Methodology appendix

Shifting career motivations are not to blame for worsening teacher shortages

National Foundation for Educational Research (NFER)

Shifting career motivations are not to blame for worsening teacher shortages: Methodology appendix

Dawson McLean and Jack Worth

Published in July 2024

By the National Foundation for Educational Research,

The Mere, Upton Park, Slough, Berkshire SL1 2DQ

https://www.nfer.ac.uk/

ISBN: 978-1-916567-12-2

How to cite this publication:

McLean, D. and Worth, J. (2024). *Shifting career motivations are not to blame for worsening teacher shortages: Methodology appendix*. Slough: NFER.

Contents

Data source

Our analysis relied on data from the UK Household Longitudinal Study (UKHLS) and harmonised British Household Panel Survey (BHPS). We focussed mainly on those who responded to the young adult survey when they were aged 16-21, using the harmonised BHPS/UKHLS to maximise sample size and get as long of a time series of young adults as we could. The waves of data we used in the analysis were BH12-BH18 and UKHLS waves 3-13 for the analysis of interest in teaching.

We used a different set of waves for the career preferences analysis as career preferences information was not collected in every wave. Specifically, we used waves BH12, BH13 and BH17 and UKHLS waves 2, 3, 5, 7, 9, 11 and 13.

Variable definitions and descriptive analysis

Our descriptive analysis of interest in teaching and reported career preferences focussed only on young people aged 16-21 because these variables were only observed for young adults in this age group. Our analysis of interest in teaching by age group was based on whether a respondent had ever said they wanted to be a teacher where they had responded to the survey multiple times within the same age group.

We analysed interest in teaching based only on those who were in full-time studies when they responded to the survey and had a firm idea of what they wanted to do after they completed their studies. Anyone who reported they did not know what they wanted to do, or otherwise had missing data, were removed from our analysis sample.

Our analysis of career preferences was similarly based on those with a valid response to all seven career preferences variables. We removed anyone with missing data or who responded 'don't know' to any of the characteristics. Differences in career preferences between teachers and non-teachers were estimated based on whether a respondent had ever said they considered each characteristic to be 'very important' when they were in the 19-21 age group.

We focussed our analysis primarily on those who said each characteristic was 'very important' for two reasons. First, a young person reporting that a particular career characteristic is 'very important' is a stronger signal than reporting it is 'important'. Secondly, for most career characteristics we considered, the majority of young person considered them to be either 'important' or 'very important'. Focussing on 'very important' only therefore provided us with more variation to use in the statistical modelling.

The descriptive analysis of interest in teaching and career preferences used the Understanding Society cross-sectional weight applicable to each survey wave, to ensure that the results were representative.

We combined a young person's records with their records from age 22-25 to determine how many respondents became a teacher by age 25. Anyone who did not respond to the survey between age 22-25 was removed from this stage of the analysis. We coded anyone who did respond to both the young adult and adult survey but were recorded as not working in the adult survey as 'not working as a teacher'.

We counted someone as ‘working as a teacher’ in a particular year if they held at least an undergraduate degree (variable *hiqual_dv*) and were recorded as working as a teacher (variable *jbsoc90*, *jbsoc00*, *jbsoc10*). We considered teachers to be anyone whose standard occupational classification (SOC) 2010 code was 2314, 2315 or 2316 (and their equivalents for SOC90 and SOC2000). For our analysis, we then considered anyone to have ‘become a teacher by age 25’ if they had worked as a teacher for at least one year before age 25. If an individual was recorded as never having worked as a teacher before age 25, we considered them to have never worked as a teacher.

Weighting to correct for longitudinal attrition

In principle, this analysis necessitated use of a longitudinal weight to correct for longitudinal attrition in the survey (i.e. not everyone who responded to the young adult survey then responded later to the adult survey). Longitudinal attrition could have the potential to bias our results if the likelihood that an individual does not respond to the survey is correlated with the likelihood they became a teacher later in life.

The Understanding Society data provides longitudinal weights for this purpose, however we found that the weights were not suitable for our analysis. This was mainly because the weights were not designed to combine different waves of BHPS/UKHLS data together, nor were they able to handle changes to the survey sample over time (e.g. when boost samples were added to the survey). There was also a high number of zero weights in the provided longitudinal weights (added due to survey design considerations), which would have drastically reduced our sample size had we included them in our analysis.

All of our descriptive and regression analysis that involved combining a young person’s records with their adult records were therefore unweighted. This had the potential to introduce some degree of attrition bias into our estimates. However, we conducted stringent sensitivity analysis of our results to ensure that there was a minimal likelihood that longitudinal attrition was driving our main results. In addition to the unweighted descriptive and regression estimates, we also generated estimates that used:

- The provided BHPS/UKHLS longitudinal weights
- The provided BHPS/UKHLS cross-sectional weights.
- Weighted/unweighted versions of our estimates across different sub-samples of the data (e.g. UKHLS core respondents only, BHPS core respondents only, UKHLS core and boost respondents only)

The sensitivity analysis suggested that our results were not strongly impacted by the inclusion or exclusion of weights, the choice of weights, or the choice of sample. We were therefore confident that longitudinal attrition was not a main driver of our results and we reported the unweighted estimates as they had the highest sample size.

Regression analysis

Our analysis of the impact of career preferences on interest in teaching was based on those who responded to the young adult survey only, and in the waves where both the interest in teaching and career preferences variables were observed (see above). We estimated a logistic regression model with the left-hand side variable showing whether an individual reported that they wanted to be a teacher each year.

We included respondents' reported career preference variables on the right-hand side of the model. Anyone who reported that they did not know what they wanted to be, or who did not know whether one or more career characteristic was important were removed from the analysis.

Instead of including 'helping others' and 'contributing to society' separately as factors, we included a simplified 'pro-social motivation' factor which took on a value of one if a respondent said that either helping others or contributing to society was 'very important' and a zero otherwise. This was because 'helping others' and 'contributing to society' were very highly correlated with each other. We took this approach because we conducted factor analysis of these two variables, which, suggested that a 'pro-social' motivation factor had roughly equal factor loadings on each of these characteristics were roughly equal.

In addition to the career preferences variables, our regression model included controls for a respondent's age, gender and whether they were working in the specification, alongside fixed effects for the survey wave. Since respondents respond to the survey across multiple waves, we also estimated standard errors clustered at the respondent level. Results were unweighted but we conducted sensitivity analysis that including the Understanding Society-provided weights did not significantly alter the main findings (see above). The estimated impacts we reported reflected marginal effects (in percentage point terms) based on the estimated coefficients.

Our analysis of the impact of career preferences on whether a respondent becomes a teacher was based on the same specification. However, the variable on the left-hand side was a dichotomous variable recording whether a young person ever worked as a teacher by the age of 25. Since this variable was based on combining the young adult and adult surveys, the analysis was restricted to only those who responded to both survey modules. Results were unweighted but we conducted sensitivity analysis that including the Understanding Society-provided weights did not significantly alter the main findings (see above).

Estimating what proportion of young people have career preferences that predict they might become a teacher

Finally, our estimates of the proportion of young people whose career preferences predict they enter teaching was based on a separate regression that included the career preferences variables only. This was to ensure that our predictions reflected mainly the most likely effect of overall changes in young peoples' career motivations. We also removed the wave fixed effects from this specification as the coefficients were similar to the original model and it would not be possible to use the wave fixed effects to generate predictions on a sample of data outside of the original training sample. Standard errors were clustered at the respondent level.

To generate the predicted proportions, we first saved the coefficients from the 'prediction' regression model. We then generated a dataset of all respondents to the Understanding Society survey with valid career preferences data (i.e. removing any respondents who said that they did not know whether one or more career characteristic was important to them). We applied the saved coefficients from the prediction model to all of these respondents to generate a predicted probability that each individual would become a teacher by age 25. To generate our final estimates of the proportion of respondents whose career preferences predict they become a teacher, we calculated the average predicted probability of becoming a teacher for all respondents in each birth year cohort.

We estimated bootstrapped standard errors for our average predicted probabilities over 800 bootstrap replications. We saved the coefficients from each bootstrap replicate and generated the average predicted probability in the prediction dataset for each set of coefficients. We then estimated the standard error of the average predicted probability across all 800 replications to generate our final estimate of the overall prediction standard error.

Evidence for excellence in education

Public

The Mere, Upton Park, Slough, Berks SL1 2DQ
T: +44 (0)1753 574123 • F: +44 (0)1753 691632 • enquiries@nfer.ac.uk

www.nfer.ac.uk

NFER ref. NUTR
ISBN: 978-1-916567-12-2