

# QUALITY MATTERS : THE CENSUS AND HESA WIDENING PARTICIPATION DATA

TEJ NATHWANI, PRINCIPAL RESEARCHER (ECONOMIST)  
JENNY BERMINGHAM, LEAD STATISTICAL ANALYST  
SIOBHAN DONNELLY, STATISTICAL ANALYST  
MARCH 2021

HESA



[Published online 2021-03-16](#)

HESA's Data and Innovation team use Census statistics to assess the quality of HESA's parental education and socio-economic classification data.

## INTRODUCTION

A primary purpose for HESA undertaking research activity, as outlined in our [strategy](#), is to utilise external sources of information to assist with reviewing the quality of our own data. This aligns with the recommendations of the [Office for Statistics Regulation \(OSR\)](#) who highlight that one of the methods by which to evaluate accuracy is to utilise other datasets where possible.

With the 2021 Census taking place in England, Wales and Northern Ireland this month, the Data and Innovation team have undertaken research to explore the potential of Census (2011) data to facilitate a further appraisal of the [parental education](#) and [occupation](#) information that HESA collects.

We hope the findings will be useful to a wide range of our key stakeholders, including widening participation (WP) practitioners within higher education providers and researchers/policymakers who utilise these two variables as part of their analysis of HESA data.

## OVERVIEW

The data we hold on parental education and occupation predominantly originates from the [Universities and Colleges Admissions Service \(UCAS\) application](#) that prospective students are asked to submit as part of the entry process into higher education. Individuals are requested to confirm whether either of their (step-) parents or guardians hold a higher education qualification, as well as the occupation of the highest earner. The latter detail is then used to assign individuals/households into a particular group - based on the [National Statistics Socio-economic Classification \(NSSEC\)](#) – which subsequently offers one potential means of determining socioeconomic background.<sup>1</sup> WP remains a key focus of policy across the UK, with several universities using parental education and/or occupation to determine their target groups when carrying out WP activity (as evidenced, for example, by their access and participation plans). However, there continues to be uncertainty around the accuracy of the data, given the questions in

---

<sup>1</sup> HESA also encourages providers to submit information on parental education/occupation for any full-time undergraduate entrants who enter higher education through a different route.

the UCAS application are optional and self-reported. The Office for Students (OfS) have considered completeness through their investigation into the extent of missing data among these two variables, alongside looking at how sector and population proportions compare.<sup>2</sup>

We complement the work undertaken by OfS as follows;

- 1) Using linked HESA-Census 2011 data, we provide a further assessment of the likely accuracy of these two variables.
- 2) We explore whether there is any relationship between the amount of 'missing' data and socioeconomic status (e.g. are those from more disadvantaged backgrounds less likely to supply this information?).

Our analysis does suggest that individuals are reporting this information accurately, though there is some evidence that disadvantaged students are less likely to know the answer or respond to these questions.

<sup>2</sup> See <https://www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/> for more information on these data quality checks.

<sup>3</sup> Qualifications at level 4 and above encompass various types of higher education qualifications. See

## CENSUS 2011

The Census in the UK takes place every ten years and attempts to provide a detailed picture of the entire population. Data is gathered in as consistent a manner as possible across all nations, with households obliged to supply information relating to their socioeconomic background, such as the qualifications they hold and their occupation. These statistics are then aggregated at various geographical levels and made available to the general public, with the smallest level available being the output area (referred to as 'small areas' in Northern Ireland). An example [output area report](#) from the Census is available here.

Using this data, we developed a small UK-wide Census dataset containing the following fields;

- Output area code
- Proportion of individuals in the output area with level 4 qualifications or above (%)<sup>3</sup>
- Proportion of individuals in the output area classified as being in NSSEC groups 1 or 2 (%)<sup>4</sup>

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/qualificationsandlabourmarketparticipationinenglandandwales/2014-06-18#background-notes> for further information.

<sup>4</sup> 'Unknown' and those who are 'long-term unemployed or never worked' are coded as missing data. Consequently, the

## LINKING TO HESA DATA

To carry out the analysis required, we use output area codes to link the compiled Census dataset to a relevant HESA data extract containing the parental education and occupation fields from the Student record. We were able to assign output areas to each individual in our database by using their postcode information. The population of interest is restricted to UK (excluding Jersey, Guernsey and the Isle of Man) domiciled full-time first degree entrants aged 20 or under at the time of starting university, with the academic years covered being 2011/12 to 2016/17. The rationale behind choosing this age group is that the occupation data UCAS gathers only relates to the parent if the individual was under 21 when beginning their degree. Please note that we also include any students who entered via a non-UCAS route in this study. While we cannot be sure that such students are asked the same questions as those that appear on the UCAS application, we find only a few percent of our total population in each academic year enrolled into university through a pathway other than UCAS and their inclusion is therefore unlikely to materially influence our conclusions.

With our linked HESA-Census dataset, we firstly partition each of the two continuous Census fields into deciles for each academic year. Decile 10 contains those areas where the proportion of individuals with a level 4 qualification or above (in NSSEC groups 1 and 2) is highest, with decile 1 comprising the localities with the lowest proportions. By academic year of entry, we then cross-tabulate the HESA parental education and occupation fields against the appropriate corresponding categorical Census variable. The NSSEC field in the HESA data had to be transformed prior to doing this, so that it mimics the Census variable as closely as possible. Specifically, we begin by recoding those who fall into the 'never worked and long-term unemployed' or 'not classified' groups as missing data. [Guidance from the Office for National Statistics \(ONS\)](#) suggests that care should be taken on whether to amalgamate the 'never worked and long-term unemployed' group with other categories or to analyse this separately. In the HESA data extract, we find that approximately 0.30% of individuals are allocated to this NSSEC category, limiting the potential to examine this group on its own. For this study, we have therefore placed them into our 'missing information' category in the knowledge that the addition of a tiny proportion of individuals

---

proportion in each output area in NSSEC groups 1 or 2 is calculated as follows: (Total in NSSEC groups 1 or 2 / Total in NSSEC groups 1-7) \* 100.

to this group will not fundamentally modify our results. This is followed up by generating a binary marker that indicates whether or not the highest earning parent was based in an occupation that falls within NSSEC groups 1 or 2.

While we are able to explore the relationship for parental education between HESA and Census data from the 2011/12 academic year, we can only begin our examination of parental occupation in 2015/16, as this was the first instance in which parental occupation in HESA records was based on SOC 2010 and therefore mirrored the coding framework used in the 2011 Census. For this piece of analysis in particular, we must therefore work under the presumption that there is not a great deal of alteration in the characteristics of individuals and households living in specific output areas over time (i.e. the Census data by output area would be very similar in 2015 and 2016 to that seen in 2011). Using the 2001 and 2011 Census, [work](#) by Rebecca Tunstall at the University of York has evaluated the extent of variation across the decade (within a lower super output area) in the proportion of residents that were based in NSSEC groups 1 or 2. The study finds little evidence of meaningful change. Though we use a different geographic level, this research provides some support for our assumption.

## HYPOTHESES

One would hypothesise that as we move from decile 1 to 10 of the relevant Census variable, we should observe:

- An increase in the proportion of individuals who self-report that they have a parent who has experience of higher education.
- A rise in the percentage of individuals who self-report that their highest earning parent works in an occupation that falls within NSSEC categories 1 or 2.

## FINDINGS

Beginning with parental education, we see that in the first two academic years we consider (2011/12 and 2012/13), approximately 78% of individuals have responded with either a 'yes' or a 'no' to the associated question on the UCAS application, which is in line with [findings reported by the OfS](#). We also detect the expected trend when cross-tabulating the HESA and Census variables, with the proportion of individuals who have a parent with a higher education qualification rising as we move up the deciles. A noticeable change then occurs from 2013/14 onwards, with around 85% of

individuals now supplying either a 'yes' or a 'no' response, which appears to be largely driven by the decline in the proportion who refuse to provide this information.

Furthermore, we continue to observe the hypothesised pattern, with the results being fairly consistent between 2014/15 and 2016/17.<sup>5</sup>

As with the parental education variable, we note that the extent of missing data in the HESA parental occupation field stood at approximately 15% in 2015/16 and 2016/17. Furthermore, we see that the proportion of individuals reporting that their highest earning parent is based in an occupation that sits within NSSEC groups 1 or 2 increases as we move from decile 1 to 10, with very similar findings evident in both 2015/16 and 2016/17.

One issue that emerges from our examination is that, for both the parental education and occupation fields we collect, it is within the lower deciles of the categorical Census variables that we find a greater share of data that could be described as 'missing'. These deciles will consist of areas where the local population will have fewer educational qualifications and are employed in occupations that are more likely to be associated with lower pay. Hence, these are

localities that could be considered to be more economically and socially deprived.

In the case of parental education, the 'missing' data seems to be driven by the fact that individuals from poorer backgrounds are less likely to know whether or not their parents hold a higher education qualification, rather than due to refusing to supply this information or not responding at all.

This analysis leads us to make the following overarching comments about our work:

- By 2016/17, the proportion of missing data in both the parental education and occupation fields stood at around 15%.
- The extent of 'missing' data is greater among those that seem to be from more disadvantaged backgrounds.
- Aggregate statistics available from the Census data have, however, helped to provide further support for the accuracy of these two variables.

---

<sup>5</sup> 2013/14 also represented the year in which the coding manual for the parental education variable was modified to explicitly include a 'No response given' option. In our analysis, those in the 'No response given' category prior to this

academic year are individuals for whom this field was blank. From 2013/14 onwards, this group incorporates both those with a 'No response given' entry and blank fields.

## FINAL REMARKS

The data from Census 2021 will be published over the course of the next few years, which will offer us another opportunity to further this analysis and quality assure more recent data from the HESA student record. Given the

ongoing importance of WP to providers and policymakers, we shall continue to look for ways in which we can assess and/or improve the quality of the variables we collect relating to this matter. Comments and/or questions on this piece are most welcome and can be sent to [pressoffice@hesa.ac.uk](mailto:pressoffice@hesa.ac.uk).

## TABLES

**Table 1a - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2011/12**

Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	53.8	22.1	9.1	14.8	0.2	30,820
2	47.3	29.3	8.8	14.5	0.2	30,800
3	42.6	34.0	8.2	15.0	0.3	30,825
4	38.4	38.7	8.0	14.7	0.3	30,790
5	35.6	42.5	7.4	14.2	0.2	30,875
6	31.6	46.9	7.0	14.3	0.2	30,750
7	28.5	50.7	6.4	14.2	0.2	30,955
8	24.6	54.9	5.8	14.5	0.2	30,660
9	20.9	59.1	5.4	14.4	0.2	30,790
10	14.9	66.4	3.9	14.7	0.1	30,800
<b>All deciles</b>	<b>33.8</b>	<b>44.5</b>	<b>7.0</b>	<b>14.5</b>	<b>0.2</b>	<b>308,065</b>
<b>Total</b>	<b>104,205</b>	<b>136,920</b>	<b>21,560</b>	<b>44,775</b>	<b>605</b>	

**Table 1b - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2012/13**

Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	54.5	21.4	9.7	14.2	0.2	27,525
2	48.4	28.3	8.8	14.4	0.2	27,565
3	43.4	33.5	8.2	14.6	0.4	27,430
4	40.2	37.5	8.0	14.1	0.3	27,530
5	36.5	41.4	7.2	14.7	0.3	27,485
6	33.0	45.7	6.9	14.2	0.3	27,515
7	29.3	50.3	6.4	13.9	0.2	27,495
8	26.1	54.2	5.6	14.0	0.2	27,505
9	22.1	58.7	4.7	14.3	0.2	27,510
10	15.3	64.8	3.7	16.1	0.1	27,505
<b>All deciles</b>	<b>34.9</b>	<b>43.6</b>	<b>6.9</b>	<b>14.4</b>	<b>0.2</b>	<b>275,070</b>
<b>Total</b>	<b>95,890</b>	<b>119,825</b>	<b>19,000</b>	<b>39,710</b>	<b>655</b>	



**Table 1c - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2013/14**

Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	60.2	23.2	9.5	3.4	3.7	30,125
2	53.5	30.3	9.0	3.4	3.9	30,005
3	48.7	35.1	8.6	3.7	4.0	30,040
4	44.9	39.2	8.2	3.8	4.0	30,065
5	40.9	43.6	8.0	3.5	4.2	30,075
6	37.0	47.7	7.6	3.7	4.1	30,035
7	33.2	52.0	6.8	3.6	4.3	30,045
8	29.1	56.7	6.4	3.4	4.5	30,065
9	23.9	62.4	5.8	3.7	4.3	30,045
10	16.7	69.6	5.3	4.0	4.4	30,055
<b>All deciles</b>	<b>38.8</b>	<b>46.0</b>	<b>7.5</b>	<b>3.6</b>	<b>4.1</b>	<b>300,555</b>
<b>Total</b>	<b>116,610</b>	<b>138,150</b>	<b>22,560</b>	<b>10,840</b>	<b>12,395</b>	

**Table 1d - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2014/15**

Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	61.9	22.9	8.6	1.8	4.9	30,925
2	55.8	29.6	8.1	1.7	4.9	30,910
3	51.0	34.4	7.7	1.9	5.0	30,875
4	46.4	39.2	7.4	2.0	5.1	30,875
5	42.0	44.2	7.0	1.9	4.9	30,900
6	38.9	47.7	6.4	1.9	5.1	30,890
7	34.3	53.1	6.1	1.8	4.9	30,900
8	30.0	58.0	5.4	1.6	4.9	30,895
9	25.1	63.4	4.9	1.8	5.0	30,915
10	17.1	71.7	3.8	2.0	5.3	30,875
<b>All deciles</b>	<b>40.2</b>	<b>46.4</b>	<b>6.6</b>	<b>1.8</b>	<b>5.0</b>	<b>308,950</b>
<b>Total</b>	<b>124,305</b>	<b>143,380</b>	<b>20,225</b>	<b>5,625</b>	<b>15,420</b>	

**Table 1e - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2015/16**

Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	62.0	23.2	7.8	1.6	5.4	31,975
2	56.2	29.2	7.2	1.8	5.6	31,950
3	50.8	34.5	7.0	1.9	5.9	32,000
4	46.9	39.0	6.6	1.9	5.7	31,955
5	43.0	43.1	6.1	1.9	5.9	31,925
6	39.0	47.9	5.9	1.8	5.4	31,960
7	34.5	52.7	5.2	1.9	5.7	31,960
8	30.0	57.4	4.9	1.9	5.8	31,965
9	25.1	62.7	4.4	1.9	6.0	31,955
10	17.2	70.5	3.7	2.0	6.6	31,950
<b>All deciles</b>	<b>40.5</b>	<b>46.0</b>	<b>5.9</b>	<b>1.9</b>	<b>5.8</b>	<b>319,590</b>
<b>Total</b>	<b>129,390</b>	<b>147,050</b>	<b>18,755</b>	<b>5,940</b>	<b>18,455</b>	

**Table 1f - The association between the distribution of qualifications in UK output areas and responses to the HESA parental education field 2016/17**

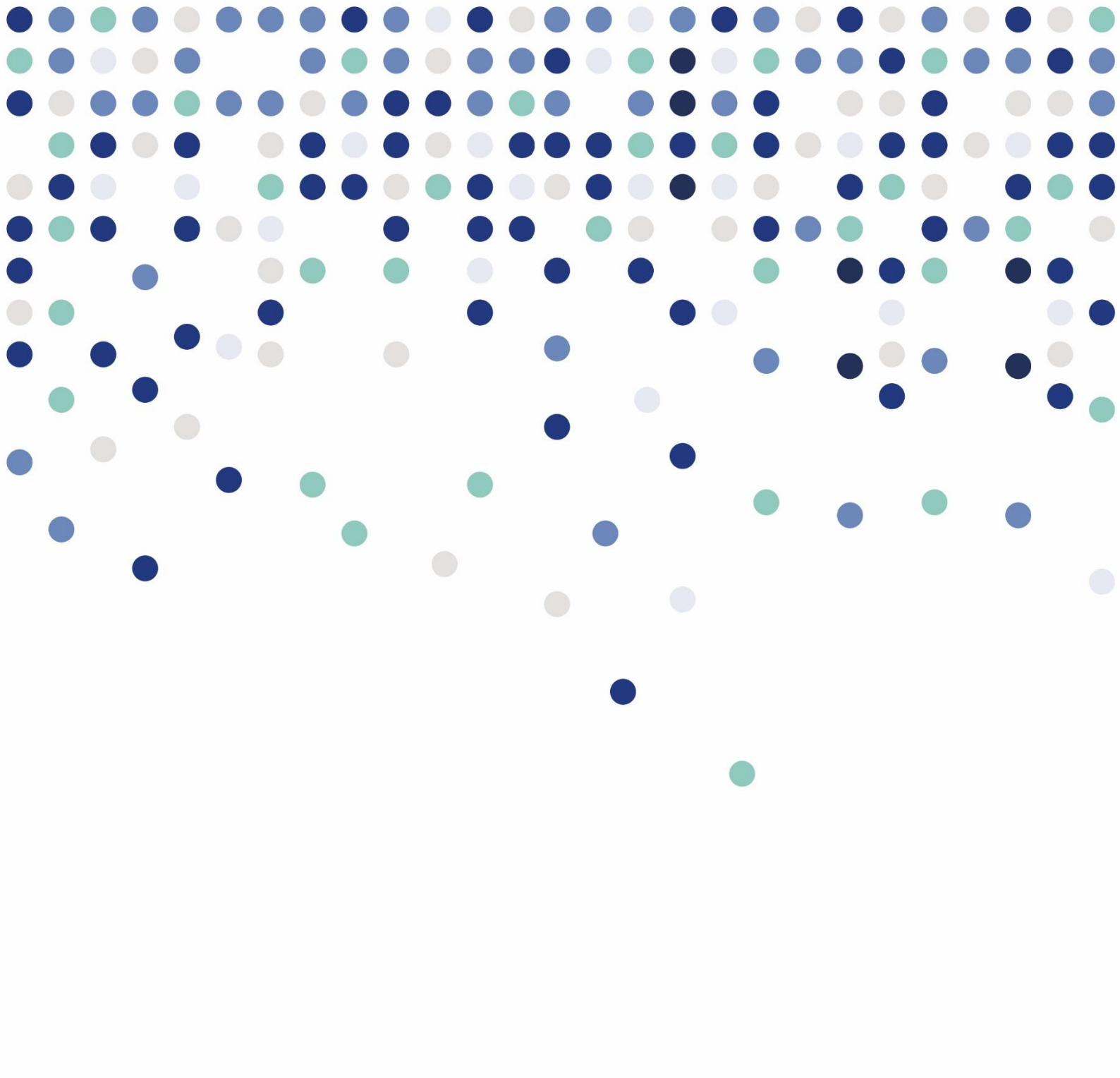
Decile	Parent doesn't have HE qualification (%)	Parent has HE qualification (%)	Don't know (%)	Information refused (%)	No response given (%)	Total
1	62.5	23.4	8.1	1.8	4.2	32,375
2	55.2	29.9	8.0	2.2	4.7	32,365
3	50.4	35.4	7.3	2.0	5.0	32,340
4	46.3	39.8	6.7	1.9	5.2	32,345
5	42.4	44.1	6.2	1.9	5.3	32,385
6	39.1	47.4	5.8	2.0	5.6	32,330
7	33.7	53.8	5.4	1.9	5.3	32,370
8	29.5	58.2	4.8	1.9	5.6	32,360
9	24.5	63.5	4.2	1.9	5.9	32,340
10	16.6	71.6	3.5	1.9	6.5	32,355
<b>All deciles</b>	<b>40.0</b>	<b>46.7</b>	<b>6.0</b>	<b>2.0</b>	<b>5.3</b>	<b>323,565</b>
<b>Total</b>	<b>129,480</b>	<b>151,135</b>	<b>19,445</b>	<b>6,305</b>	<b>17,200</b>	

**Table 2a - The association between the distribution of NS-SEC in UK output areas and responses to the HESA parental occupation field 2015/16**

Decile	Not in a professional occupation (%)	In a professional occupation (%)	Missing information (%)	Total
1	59.5	20.6	19.9	<b>31,960</b>
2	54.6	28.3	17.1	<b>31,970</b>
3	49.7	34.6	15.7	<b>32,115</b>
4	45.7	39.9	14.4	<b>31,815</b>
5	41.2	45.6	13.3	<b>31,945</b>
6	36.3	50.8	13.0	<b>31,985</b>
7	31.8	55.2	13.0	<b>31,930</b>
8	27.2	59.6	13.3	<b>31,950</b>
9	23.2	63.7	13.2	<b>31,960</b>
10	17.1	68.5	14.5	<b>31,960</b>
<b>All deciles</b>	<b>38.6</b>	<b>46.7</b>	<b>14.7</b>	<b>319,590</b>
<b>Total</b>	<b>123,420</b>	<b>149,110</b>	<b>47,060</b>	

**Table 2b - The association between the distribution of NS-SEC in UK output areas and responses to the HESA parental occupation field 2016/17**

Decile	Not in a professional occupation (%)	In a professional occupation (%)	Missing information (%)	Total
1	60.0	20.9	19.1	<b>32,740</b>
2	55.3	28.2	16.5	<b>32,005</b>
3	50.2	34.6	15.3	<b>32,335</b>
4	45.5	40.3	14.2	<b>32,350</b>
5	41.1	45.6	13.3	<b>32,385</b>
6	36.3	51.2	12.5	<b>32,330</b>
7	31.4	55.3	13.3	<b>32,355</b>
8	27.1	59.9	13.0	<b>32,360</b>
9	22.7	64.0	13.4	<b>32,360</b>
10	16.9	68.9	14.2	<b>32,350</b>
<b>All deciles</b>	<b>38.7</b>	<b>46.9</b>	<b>14.5</b>	<b>323,565</b>
<b>Total</b>	<b>125,090</b>	<b>151,655</b>	<b>46,820</b>	



HESA  
95 Promenade  
Cheltenham  
GL50 1HZ  
E [pressoffice@hesa.ac.uk](mailto:pressoffice@hesa.ac.uk)  
T +44 (0) 1242 388 513 [option 6]  
W [www.hesa.ac.uk](http://www.hesa.ac.uk)

Higher Education Statistics Agency Limited is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ. Registered No. 02766993. Registered Charity No. 1039709. Certified to ISO 27001. The members are Universities UK and GuildHE.

