

Research Insight from HESA (Higher Education Statistics Agency – part of Jisc)
Tuesday 14 February 2023
www.hesa.ac.uk/insight/14-02-2023/parental-education-why-students-respond-dont-know
+44 (0) 1242 388 513 [option 6]

Parental education data: Can we ever know why students might respond ‘I don’t know’?

Authors: Archie Bye, Lorraine Crofts, Sophie Gayne, Tej Nathwani

In previous research, HESA highlighted that the parental education field it collected contained around 15% of data that could be classified as ‘missing’. In this insight, HESA researchers explore whether family composition may be a factor that explains why some students respond with ‘I don’t know’ when asked about the qualifications attained by their parents in the UCAS form.

This study will therefore be of particular relevance to those individuals or organisations who utilise parental education data as part of their analysis and/or decision-making processes.

Key findings

- In previous research, HESA highlighted that the parental education field it collected contained around 15% of data that could be classified as ‘missing’.
- This may be the result of individuals refusing to supply this information, choosing to skip the question entirely in the UCAS form or not knowing the answer.
- A possible reason for students responding ‘I don’t know’ could be that they live in single parent households and are therefore not aware of the qualification levels of the other parent.
- By linking HESA data to Census 2011 records, we can identify whether students were living in localities with higher or lower proportions of lone parent households.

- One would anticipate that those residing in areas with greater percentages of lone parent families have a higher probability of responding 'I don't know'.
- Furthermore, Census data reveals that Black African and Black Caribbean individuals are more likely to live in lone parent families.
- We would therefore also expect a larger percentage of 'I don't know' responses among Black African and Black Caribbean students.
- Linked HESA-Census 2011 data finds evidence in support of both of these hypotheses.
- Future research of a qualitative nature could help with confirming that family composition is a reason why students select the 'I don't know' option to the parental education question.
- Should research continue to indicate that household structure is a driver of 'I don't know' responses, it may be worth investigating whether the wording of (and/or guidance associated with) the parental education question might benefit from being changed to make it clearer how individuals from such households should respond.

Introduction

In their latest State of the Nation report, the Social Mobility Commission defined social mobility from an intergenerational perspective and see it as being the difference in the outcomes of an individual when compared with their mothers and fathers.¹ From an educational viewpoint, evidence suggests there is a positive causal relationship between the attainment of parents and their children. Indeed, in the context of higher education, individuals whose parents do not have a higher education qualification are far less likely to attend university.² Consequently, to try to ensure there is equal opportunity for all and improve levels of social mobility within society, higher education providers continue to use data on parental education as part of their outreach activity. These initiatives aim to raise aspiration and attainment among under-represented groups through means such as running summer schools or campus visits. Eligibility criteria will often be set by providers for such programmes and one of these may be that the prospective student is the first in their family to attend higher education. Data on this matter may also be used by researchers or

¹

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1084566/State_of_the_Nation_2022_A_fresh_approach_to_social_mobility.pdf

² <https://www.sciencedirect.com/science/article/pii/S0272775720305240>

sector organisations who are looking to evaluate equality trends. For instance, the Office for Students (OfS) – the regulator in England – utilise this field as part of the regular insights they disseminate on equality and diversity.³

HESA have been gathering data on parental education for approximately fifteen years, with this being sourced from the UCAS application form in which individuals are asked about the qualifications held by their (step-) parents or guardians. Recently, we have published studies looking at the quality of this variable, with one of the findings we have noted being that around 15% of students have ‘missing information’, as a result of either refusing to provide this data, choosing not to respond or simply not knowing the answer to the question.⁴

There is merit in examining the reasons behind missing data. Alongside enabling better quality statistics to be published in this area, understanding the drivers of this issue may help with improving the way the question is asked to potential students in future. This could bring the twin benefit of reducing the extent of missing data we observe, but also assisting individuals to respond more accurately too.

Individuals who apply for higher education via UCAS are optionally asked to answer the following question (in italics) about the education levels of their parents. Alongside the categories of ‘yes’ or ‘no’, respondents can select ‘don’t know’ or ‘I prefer not to say’, as well as skipping this part of the form altogether.

Do any of your parents, step-parents or guardians have any higher education qualifications, such as a degree, diploma or certificate of higher education?

Given the wording of the question, a potential reason for ‘don’t know’ responses is that prospective students are part of a single parent household and therefore may have limited or no knowledge about their other parent. Assessing this hypothesis through the use of data shall be the primary purpose of this study.

³ https://www.officeforstudents.org.uk/media/79a7bb57-83cf-4c50-a358-6bcfe80f165c/ofs2022_29.pdf

⁴ <https://www.hesa.ac.uk/insight/16-03-2021/quality-matters-census-widening-participation>

Data and methods

To carry out the exploration, we draw upon two sources of data – the HESA Student Record and Census 2011. We concentrate on a sample of UK-domiciled (excluding Jersey, Guernsey and the Isle of Man) full-time first degree entrants aged 20 or under at the time of starting university in the academic year 2011/12. While the Student Record does not contain any information that informs us of the type of household that the individual lived in prior to starting their study, we can introduce the 2011 output area code into our HESA dataset by linking their postcode information to higher level geographic data using the Office for National Statistics (ONS) postcode directory. An output area is the building block of Census geography and is a type of locality that generally consists of less than 500 individuals.⁵ In the 2011 Census, there were 232,296 of these areas across the UK. We limit our HESA dataset to three fields – output area code, parental education and ethnicity (the rationale behind the inclusion of this final variable will be explained later in this piece).

One of the ways in which the various administrations responsible for collecting Census data across the four nations of the UK release information to the public is through a set of key statistics (KS) at output area level – each of which covers one of the range of topics covered through the questionnaire. KS105 relates to household composition and supplies data on the proportion of lone parent households in an output area.⁶ In each nation, we therefore downloaded this key statistic and created a UK-wide dataset containing two fields – the 2011 output area code and the corresponding proportion of lone parents with dependent or non-dependent children.

The presence of output area code in both the HESA Student dataset we have compiled and our bespoke Census file containing information on household composition allows us to link these two sources together to form a combined HESA Student Record-Census 2011 dataset. The final file consists of a total of 308,065 individuals for whom we had the corresponding family type information from the Census. The dataset was ordered using the ‘proportion of lone parents in the output area’ field and a new derived variable was created in which students were allocated a decile based on this percentage. Those in the first decile were living in output areas containing the lowest proportion of lone parent households prior to entering higher education, with students in decile ten residing in localities with the highest percentages of single parent families.

⁵ <https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/outputgeography/outputareas>

⁶ See <https://www.nomisweb.co.uk/reports/localarea?compare=E00112425> for an example of the information contained within KS105 across the UK.

Hypotheses

Earlier, we noted that we might anticipate that those living in single parent households are more likely to respond 'I don't know' to the parental education question in the UCAS application form. Though we don't have family composition information at the individual-level, we do know whether a student was living in an area with a higher proportion of lone parent families prior to beginning their course through our linked HESA-Census dataset. Hence, our first hypothesis is as follows;

Hypothesis 1: A greater proportion of students living in areas with a higher share of lone parent families (i.e. those in the top deciles) will respond 'I don't know' to the parental education question in the UCAS form when compared with those residing in localities with a lower percentage of single parent households (i.e. those in the bottom deciles).

Statistics released on the Census looking at the association between ethnicity and household type illustrate that a higher proportion of Black African and Black Caribbean people live in lone parent families.⁷ Consequently, we may also expect to see a higher proportion of individuals within these two ethnic groups respond 'I don't know' to the parental education question. So, our second key hypothesis is;

Hypothesis 2: A greater percentage of Black African and Black Caribbean students will respond 'I don't know' to the parental education question.

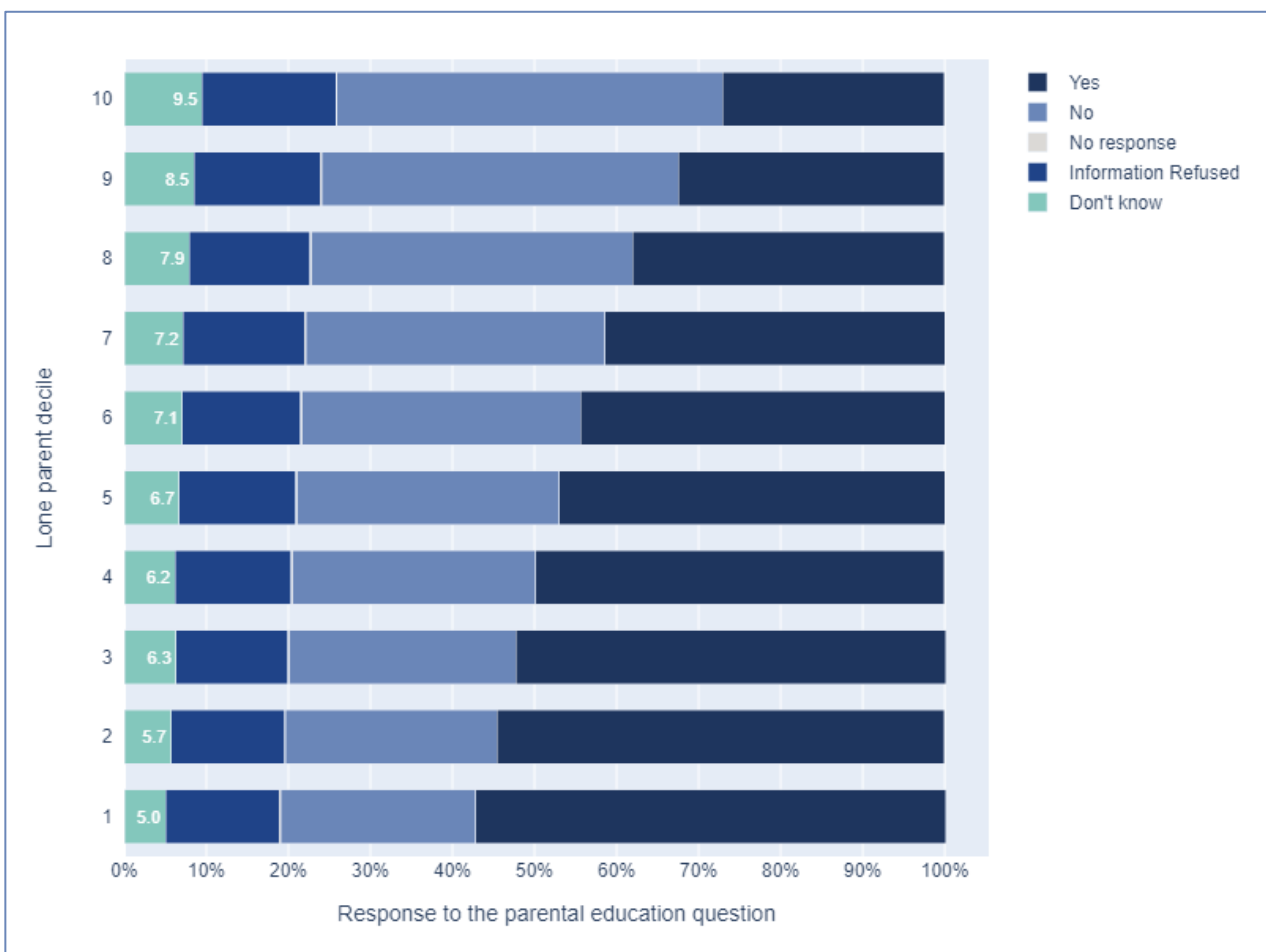
Testing this hypothesis using data was the principal reason for including ethnicity in our extract of HESA records.

⁷ <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/families-and-households/latest#household-types-by-ethnicity>

Results

Figure 1 plots the response to the parental education question by the categorical Census variable indicating the proportion of lone parents in an output area. The first thing that we observe is that the data appears to support hypothesis 1. That is, as we move from decile 1 to decile 10, the percentage of ‘I don’t know’ responses rise.

Figure 1: Responses to the parental education question in the UCAS form by our Census variable on family composition in output areas



Decile 1 represents students residing in places with the lowest proportion of lone parent households, while decile 10 indicates localities with the highest percentages.

Secondly, going from decile 1 to decile 10, we see a clear increase in the proportion of individuals with parents who do not have a higher education qualification. Given that adults with low levels of education are more likely to experience persistent levels of low income, it is probable that those individuals who fall into the higher deciles of our lone parent variable also have a greater probability of being from a disadvantaged background.⁸ Indeed, in our recent work on creating a new UK-wide measure of socioeconomic disadvantage, we showed that areas of high deprivation tend to also have a larger proportion of lone parent families.⁹

In our 2021 insight on the quality of the parental education variable, we stated that the extent of missing information for this field seemed to be higher among those from socioeconomically disadvantaged backgrounds. The evidence presented here suggests that one of the reasons for this may be that there are often a higher proportion of lone parent families in deprived neighbourhoods. Children in such households are less likely to know the education levels of both of their parents, resulting in a greater percentage of 'I don't know' responses.

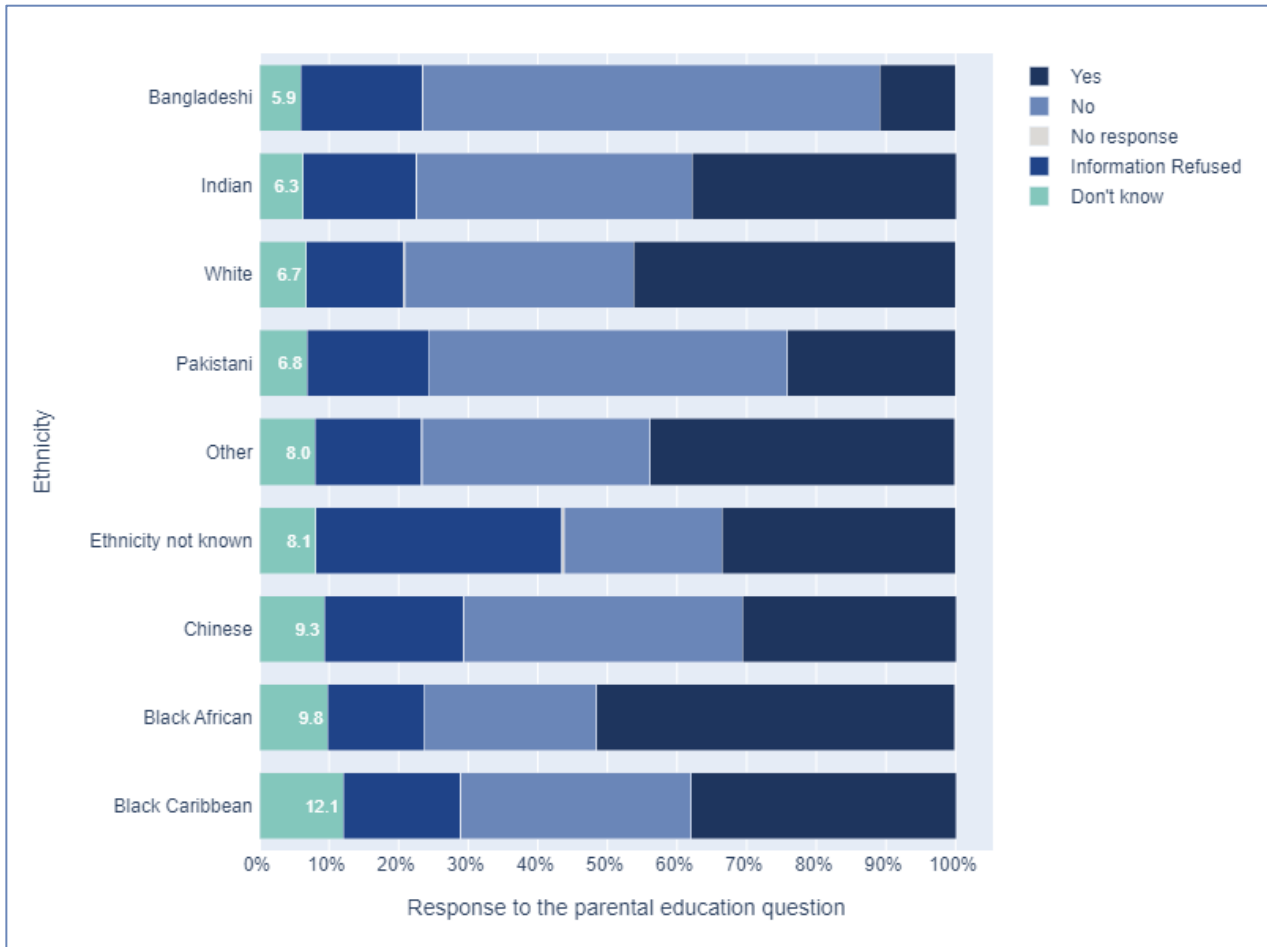
In Figure 2, we show how the responses to the parental education question vary by ethnicity. As set out in hypothesis 2, both the Black Caribbean and Black African groups have a higher proportion of 'don't know' responses, with one of the possible reasons behind this being that individuals from these two ethnic groups are more likely to be part of a single parent household.

8

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/285389/Cm_8781_Child_Poverty_Evidence_Review_Print.pdf

⁹ <https://www.hesa.ac.uk/insight/08-11-2022/new-area-based-measure-deprivation-s4-income>

Figure 2: Response to the parental education question in the UCAS form by ethnicity



Concluding remarks

In this insight, we have illustrated that a potential driver of missing data in the parental education field could be that individuals living in lone parent households are not aware of the qualifications possessed by the other parent and therefore conclude that they should respond by saying 'I don't know'. While the patterns that emerge do align with our hypotheses, future research may wish to strengthen the evidence base we have started to develop here by carrying out qualitative interviews with those who provide such a response and to explore whether family composition is a reason for their answer.

Should further work in this area supply more support in favour of our hypotheses - and given the ongoing utilisation of this variable in the sector - there could be value in considering whether the way the question is asked can be made clearer in future. This may mean that those living in such circumstances are supplied with better guidance on how they should answer this part of the UCAS form, which will assist with reducing the extent of missing data, alongside mitigating the risk of inaccurate responses being given by the individual.

Appendix: Tables of summary statistics

Table A1: Response to parental education question by lone parent decile

	Yes	No	Don't know	Information refused	No response given	Total
1	57.3	23.7	5.0	13.9	0.2	30,965
2	54.5	25.7	5.7	13.9	0.1	30,670
3	52.4	27.6	6.3	13.6	0.2	30,790
4	49.8	29.6	6.2	14.1	0.2	31,905
5	47.0	31.9	6.7	14.2	0.2	30,685
6	44.4	33.9	7.1	14.4	0.2	29,845
7	41.4	36.4	7.2	14.8	0.2	30,785
8	37.9	39.2	7.9	14.7	0.2	30,810
9	32.4	43.4	8.5	15.4	0.2	30,815
10	27.0	47.0	9.5	16.3	0.1	30,795
Total	44.4	33.8	7.0	14.5	0.2	308,065

Table A2: Response to parental education question by ethnicity

	Yes	No	Don't know	Information refused	No response given	Total
White	46.2	32.9	6.7	14.0	0.2	243,585
Indian	37.9	39.6	6.3	16.2	0.1	12,145
Pakistani	24.1	51.6	6.8	17.5	0.0	9,135
Bangladeshi	10.9	65.6	5.9	17.5	0.1	3,565
Chinese	30.7	40.0	9.3	20.1	0.0	2,920
Black African	51.5	24.8	9.8	13.8	0.0	10,270
Black Caribbean	38.1	33.0	12.1	16.9	0.0	3,915
Other	43.8	32.7	8.0	15.3	0.1	20,100
Not known	33.5	22.7	8.1	35.3	0.4	2,435
Total	44.4	33.8	7.0	14.5	0.2	308,065