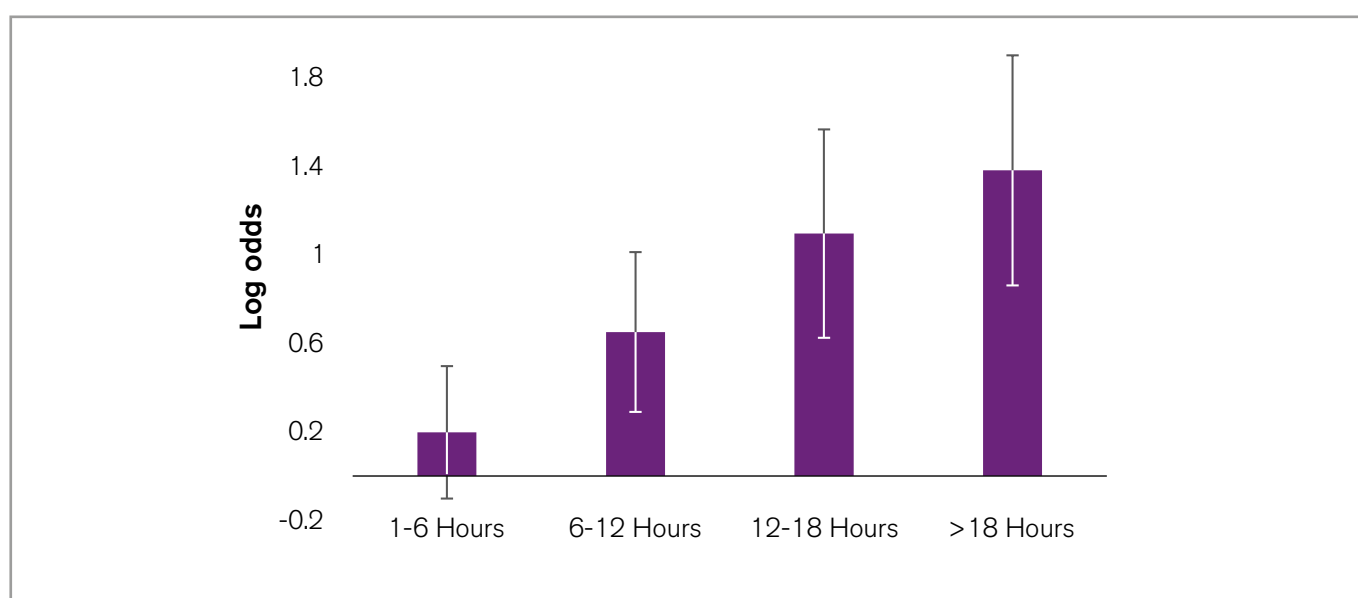
Note. The interaction between contact hours (2017-18) and year group, $\chi^2[5] = 5.96$, $p = .31$), and the quadratic effect of contact hours (2017-18) ($\chi^2[1] = 0.21$, $p = .64$) were both non-significant in this case, and therefore omitted from the model.

Exploratory Analysis: Optimal Contact Hours

Given the evidence for a non-linear effect of contact hours after controlling for baseline survey responses ([see Table 4](#)), exploratory analysis was conducted to provide an indication of what range of contact hours may be most effective. The contact hours variable was split into groups (less than 1 hour

[reference category], 1-6 hours, 6-12 hours, 12-18 hours, over 18 hours), and entered into a multilevel binary logistic regression on HE access, alongside sex, UCP status, IMD quintile, baseline year group, baseline likelihood of applying to HE, study motivation, and know someone at HE. Relative to students that received less than one hour of contact, students that received 6-12, 12-18 or over 18 hours of contact, but not students that received 1-6 hours, had significantly higher odds of progressing to HE. However, there was no significant difference in the odds of HE progression for students receiving over 18 hours relative to students receiving 12-18 hours ($b = 0.29$, $SE = 0.28$, $p = .31$, 95% CI [-0.26, 0.83]) (Figure 5).

Figure 5. Log odds of progression to HE by grouped contact hours variable



Note. Reference category = < 1 hour. Sample sizes per group: < 1 hour = 362, 1-6 = 1182, 6-12 = 450, 12-18 = 164, > 18 = 145. Error bars reflect 95% confidence intervals.

Summary

As hypothesised, and providing support for the effectiveness of HeppSY outreach, this analysis found a positive association between the number of HeppSY contact hours received by students and their odds of progressing to higher education. Importantly, this association remained significant after controlling for students' baseline expectations of applying to HE and study motivation. There was, however, evidence of non-linearity in the effect of contact hours, with a diminishing incremental benefit of each additional contact hour at higher levels of overall contact. Exploratory analysis suggested that 12-18 hours of contact across a three-year period may be beneficial in increasing the likelihood of a student progressing to HE, but there was no clear evidence of any incremental benefit of engagement beyond this point. Additionally, the association between engagement in HeppSY activity and HE access was significantly stronger among Year 11 students relative to Year 12 and Level 3 Year 1 students, which could indicate that there is less potential for outreach to effect HE outcomes among post-16 students (though this particular result could potentially be explained by a survivorship bias – see limitations).

Limitations

There are several limitations of the present report. Firstly, it was not possible to control for students' pre-intervention academic ability or attainment. It is plausible that more academically able students, who are more likely to access HE, were selected by schools to participate in more activity, which could have produced a spurious relationship between contact hours and probability of HE access. However, this issue may have been partially mitigated by controlling for students' baseline expectations of applying to HE and study motivation. Secondly, since the baseline CFE survey was administered throughout academic year 2017-18, while HeppSY activity was ongoing, some students are likely to have participated in the programme prior to providing a "baseline" survey response. However, this seems more likely to attenuate rather than exaggerate the relationship between programme engagement and HE access. This is because positive impacts of HeppSY activity that occurred prior to baseline survey measurement would not be attributed to programme engagement, since the analysis determined the relationship between outreach participation and HE access after statistically controlling for self-reported expectations of applying to HE in the baseline survey. Thirdly, the analysis provided evidence of overall programme impact, but not the particular aspects of the programme that may have been most (and least) effective. Evaluation of specific programmes of activity will be required to establish the effectiveness of particular interventions, grounded in an appropriate theory of change model. Fourthly, as considered in the results section, a form of survivorship bias could have contributed toward the relationship between programme engagement and probability of HE access, whereby Year 11 students who progressed to a Level 3 course (and therefore remained on the pathway to HE) would have been eligible to receive three years of HeppSY activity, whereas those that did not progress to a Level 3 course after Year 11 were likely to only have received one year of

activity. This did not appear to fully account for the observed effect, since higher contact hours within a single academic year (2017-18) was also associated with increased odds of HE access. However, this survivorship bias could have contributed to the observation that the effect of contact hours was relatively stronger among students who were in Year 11 at baseline relative to post-16 students.

References

Aimhigher West Midlands (2020). Does Engagement in Aimhigher Interventions Increase the Likelihood of Disadvantaged Learners Progressing to HE? A mixed methods approach employing a quasi-experimental design and case studies [online]. Available from: <https://aimhigherwm.ac.uk/wp-content/uploads/2020/04/Aimhigher-Mixed-Methods-Impact-Evaluation-Study-2020.pdf>

HeppSY, (2020a). HeppSY Cohort Evaluation: Contact Hours and Expectations of Applying to HE Aged 18 or 19 [online]. Available from: <https://heppsy.org/assets/2021/02/201020-HeppSY-Cohort-Evaluation-Contact-Hours-and-Expectations-of-Applying-to-HE-Aged-18-or-19-Full-Report-v5.pdf>

HeppSY, (2020b). Cohort Evaluation: HE Knowledge Matched Analysis [online]. Available from: <https://heppsy.org/assets/2021/02/201204-HE-Knowledge-Matched-Analysis-Report-Full-Report-v3.pdf>

HeppSY, (2022). Matched Analysis of Student Survey Responses [online]. Available from: <https://heppsy.org/assets/2022/10/221014-Matched-Analysis-of-Student-Survey-Responses-October-2022-v3.pdf>

The Sutton Trust (2008). Increasing higher education participation amongst disadvantaged young people and schools in poor communities [online]. Available from: <https://www.suttontrust.com/our-research/report-national-council-educational-excellence/>

UCAS (2022). UCAS Undergraduate End of Cycle Data Resources 2022 [online]. Available from: <https://www.ucas.com/data-and-analysis/undergraduate-statistics-and-reports/ucas-undergraduate-end-cycle-data-resources-2022>



To find out more please contact:

Gemma Styles
Hepp Director
G.Styles@shu.ac.uk

Vicki Spink
HeppSY Programme Manager
V.Spink@shu.ac.uk

Helen Zhu
Evaluation and Data Manager
C.Zhu@shu.ac.uk

