STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
BE551 Degree Project in Economics
Spring 2023

Like Parents, Like Children – The Effects of Social Origins on Children's Higher Educational Attainment in the UK

Jonathan Stefansson (25051) and Peiyu Wang (25103)

Abstract. An educated populace is a prerequisite for a functioning modern society. Educational inequality based on social origins might therefore have large economic consequences for both individuals and society at large. This paper studies how two measures of social origins, social class and parental education, affect the probability of individuals completing an undergraduate degree in the UK. We have used panel data from three cohort studies administered by The Centre for Longitudinal Studies to perform linear probability regressions on how social origins affect degree completion over time. Our results indicate that the effect of social origins largely remained constant between cohorts born 1958 and 1970, but massively decreased in importance for individuals born in 1990. Furthermore, we are able to show how most of this effect seems to be mediated through educational attainment prior to higher education, and how social origin remains highly important in determining who graduates from the most selective universities in the 1990 cohort. Taken together, our study suggests that social origins appear to have become less important in determining whether an individual completes an undergraduate degree in the UK, but that it still has a large effect on from which institution an individual graduates.

Keywords: educational inequality, social origin, higher education

JEL: D62, I23, I24

Supervisor:	Abhijeet Singh
Date submitted:	2023.05.15
Date examined:	2023.05.31
Discussants:	Cecilia Sacklén and Emma Bergman Karlsson
Examiner:	Johanna Wallenius

Acknowledgements

The completion of this study could not have been possible without the support and guidance from our supervisor Abhijeet Singh. We are sincerely grateful for his help. Furthermore, we would like to thank the UK Data Service for providing us access to the datasets that made our analysis possible, and The Centre for Longitudinal Studies for compiling those datasets. Finally, we also appreciate the feedback received from the students in our Midterm Seminar group. It has been of great help to the writing process.

Table of Contents

Introduction

It is widely acknowledged that societies reap both social and economic rewards through an educated population. A higher intensive margin of educational attainment improves labor market outcomes for individuals by, inter alia, increasing their competitiveness and the likelihood of finding jobs, thus leading to higher life-time incomes (Farquharson et al., 2022; OECD, 2022). At a macro level, it enables governments to not only reduce inequality and poverty but also collect more taxes and reduce public spending on social entitlements (OECD, 2022). Ensuring that the citizens of a nation are able to reach their educational potential is therefore vital for a well-functioning society.

The benefits of higher education are intergenerational

The importance of education, and higher education (HE) in particular, is widely accepted in academic discourse and society more broadly. Although entangling the casual and selection effects of HE might be difficult, reviews from Institute for Fiscal Studies (IFS) conclude that completing higher education has a large positive effect on life-time earnings and aspects such as mental and physical health (Britton, 2020; Farquharson et al., 2022).

Furthermore, the IFS review also shows that the benefits of HE are intergenerational: parents with higher levels of education may not only possess better knowledge about studying and educational systems but also have higher incomes to provide their children with the resources needed, e.g. private tutoring. Some of the effects appear to be mediated through the increased likelihood of a child reaching different educational thresholds, stemming from parental education level (Bukodi & Goldthorpe, 2013). An increasing population share of HE completion in the UK might, therefore, be partially explained by the self-reinforcing causal effect of parental education.

Relative educational inequality is persisting

The population share of university degree holders is increasing in the UK and other OECD countries, which is a promising development (Farquharson et al., 2022). However, while the share of the population receiving an undergraduate degree has increased in all social groups, the

gap between different social groups has persisted since at least the middle of the 20th century (Farquharson et al., 2022). The potential issues with such a phenomenon could be economic, concerning "a wastage of talents" and inefficient matching processes in the labor markets, as well as political, as it highlights societal injustices (Breen, 2022). More concretely, since individuals' incomes and education are highly correlated with each other, a persisting relative educational inequality could be translated to a persisting relative income inequality (Breen, 2022). From an individual point of view, educational inequality is, thus, damaging to socio-economically disadvantaged but highly talented individuals, as unequal access to education results in foregone future earnings (Bowman, 1991). For societies, the consequences of inefficiently allocated labor capital could result in, among other things, higher recruiting expenses, higher unemployment rates, as well as spill-over costs on unemployment insurance and placement agencies (Michaillat, 2012). Furthermore, previous research indicates that the rate of innovation, which is a key driver in economic development, is negatively affected by the socio-economic gap in education, and this effect is found to be compounded over time (Bell et al., 2018). According to Bell et al. (2018), "there are many 'lost Einsteins' – people who would have had high-impact inventions had they become inventors" which reflects the consequence of educational inequality and the negative effects of it on the economy.

However, the desirability of educational inequality for economists also depends on whether the definition used is more in line with that of equity or equal opportunity, which is discussed by Bowman (1991). Equity, according to its dictionary definition and in the context of education, entails supporting individuals until everyone reaches the same educational outcomes. Such a policy is not desirable from an economist's point of view, as it must, among other things, lead to substantial matching problems in the labor market. By contrast, equality in opportunities is defined as giving everyone the same opportunities, and thus not necessarily concerned about equal educational outcomes. The concept of educational equality in opportunities is thus more in line with the principles of meritocracy and is commonly justified on the grounds of being socially efficient, whereas equity is usually justified on philosophical grounds (Bowman, 1991). Thus, unequal outcomes in education are not necessarily a problem from an economic perspective since a perfectly meritocratic system with equal opportunities could potentially lead to both perfect allocation of labor and divergence in education levels among a population.

However, if other factors, such as social origins, drive inequality without being mediated through individual abilities, e.g. cognitive ability, the resulting educational inequality would likely imply the presence of a lack of equal opportunities and of the "lost Einsteins" described above.

In 2013, Bukodi and Goldthorpe published a paper showing how the effect of social origins on education has changed throughout the second part of the 20th century (Bukodi & Goldthorpe, 2013). Understanding how this period experienced both general increases in educational attainment and persistent inequality in educational outcomes must be considered vital for properly constructing any potential reforms aimed at closing the gap. Furthermore, the reasons why some gifted students did not study at a higher level in the 1970s might not be the same as today, which implies that looking at whether the effects of social origins change over time is necessary. When conducting this study, the authors used cohort study data from generations born in 1948, 1958 and 1970. Recently, a corresponding study for people born in 1990 released data from when the individuals were 25 years old, which includes questions on whether they had finished an undergraduate degree or not. This allows us to investigate how the effects studied by Bukodi and Goldthorpe have changed between the cohorts born in 1970 and 1990. In particular, we look at how these social origins affect the probability of completing an undergraduate degree, which, so far to our knowledge, no other study has done, likely due to the recent availability of the data.

Previous educational policies have had mixed effects on educational inequality

As laid out in the introduction, even though the UK experiences an increase in the share of degree holders, the gaps in the likelihood of individuals of different social origins doing so seem fairly constant (Breen, 2022). To explain this apparently static level of inequality, Breen (2022) highlights the importance of detangling the causal effects of educational and social background. He presents how previous government programs meant to target educational inequality often have the opposite effect due to the failure of addressing within-school variations in student outcomes, namely hereditary and environmental factors related to social inequality. He concludes that there is a limit to how much an educational policy can do to reduce inequality, both educational and social, and that future reforms should keep that in mind.

However, Breen has also presented that, in contrast to HE, social gaps in lower secondary schools have narrowed over time, implying that closing gaps in HE attainment might not be entirely out of reach. Additionally, curbing inequality at an earlier educational stage could potentially have substantial effects later on, as prior achievements in secondary schools appear to explain a large proportion of HE participation rate variations for students from disadvantaged families (Chowdry et al., 2013).

Literature Review

Definition and measurement of social origin

The International Labour Organization (ILO) defines social origin as "social class, socio-occupational category and caste" (ILO, N.D.). Nonetheless, besides ranking the social status of individuals' backgrounds, parental education is commonly appended to this definition in the field of research. Bukodi and Goldthorpe argue that, to construct a detailed measurement of social origins, separate indicators of social class and three types of family capital – economic, cultural, and social – should ideally be included (Bukodi & Goldthorpe, 2013). However, in the same paper, they also argued for their use of social class schema as a measure for both social class and family economic capital since the construction of it encapsulates information on income levels, employment status and social positions within the labor markets. In the same paper, they have, therefore, chosen to divide social origins into three different measurable components: parental education, parental social class, and parental social status. These should be interpreted as measuring, respectively, the ability to offer direct educational assistance or navigate through the educational system, economic ability, and socio-cultural influence among peers.

They then show how these three variables have large, independent, effects on the likelihood of reaching different educational thresholds (Bukodi & Goldthorpe, 2013), and how adding a control measure of cognitive ability during childhood does not eliminate such effects (Bukodi et al., 2014). This means that an analysis not including one of these variables, such as our own which lacks a variable for parental social status and cognitive ability, will likely result in an overstated effect of the other two variables as a result of omitted variable bias.

Furthermore, cognitive ability only seems to mediate about 30 – 50 % of the effect of social origins on educational performance (Bukodi et al., 2017). In alignment with this, other authors have found that cognitive ability alone plays a lesser role in explaining variations in educational attainment over time while family background has been gaining importance (Galindo-Rueda & Vignoles, 2005). This implies that the majority of the benefits of preferential social origins seem to operate through other channels than by directly increasing the cognitive ability of children. This further underscores the point that a lot of the benefits gained through social origins might be difficult or even impossible to compensate for even if a more equal educational system was constructed. It also raises serious objections to arguing that educational inequality is not important by referring to meritocracy. If hereditary factors render individuals with more educated parents smarter on average, then a large gap in education between different social groups might not imply a large wastage of talents. However, if less than half of the effect of social origins is mediated through cognitive ability, such an argument becomes weaker, as it has to rely on other non-cognitive factors being hereditary and causing the division. Alluding to this argument, Bukodi et al. (2014) also find that social origins seem to become more important for future educational success the higher a child's cognitive ability is, implying that some individuals with lower levels of education might be intelligent but receive less education than they should due to their social background.

Previous findings

Bukodi & Goldthorpe (2013) find that the effect of social class has been constant for cohorts born in 1946, 1958 and 1970, whereas the effect of parental social status is constant between 1946 and 1958 but decreased between 1958 and 1970. The effect of parental education seemingly decreased between 1946 and 1958 but increased between 1958 and 1970 (Bukodi & Goldthorpe, 2013). However, in contrast to our study, the parental education levels are not directly comparable across cohorts, but rather a relative measure of how educated the parents are compared to other parents from the same cohort.

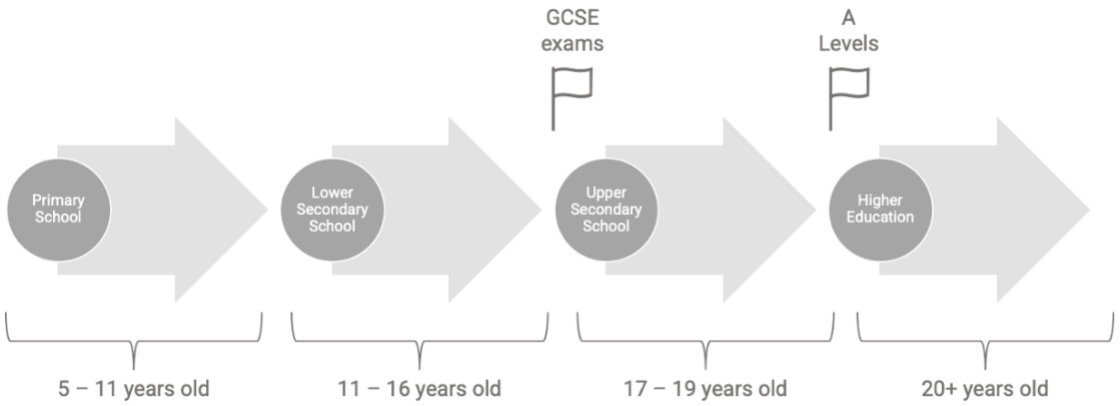
Finally, Bukodi et al. (2017) have carried out a study which looks at the effect of both social origins and early cognitive ability for cohorts born in 1958 and 1970, as well as data on individuals born in 1991/1992 from the Avon Longitudinal Study of Parents and Children

(ALSPAC). As they did not use the same study as we do for the last cohort due to its lack of data on early cognitive ability, the main results might be of less importance for our purposes. Nevertheless, according to their appendix, regressions that use the same three studies we use are available up until the students finish upper secondary school. They seem to imply that the effect of parental social class has not changed, while the effect of both parental education and parental social status has increased (Bukodi et al. 2017, Appendix B).

The education system in the United Kingdom

The education system in the United Kingdom is, for our purposes, divided into four different levels (Farquharson et al., 2022), as illustrated in **Figure 1**:

Figure 1. Simplified illustration of the UK education system.



Note: Authors' illustration.

At the end of lower secondary school, children usually take GCSE exams which determine their future educational paths. Students that do well on these tests and choose to study at upper secondary school usually study for several A-level qualifications, which determine whether they are eligible for higher education or not (Farquharson et al., 2022).

However, the school system of the UK went through substantial changes during the time that our cohort members and their parents attended school. Below is a summary of the, for the purpose of understanding this paper, most important changes of the period:

- The lowest school leaving age was increased from 15 to 16 years of age in 1973 (Sullivan et al., 2014). See **Appendix 3: Construction of *Parental Education*** for a discussion about the effect of this change.
- A transition was made from a system that, at an early age, divided students into different types of schools based on their abilities into a "comprehensive" system during the 1970s and 1980s (Sullivan et al., 2014).
- GCSE exams replaced O-levels at the end of lower secondary school in 1988 (Pearson, N.D.). See more discussions in **Appendix 3: Construction of *Parental Education***.
- The 1992 Higher Education Act substantially increased the number of available places in HE (Sullivan et al., 2014).
- Tuition fees for HE were abolished for most students in the Education Act of 1962 but reintroduced in 1998 (Sullivan et al., 2014).

Russell Group

Established 1994 and incorporated 2007, The Russell Group consists of 24 research-intensive universities in the UK (Russell Group, N.D.). Its members range from the most prestigious institutions both national- and worldwide, e.g. University of Oxford, University of Cambridge and Imperial College London, to their less prominent counterparts, e.g. Newcastle University and Queen Mary University of London. Although varying in perceived prestige and educational quality, a commonly shared focus on research and a large share of research grants received render the Russell Group universities attractive to academics (Study in UK, N.D.). Consequently, these institutions are highly selective.

Data and Variables

Data

Our main data sources are three large cohort studies extracted from the Center for Longitudinal Studies: National Child Development Survey (NCDS), British Cohort Study (BCS), and Next Steps (NS) which, respectively, follow individuals born in 1958, 1970, and 1990.

The survey participants, namely the cohort members, answer survey questions on various topics at different points in time which enables us to follow their individual characteristics over time. The topics include, inter alia, family, education, social behavior, and attitudes. Within a cohort study, a survey conducted at a particular point in time is termed a wave or a sweep. For the 1958 and 1970 cohort studies, the first wave started at the time of cohort members' birth, whereas for the 1990 study, it started at the time of age 14. The parents of cohort members assist in answering the questions either entirely or partially, depending on the age of the cohort members, among other conditions.

Since these cohort studies are not identically constructed and, hence, not always completely comparable, a presentation of how we have compiled and standardized the variables is available in **Appendix 2: Construction of Dependent Variable**, **Appendix 3: Construction of *Parental Education*** and **Appendix 4: Construction of *Social Class***. In addition to these three cohort studies, we also use National Child Development Study and 1970 British Cohort Study Educational Qualifications Histories (1981-2009) in our study. These separate data sets are compiled by Bukodi and contain standardized data on education for the 1958 and 1970 cohorts.

Variables

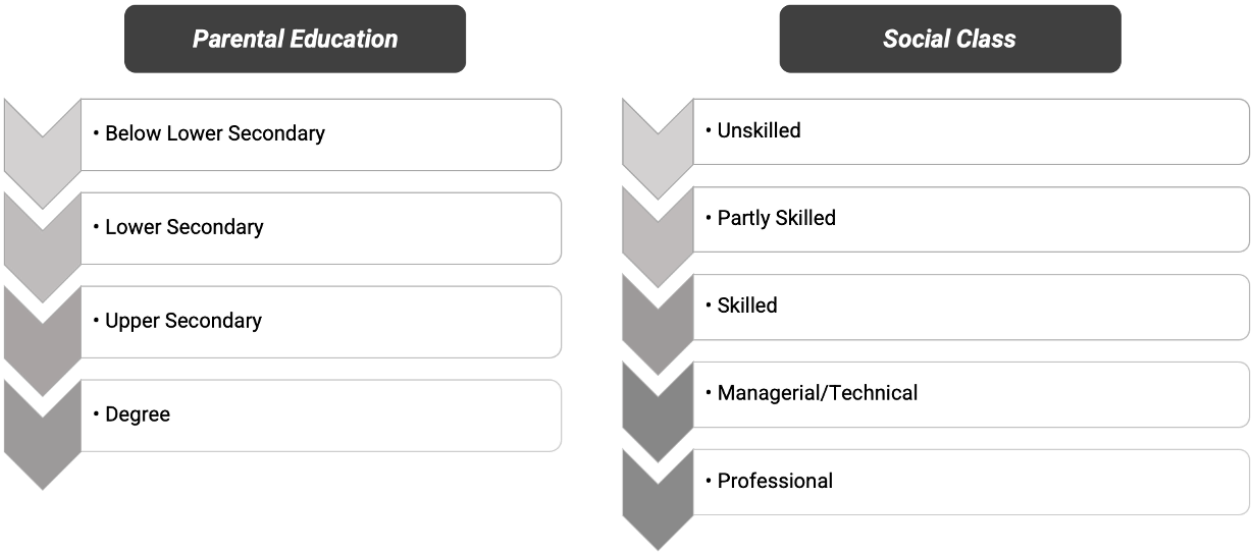
Dependent Variables

Our main dependent variable is *Undergraduate Degree Attainment* which is a binary variable. It takes on the value 1 when a cohort member has obtained an undergraduate degree and 0 otherwise. Additionally, we introduce another binary dependent variable, *Russell Group*, which specifies whether an individual's undergraduate degree is from a member university in the Russell Group. Similar to *Undergraduate Degree Attainment*, *Russell Group* takes on the value 1 when a cohort member has obtained an undergraduate degree from a Russell Group university and 0 otherwise. This analysis is only applicable to the 1990 cohort as the other cohort studies lack data on which university an individual has attended.

Independent Variables

Our two main independent variables are *Parental Education* and *Social Class* which are ordinal variables. The value ranges which *Parental Education* and *Social Class* take on can be seen below in **Figure 2**:

Figure 2. Value ranges of Parental Education and Social Class.



Note: Authors' illustration.

Professional and *Degree* are considered the most advantageous levels for *Social Class* and *Parental Education*, respectively.

For the construction of *Parental Education*, we use the absolute approach to measure parental education, i.e. whether a parent has reached a certain educational level or not, instead of how educated they are relative to other parents of the cohort members. A discussion on the choice of the absolute approach is available in **Appendix 3: Construction of *Parental Education***. For the mediation analysis, we have chosen to include A levels study as a mediating variable which is denoted by *A Levels* in this paper. *A Levels* is a binary independent variable which takes on the value 1 when a cohort member has obtained A Levels or enrolled in A Levels study, and 0 otherwise.

Variable Choice

Besides having been inspired by previous papers done on the same topic, the choice of dependent and independent variables is also influenced by the definition of social origin and its conventional measurements. Due to the lack of data for constructing parental social status and individuals' cognitive ability in one or more cohort studies, we have not included such measures.

For the 1958 and 1970 cohorts, we were able to use a variable which indicates whether an individual has completed A-levels or not, whereas the variable used for the 1990 cohort only shows if an individual has enrolled in A-levels study or not. Nevertheless, both these variables allow us to better understand how social origins affect undergraduate degree attainment by showing if they still matter after having finished upper secondary school. If the effect is fully mediated, it indicates that the effect on degree attainment mostly stems from socially advantageous individuals being more likely to finish upper secondary school, which is a prerequisite for HE.

Missingness Analysis of Key Variables

Due to the nature of longitudinal studies and the fact that our key variables are retrieved from different survey waves, our final data sets suffer from sample attrition. There are various explanations for the presence of sample attrition and it is important to explore whether the potential, relevant causes indicate that our data is missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). MCAR refers to the missing data being dependent on neither the observed characteristics nor the unobserved characteristics, while MAR refers to missingness caused by observed characteristics rather than unobserved characteristics. Lastly, the MNAR mechanism refers to missing data resulting from unobserved characteristics (Metten et al., 2022).

In **Table 1** we present a missingness analysis of our key variables for each cohort study. Starting from the initial, unfiltered sample, each row in the table shows the size of the samples after filtering out individuals whose information related to a particular variable lacked. **Table 2** shows the cumulative result of filtering out all the individuals who had missing data related to one of our key variables. For all three of our cohort studies, this removes a large share of the individuals

that participated in the first wave. We present how this alters the formation of our sample in **Table 3**, where we have shown how the share of individuals of different social origins changes between our initial and final samples. Furthermore, we have also conducted a t-test on whether these effects are statistically significant. As can be observed, almost all the changes turn out to be significant.

It is therefore plausible that our degree data suffers from either between-wave or within-wave sample attrition, or both simultaneously. The between-wave sample attrition means that some individuals drop out entirely from later survey waves while the within-wave sample attrition implies that some individuals continue to participate in later surveys but do so without answering all questions. The main reason we have for suspecting that within-wave attrition might correlate with individual characteristics is our undergraduate degree indicator for the 1990 cohort. According to an IFS review, 43 % of the UK population between 25 and 34 had a degree (Farquharson et al., 2022), which makes the 59 % degree completion rate in our data seem inflated. A possible explanation for this is that people without a degree are more likely to skip this question, or the section of questions that discuss educational qualifications, see **Appendix 2: Construction of Dependent Variable**. This could pose a serious problem for our results if we assume the data to be missing completely at random whereas it was conditionally missing at random. Therefore, we have chosen to use inverse probability weighting to ameliorate these issues.

Inverse Probability Weighting

To ameliorate the effects of sample attrition, we will employ a statistical method called inverse probability weighting (IPW). The method starts by estimating the probability of participation for all complete individuals, i.e. the individuals that participate in a particular survey wave. During the estimation step, the complete individuals are divided into different characteristic groups. For this purpose, we have chosen to perform probit regressions. After having obtained the probabilities, it is then possible to calculate the representative weights of the subgroups by taking the inverse of the probabilities. We will demonstrate this concept with the following hypothetical example. If the survey participation rate would be 80 % for females and 50% for males, the respective weights of female and male participants would be 1.25 and 2. One could therefore say

that if an individual with a certain characteristic has a comparatively low participation probability, his/her response in a survey corresponds to a higher weight, as this individual carries "the weight of non-respondents" (Metten et al., 2022).

The literature has suggested that IPW should be used when MAR mechanisms are present, which is the case in our study as we believe that our sample attrition is correlated with survey participants' educational level. Thus, we will present IPW regression results in **Results** and non-IPW regression results in **Appendix 5: Main Regression Results without IPW** for comparison. IPW has also been used on the samples we use to conduct our mediation and Russell Group regressions, as they both suffer from even more sample attrition due to the addition of more variables. However, it should be clarified that IPW cannot solve issues relating to non-random missingness, namely MNAR. This implies that even IPW results could be subject to bias if some unobserved characteristics cause the missingness.

Methodology

Our initial approach follows very closely to that of Bukodi & Goldthorpe (2013), which will be referred to as the BG paper onwards in our work. Our first two research questions are:

1. *Is there an independent effect of Parental Education and Social Class on children's higher education attainment for individuals born in 1958, 1970 and 1990?*
2. *How do these effects change over time?*

To investigate our first research question, we have constructed the following model:

$$\begin{aligned} Undergraduate\ Degree\ Attainment_i \\ = \beta_0 + \beta_1 * 1958\ Cohort + \beta_2 * 1990\ Cohort \\ + \beta_3 * Partly\ Skilled \\ + \beta_4 * Skilled \\ + \beta_5 * Managerial\ /\ Technical \\ + \beta_6 * Professional \\ + \beta_7 * Lower\ Secondary \\ + \beta_8 * Upper\ Secondary \\ + \beta_9 * Degree + \varepsilon_i \end{aligned}$$

To answer our second research question, we perform both LPM regressions at the cohort level and a LPM regression where we interact cohort indicators with our two independent variables which allows us to see if the effects of social origin change across the three studies. The interaction effect model specification is:

$$
\begin{aligned}
&Undergraduate\ Degree\ Attainment_i \\
&\quad = \beta_0 + \beta_1 * 1958\ Cohort + \beta_2 * 1990\ Cohort \\
&\quad + \beta_3 * Partly\ Skilled \\
&\quad + \cdots \\
&\quad + \beta_6 * Professional \\
&\quad + \beta_7 * Lower\ Secondary \\
&\quad + \cdots \\
&\quad + \beta_9 * Degree \\
&\quad + \beta_{10} * Partly\ Skilled * 1958\ Cohort \\
&\quad + \cdots \\
&\quad + \beta_{13} * Professional * 1958\ Cohort \\
&\quad + \beta_{14} * Lower\ Secondary * 1958\ Cohort \\
&\quad + \cdots \\
&\quad + \beta_{16} * Degree * 1958\ Cohort \\
&\quad + \beta_{17} * Partly\ Skilled * 1990\ Cohort \\
&\quad + \cdots \\
&\quad + \beta_{20} * Professional * 1990\ Cohort \\
&\quad + \beta_{21} * Lower\ Secondary * 1990\ Cohort \\
&\quad + \cdots \\
&\quad + \beta_{23} * Degree * 1990\ Cohort + \varepsilon_i
\end{aligned}
$$

All variables in both models are binary variables and their estimated effects will be compared to the effect of *Unskilled* and *Below Lower Secondary* in the 1970 cohort.

In addition to replicating the BG paper with our new dataset, an analysis of whether previous educational attainment mediates the effect of social origins on undergraduate degree attainment will also be conducted. Finally, we will also study how social origins affect the probability of graduating from a Russell Group member university compared to from a non-member university. This will be done by regressing our indicators of social origins on a binary variable indicating if an individual has a degree from a Russell Group university or not. Our third and fourth research questions are:

3. *Does previous educational attainment partially or fully mediate the effect of social origins on the probability of obtaining an undergraduate degree?*
4. *How does social origin affect the probability that an individual will graduate from a Russell Group university for individuals born in 1990?*

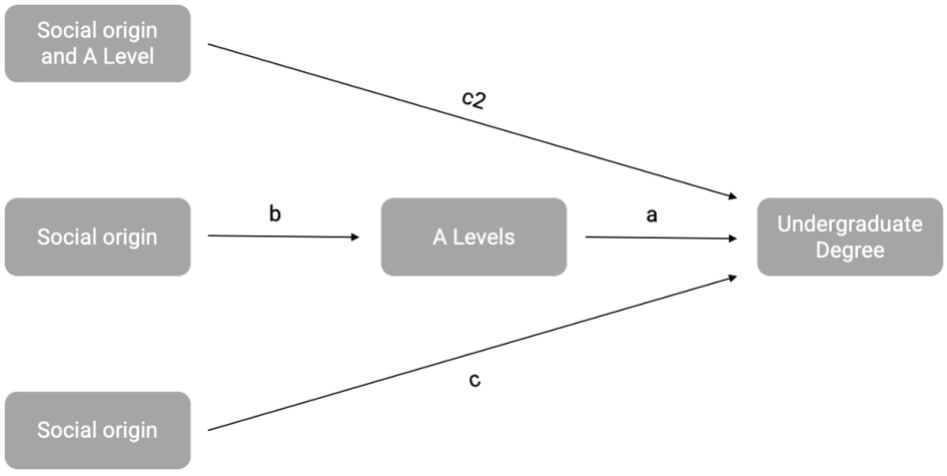
The effect of social origin might be mediated, either partially or fully, through previous educational attainment, namely *A Levels*. To investigate the presence of the mediation effect and answer our third research question, we have constructed four models as illustrated in **Figure 3**. For partial mediation to be the case, three requirements must be fulfilled:

1. Social origin must affect *Undergraduate Degree Attainment* (model c)
2. Social origin must affect *A Levels* (model b)
3. *A Levels* must affect *Undergraduate Degree Attainment* (model a)

Additionally, full mediation requires a fourth condition:

4. The effect of social origin must become insignificant when *A Levels* is included as a regressor (model c2)

Figure 3. Mediation effect models.



Note: Authors' illustration.

Lastly, the model specification for our final research question is:

$$
\begin{aligned}
&Russell\ Group\ Degree\ Attainment_i \\
&\quad = \beta_0 \\
&\quad + \beta_1 * Partly\ Skilled \\
&\quad + \beta_2 * Skilled \\
&\quad + \beta_3 * Managerial\ /\ Technical \\
&\quad + \beta_4 * Professional \\
&\quad + \beta_5 * Lower\ Secondary \\
&\quad + \beta_6 * UpperSecondary \\
&\quad + \beta_7 * Degree + \varepsilon_i
\end{aligned}
$$

Results

Summary Statistics

We begin by providing summary statistics for our main variables. Further sample analysis and in-depth statistics can be found in **Appendix 1: Additional Summary Statistics and Graphs**.

Parent Educational Level

The distribution of *Parental Education* across cohorts is shown in **Table 4**. It can be observed that parents have become more educated over time. The share of lower education levels, *Below Lower Secondary* and *Lower Secondary*, has decreased while the share of more advanced education levels, *Upper Secondary* and *Degree*, has increased.

Social Class

In **Table 4**, the distribution of *Social Class* across cohorts is presented. Between 1958 and 1970, the share of social classes stayed comparatively stable, with a slight upward shift towards the higher classes. The share of cohort members from *Skilled* plunged between 1970 and 1990 which seems to have caused all other classes, and *Managerial/Technical* in particular, to increase for the 1990 cohort.

Undergraduate Degree Attainment

As shown in **Table 4**, the share of cohort members obtaining an undergraduate degree has, similar to *Parental Education*, increased over time. Examining the outputs from the same table divided by sex in **Appendix 1: Additional Summary Statistics and Graphs**, this overall

increase is apparent in both male and female subgroups and the increase for females is larger than that for males.

Cohort Member Sex

The distribution of cohort member sex is shown in **Table 4**. The distribution of men and women within cohorts has been approximately even but the share of female cohort members has slightly increased over time.

Graphs

In **Figure 4**, the within-social class relation between time and *Undergraduate Degree Attainment* is illustrated. It can be observed that, going from the 1958 cohort to the 1970 cohort, the share of undergraduate degree holders has increased for cohort members from all social classes, except for *Unskilled* whose share slightly decreased. Comparing the 1970 and 1990 cohorts, a unanimous upward trend in *Undergraduate Degree Attainment* can be observed and it is most sizable in the less advantageous social classes. Breaking down this analysis to gender level, as shown in **Appendix 1: Additional Summary Statistics and Graphs**, we can confirm that the same trend applies to both female and male cohorts.

The within-education level relation between time and *Undergraduate Degree Attainment* is illustrated in **Figure 5**. Between the 1958 and 1970 cohorts, the share of undergraduate degree holders in each *Parental Education* level group marginally rose, compared to more pronounced surges between the 1970 and 1990 cohorts. The same trend can also be seen within the gender subgroups in **Appendix 1: Additional Summary Statistics and Graphs**.

In **Figure 6**, the time trend of *Undergraduate Degree Attainment* is presented at the *Cohort Member Sex* level. The share of degree holders has increased across cohorts, with a drastic increase between the 1970 and 1990 cohorts. Although initially starting at a lower level, the female subgroup has experienced a strong and steady upward trend, and exceeded the male subgroup by the time of the latest cohort.

Regression Results

1. *Is there an independent effect of Parental Education and Social Class on children's higher education attainment for individuals born in 1958, 1970 and 1990?*

To begin with, we present a LPM where *Social Class* and *Parental Education* are regressed on *Undergraduate Degree*, as well as the same regression for subgroups based on sex in **Table 5**, the majority of the levels of our regressors have a highly significant and positive effect on *Undergraduate Degree Attainment*. The only exception is that the effect of moving from *Unskilled* to *Partly Skilled* seems insignificant. Thus, it appears that it is more beneficial to have advantageous social origins as it increases the probability of completing a degree. This result also generally holds true for both gender subgroups. Moreover, it can also be observed that *Parental Education* seems to have a larger impact on degree attainment than *Social Class* due to a larger coefficient magnitude, which is in line with the results from the BG paper. Following our LPM regression, we apply logit regression to the same data as shown in **Appendix 7 Table 26**. Our conclusions from the LPM regression also seem to hold in the logistic one, as no apparent changes occur. **In response to our first research question, we can therefore conclude that there is an independent effect of both our independent variables on the probability of completing an undergraduate degree across all three cohorts.**

2. *How do these effects change over time?*

Table 6 shows a LPM model for each of the three cohorts, as well as male and female subgroups for the cohorts. We have also conducted an analysis with interaction terms, presented in **Appendix 7 Table 27**, where we formally test for any changes in effects over time.

Starting by comparing the results between the 1958 and 1970 cohorts, **Table 6** suggests that, following the results of the BG paper, the overall effect of *Social Class* and *Parental Education* has not changed much. The stagnant effect is partly confirmed by **Appendix 7 Table 27**, which indicates that the only significant coefficient changes appear to be between *Unskilled* and *Partly Skilled*, as well as between *Below Lower Secondary* and *Degree.* The results from **Table 6** further imply that the effect of *Partly Skilled* and *Degree* on the probability of individual degree attainment gain in magnitude for the 1970 cohort. Finally, our logistic regression in **Appendix 7 Table 28** supports the above results.

In contrast, when comparing the 1970 and 1990 cohorts, **Table 6** indicates that being born in the latter one appears to have overall significantly reduced the effect of *Social Class* and *Parental Education* on *Undergraduate Degree Attainment*. This is again confirmed in **Appendix 7 Table 27** where all variables except for *Skilled* are highly significant. Worth noting is that these changes are very large in magnitude. Finally, our logit regression analysis in **Appendix 7 Table 28** supports the same result.

In response to our second research question, our results suggest that the general importance of *Social Class* is constant between the first two generations but then decreases for the latest generation. The effect of *Parental Education* seems ambiguous between the two first generations but decreases between the last two generations.

3. Does previous educational attainment partially or fully mediate the effect of social origins on the probability of finishing an undergraduate degree?

Following the previously mentioned principles to interpret **Table 7**, **Table 8** and **Table 9**, we can obtain information about the presence and types of mediation. The summary of mediation effects is demonstrated in **Table 10**.

In response to our third research question, our results imply that the majority of the effect of social origin is mediated, either fully or partially, through educational achievements prior to attending university. Breaking down the social origin factors, *Social Class* seems to operate through the channel of prior educational achievements to a greater extent than *Parental Education*. Furthermore, it appears that *Parental Education* has become more important after upper secondary school over time, as it seemingly goes from being fully mediated to partially mediated. However, since the 1990 cohort uses a different measure of previous educational achievements, this should be interpreted with caution.

4. How does social origin affect the probability that an individual will graduate from a Russell Group university for individuals born in 1990?

Finally, we have examined whether social origin affects the probability of attending and obtaining a degree from a Russell Group University. This analysis was only possible for the 1990 cohort and between individuals that have obtained a degree because of how the Russell Group indicator was structured in that cohort study. As seen in **Table 11**, both *Social Class* and *Parental Education* have a large positive effect. Individuals of the most advantageous social class, *Professional*, are 16.4 percentage points more likely to go to such a university compared to their peers of the least advantageous social class, Un*skilled*. Similarly, individuals with the most educated parents, *Degree*, are 16.6 percentage points more likely to do so. Considering that approximately 27 % of the sample has such a degree, these effects are large. In addition to this, the difference between *Managerial/Technical* and *Professional* is approximately 8.9 percentage points, and 8.7 percentage points for *Upper Secondary* and *Degree*. **In response to our fourth research question, our results therefore imply that social origin has a large effect on the probability of attending and gaining a degree from a Russell Group university, something that is particularly true for individuals of the most advantageous social origins.**

Limitations

This study suffers a few limitations. Because of the lack of available data, we were unable to construct variables indicating parental social status and cognitive ability for our three cohorts. As social status is part of the framework for measuring social origins, devised in the BG paper, this leaves our analysis vulnerable to omitted variable bias. A measure of cognitive ability would also have been useful in order to test whether social origins are mediated through cognitive ability. Including it would have allowed us to better answer the introductory discussion about whether educational inequality resulted from an imperfect meritocratic system or not, which could in turn help us gauge the presence of inefficient labor matching.

Moreover, the inconsistency of cohort study designs, including the number of questions and questions asked, has resulted in some of our variables becoming not directly comparable across cohorts. For example, the data available on A Levels information is A Levels enrollment for the 1990 cohort, compared to A Levels attainment for the two previous cohorts.

Instead of the combination of IPW and deletion, the issues with missing values in our datasets could have been handled by using other statistical methods, e.g. multiple imputation. However, these methods lie beyond the scope of this Bachelor thesis.

We would also have liked to have an indicator for Russell Group universities across all three cohorts to analyze the time trend at an institutional level. Besides this, an indicator that is constructed to be applied to the entire sample instead of to degree holders only, would have been ideal, see **Appendix 8** for further discussion. To remedy the above-mentioned limitations and perfect our current study, more consistent and complete cohort study data is required. However, this has not been feasible for us due to the lack of data access and the nature of large, national cohort studies spanning over decades.

Conclusions and Discussion

The effect of social origin on educational attainment has both economic and political implications. A better understanding of how these effects work, and if they change over time, is therefore crucial to understanding both what types of reforms might be required for a more meritocratic society, and how future reforms might affect the educational outcomes of people with different social origins. In this study, we have added to the knowledge in this area in several ways:

1) We confirm Bukodi and Goldthorpe's result that social class and parental education have independent and positive effects on individuals' undergraduate degree attainment holds true for the cohort born in 1990.
2) We confirm that the effect of social class and an absolute measure of parental education did not seem to change much between the cohorts born in 1958 and 1970, besides an increase in importance for individuals with the most educated parents. This stands in contrast to Bukodi and Goldthorpe's findings that there was a general increase in the importance of parental education measured in relative terms, which highlights the importance of the treatment of this variable.
3) We show that the effect of both social class and parental education became less important between cohorts born in 1970 and 1990.

4) We show that a large proportion of the effect social origin has on undergraduate attainment is mediated by educational attainment prior to HE. This is particularly true for the social class factor across all generations, whereas parental education appears to become less mediated by previous educational achievements over time.
5) We show that, for the 1990 cohort, individuals with the most advantageous social origins are much more likely to graduate from a Russell Group university.

The results for the 1990 cohort study are of particular interest as they are not present in the previous literature. Social class and parental education are found to have a positive effect on undergraduate degree attainment, however, this effect has massively declined in importance compared to the previous two cohorts.

At first glance, this seems very promising as to how it indicates a break in the phenomenon of a persisting educational gap between people of different social origins. This can be seen clearly in **Figure 4** and **Figure 5**, where individuals of the second most advantageous and the less advantageous social origin levels converged between the 1970 and 1990 cohorts, despite the most advantageous subgroup maintaining its distance. Since the convergence in general educational outcomes occurs simultaneously as a large increase in educational attainment for all subgroups, it is even more encouraging from an economic perspective.

Understanding exactly why this has happened might be beyond the scope of this paper but some tentative notes can be made. Firstly, as discussed in **Missingness Analysis of Key Variables**, there are indications that the undergraduate degree rate for the 1990 cohorts is inflated. Secondly, as discussed in **Literature Review**, the educational system of the UK has undergone substantial changes between 1970 and 1990, and some of these changes might have plausibly affected our results. The two main candidates for the second explanation are the large expansion of HE in 1992, and the reintroduction of tuition fees in 1998. Although the expansion of HE might explain the general increase in the share of degree holders (Sullivan et al., 2014), and conceivably have a positive effect on narrowing the attainment gaps, the reintroduction of tuition fees seems, if anything, more likely to increase the size of such gaps (Bukodi et al., 2017). Another possible candidate could also have been the transition to a “comprehensive” education system which does

not divide students at an early age in the 1970s and 1990s, but any effect of such a reform would arguably also affect the cohort born in 1970 (Sullivan et al., 2014).

Due to the recent availability of the data, it is also hard to compare our results to that of other studies. However, as mentioned in the **Literature Review**, Bukodi et al. (2017, Appendix 2) have looked at how the importance of social origin on the likelihood of completing upper secondary school changed between our three cohorts. They find no changes in the effect of social class, and an increase in the importance of parental education, the latter being contradictory to our findings. Worth noting that they also include a measure of cognitive ability and parental status which renders our results not directly comparable. The increase in the importance of parental education is explained by an increasingly complex educational system, an explanation which should certainly apply to university studies and thus our results. The effects of social origins on upper secondary education and HE do not necessarily have to go in the same direction, but our mediation analysis suggests that this is, in fact, the case. The seemingly contradictory results of our study and the results of Bukodi et al. (2017) therefore indicate that further research on this topic might be fruitful.

Moving on to our mediation analysis, it is clear that most if not all of the effect of social class on undergraduate degree attainment seems to go through the channel of previous educational achievements. This is interesting considering that social class should mainly capture family economic resources in this framework, and that, as stated above, university tuition fees were present for the 1990 cohort.

Parental education seems to be more important than social class given that a student has enrolled in or completed A Levels. As indicated earlier, this might be the case because highly educated parents have more knowledge about the educational system which enables them to better assist their children with choosing an appropriate tertiary education and navigating through the education system. Highly educated parents might possibly also be more capable of helping their children with their schoolwork, but such effects seem more likely at a lower educational level than in tertiary education. However, even if some levels of parental education are not fully mediated, their effects are greatly reduced in size over time.

Finally, our results show how students at the very top of the social scale are much more likely to get into the more selective universities. For both our social origin variables, the difference between being in the lowest level and the second highest is about the same size as the one between the two highest levels. In other words, even if social origin has become less important for the last cohort, the individuals with the most socially advantageous origins are still much more likely to graduate from the best universities. Such results are in alignment with the finding of Britton that these universities, whose graduates face great prospects in the labor markets, accept very few socio-economically disadvantaged individuals (Britton et al., 2021). This highlights how, even if social gaps in educational attainment are vastly decreased, inequality in where individuals complete a degree does not necessarily have to disappear. Since we do not have data on how social origins affect Russell Group attendance for the older cohorts, we cannot know for certain if these gaps have decreased over time, even if it seems reasonable to expect such a change. Further study on this topic therefore seems warranted, especially on whether, in a world where more individuals complete an undergraduate degree, social origin is still an important determinant of where one completes an undergraduate degree, and thus how attractive one will become in the labor market afterwards. Additionally, it would also be interesting to further investigate why the UK university degree completion rate has increased. Denning (2022) has concluded that the US, also experiencing a surge in the share of degree holders, could seek root causes in relaxed grading standards at universities. It would also be useful to investigate whether social origin comes into effect at the subject level since the choice of university subject also significantly affects individuals' life-time earnings (Britton et al., 2021). Finally, once new data from other cohort studies is released, future studies extending the work of us and Bukodi and Goldthorpe would further facilitate the understanding of this topic.

Table 1. Per variable missingness analysis across cohorts.

	1958	**1958 %**	**1970**	**1970 %**	**1990**	**1990 %**
Initial Sample	18558	100%	17196	100%	15770	100%
Sex	18554	99.98%	11238	65.35%	15770	100%
Social Class	16476	88.78%	15773	91.72%	15133	95.96%
Parental Education	11548	62.23%	17148	99.72%	11627	73.73%
Undergraduate Degree Attainment	10089	54.36%	7825	45.5%	4701	29.81%

Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 2. Cumulative missingness analysis across cohorts.

	Observations in Final Sample	**Share of Observations Left in Final Sample**
1958	6870	37.02%
1970	7322	42.58%
1990	3622	22.97%

Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 3. Share of observations in the initial and final samples as well as t-statistics for the differences in shares across all cohorts.

Variables	**1958 Initial Sample**	**1958 Final Sample**	**1958 t-value**	**1970 Initial Sample**	**1970 Final Sample**	**1970 t-value**	**1990 Initial Sample**	**1990 Final Sample**	**1990 t-value**
Social Class									
Unskilled	9.8%	6.9%	7.3***	7%	4.4%	7.6***	22.7%	14%	11.7***
Partly Skilled	12.1%	10.9%	2.7***	15.7%	12.9%	5.4***	19.4%	17.1%	3.1***
Skilled	60.6%	61.2%	-0.83	60%	61%	-1.5	29.3%	29.9%	0.8
Managerial/Technical	12.9%	15.7%	-5.6***	12.1%	14.5%	-5.1***	23.1%	30.6%	-9.5***
Professional	4.5%	5.4%	-2.9***	5.2%	7.1%	-5.8***	5.5%	8.4%	-6.5***
Parental Education									
Below Lower Secondary	40.2%	36.7%	4.8***	32.7%	29%	5.8***	23.2%	19.1%	5.2***
Lower Secondary	47.6%	50.1%	-3.3***	44.2%	40.5%	5.4***	18.5%	17.6%	1.2
Upper Secondary	5.9%	6.8%	-2.3**	11%	13.4%	-5.2***	27.2%	27.6%	-0.43
Degree	6.2%	6.4%	-0.56	12.1%	17.2%	-10.8***	31%	35.7%	-5.2***
Female	48.3%	51.4%	-4.3***	50.9%	52.2%	-1.7*	49.1%	55.6%	-7.0***

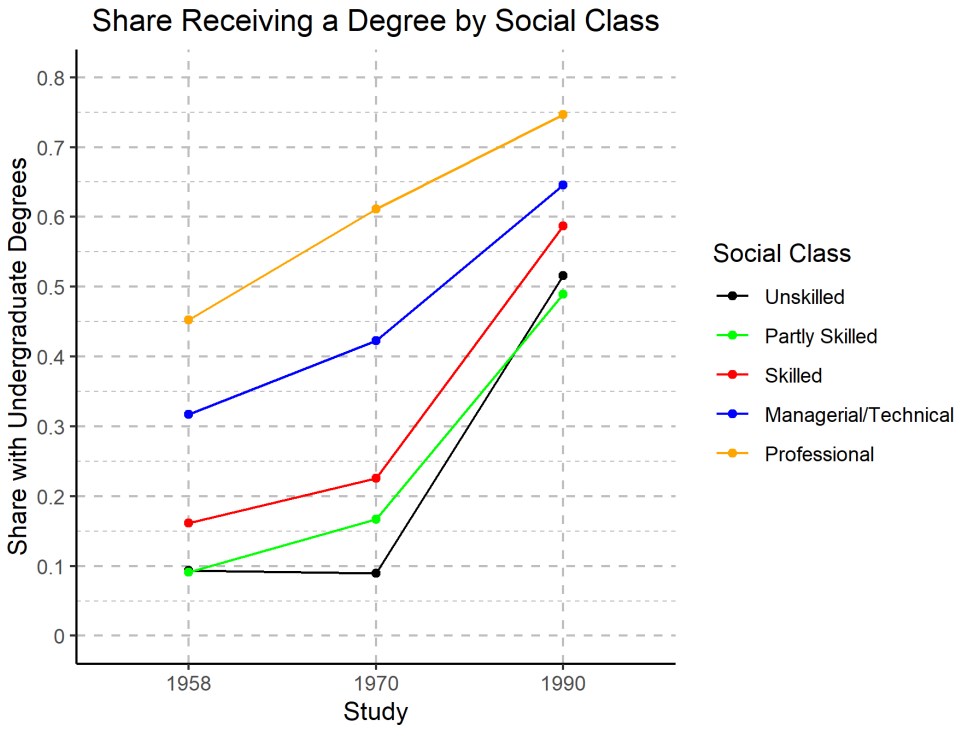
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 4. Distribution of key variables across cohorts.

Variables	**1958**	**1970**	**1990**
Undegraduate Degree Holders	18.91%	26.78%	59.19%
Female	51.35%	52.23%	55.6%
Social Class			
Unskilled	6.86%	4.41%	13.97%
Partly Skilled	10.87%	12.95%	17.09%
Skilled	61.16%	61.04%	29.9%
Managerial/Technical	15.71%	14.5%	30.65%
Professional	5.4%	7.1%	8.39%
Parental Education			
Below Lower Secondary	36.68%	28.95%	19.08%
Lower Secondary	50.1%	40.45%	17.64%
Upper Secondary	6.78%	13.38%	27.61%
Parent Degree	6.43%	17.21%	35.67%

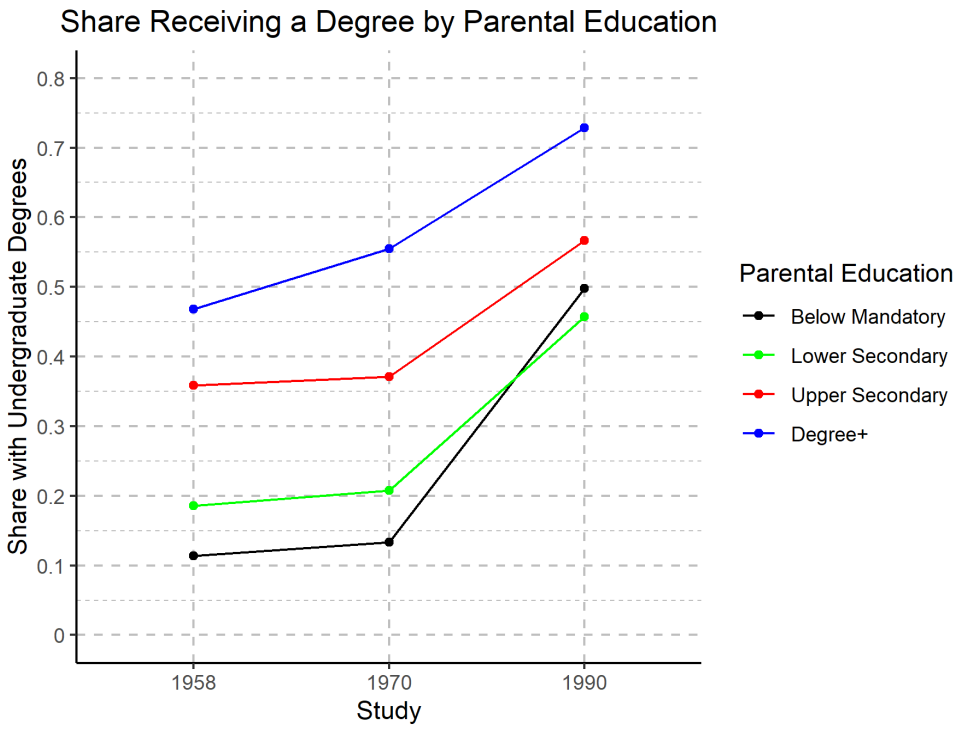
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 4. Share of cohort members receiving an undergraduate degree by Social Class over time.



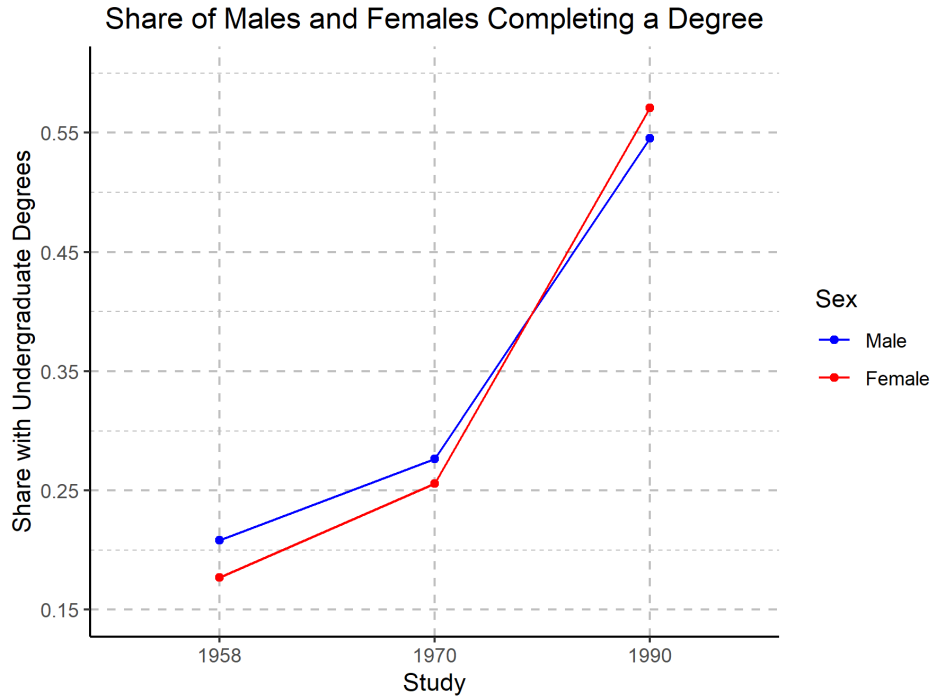
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 5. Share of cohort members receiving an undergraduate degree by Parental Education over time.



Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 6. Share of cohort members receiving an undergraduate degree by Cohort Member Sex over time.



Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 5. (1) Inverse probability weighted LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education with 1970 cohort group, Unskilled and Below Lower Secondary as baseline (2) Male subgroup (3) Female subgroup.

LPM Results

	Dependent Variable: Undegraduate Degree Attainment		
	All	Males	Females
	(1)	(2)	(3)
1958	-0.039***	-0.028***	-0.051***
	(0.007)	(0.010)	(0.009)
1990	0.258***	0.230***	0.285***
	(0.011)	(0.016)	(0.014)
Partly Skilled	-0.022	0.010	-0.051**
	(0.017)	(0.025)	(0.022)
Skilled	0.025*	0.050**	0.003
	(0.014)	(0.021)	(0.019)
Managerial/Technical	0.094***	0.119***	0.072***
	(0.017)	(0.025)	(0.022)
Professional	0.181***	0.180***	0.187***
	(0.021)	(0.032)	(0.028)
Lower Secondary	0.033***	0.048***	0.017
	(0.008)	(0.012)	(0.011)
Upper Secondary	0.139***	0.162***	0.115***
	(0.014)	(0.021)	(0.018)
Degree	0.267***	0.294***	0.239***
	(0.014)	(0.020)	(0.018)
Constant	0.147***	0.120***	0.173***
	(0.015)	(0.021)	(0.019)
Observations	17,814	8,448	9,366
R^2	0.199	0.188	0.214
Adjusted R^2	0.199	0.187	0.214
Residual Std. Error	0.571 (df = 17804)	0.596 (df = 8438)	0.547 (df = 9356)
F Statistic	491.686*** (df = 9; 17804)	216.682*** (df = 9; 8438)	283.655*** (df = 9; 9356)
Note:			*p**p***p<0.01

Table 6. (1) Inverse probability weighted LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education for 1958 cohort (2) Male subgroup 1958 cohort (3) Female subgroup 1958 cohort (4) LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education for 1970 cohort (5) Male subgroup 1970 cohort (6) Female subgroup 1970 cohort (7) LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education for 1990 cohort (8) Male subgroup 1990 cohort (9) Female subgroup 1990 cohort.

	Dependent Variable: Undegraduate Degree Attainment								
	1958	1970	1990	1958 M	1970 M	1990 M	1958 F	1970 F	1990 F
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Partly Skilled	-0.004	0.063***	-0.041**	0.040	0.089***	-0.021	-0.049**	0.035	-0.054
	(0.017)	(0.020)	(0.017)	(0.026)	(0.029)	(0.046)	(0.023)	(0.028)	(0.040)
Skilled	0.054***	0.073***	0.044***	0.077***	0.087***	0.070*	0.028	0.059**	0.023
	(0.015)	(0.017)	(0.015)	(0.020)	(0.024)	(0.043)	(0.021)	(0.025)	(0.036)
Managerial/Technical	0.155***	0.187***	0.051***	0.168***	0.191***	0.089**	0.140***	0.183***	0.020
	(0.020)	(0.023)	(0.020)	(0.028)	(0.032)	(0.043)	(0.028)	(0.032)	(0.038)
Professional	0.239***	0.262***	0.123***	0.297***	0.226***	0.118**	0.185***	0.302***	0.141***
	(0.031)	(0.030)	(0.031)	(0.045)	(0.042)	(0.054)	(0.042)	(0.043)	(0.047)
Lower Secondary	0.053***	0.063***	-0.056***	0.083***	0.075***	-0.067	0.022*	0.052***	-0.048
	(0.009)	(0.011)	(0.009)	(0.014)	(0.016)	(0.042)	(0.013)	(0.015)	(0.037)
Upper Secondary	0.174***	0.202***	0.053**	0.204***	0.220***	0.072*	0.145***	0.187***	0.030
	(0.024)	(0.018)	(0.024)	(0.035)	(0.026)	(0.039)	(0.033)	(0.024)	(0.034)
Degree	0.258***	0.329***	0.198***	0.235***	0.350***	0.230***	0.275***	0.306***	0.162***
	(0.026)	(0.019)	(0.026)	(0.038)	(0.027)	(0.037)	(0.036)	(0.027)	(0.034)
Constant	0.065***	0.061***	0.483***	0.043**	0.049**	0.430***	0.089***	0.072***	0.532***
	(0.015)	(0.016)	(0.015)	(0.020)	(0.022)	(0.039)	(0.021)	(0.023)	(0.033)
Observations	6,870	7,322	3,622	3,342	3,498	1,608	3,528	3,824	2,014
R^2	0.077	0.122	0.055	0.080	0.119	0.074	0.081	0.127	0.041
Adjusted R^2	0.076	0.121	0.053	0.078	0.117	0.070	0.079	0.125	0.037
Residual Std. Error	0.461 (df = 6862)	0.486 (df = 7314)	0.846 (df = 3614)	0.488 (df = 3334)	0.501 (df = 3490)	0.902 (df = 1600)	0.432 (df = 3520)	0.471 (df = 3816)	0.797 (df = 2006)
F Statistic	82.033*** (df = 7; 6862)	144.890*** (df = 7; 7314)	30.057*** (df = 7; 3614)	41.226*** (df = 7; 3334)	67.264*** (df = 7; 3490)	18.236*** (df = 7; 1600)	44.193*** (df = 7; 3520)	79.084*** (df = 7; 3816)	12.151*** (df = 7; 2006)

Note: *p**p***p<0.01

Table 7. Inverse probability weighted mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Levels for 1958 cohort. The regression baseline is Unskilled and Below Lower Secondary. (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1958

	Degree		A-Levels	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
Partly Skilled	-0.004		0.028	-0.020
	(0.017)		(0.023)	(0.012)
Skilled	0.054***		0.100***	-0.003
	(0.015)		(0.019)	(0.010)
Managerial/Technical	0.155***		0.260***	0.007
	(0.020)		(0.024)	(0.014)
Professional	0.239***		0.380***	0.023
	(0.031)		(0.032)	(0.024)
Lower Secondary	0.053***		0.082***	0.006
	(0.009)		(0.011)	(0.007)
Upper Secondary	0.174***		0.274***	0.018
	(0.024)		(0.025)	(0.019)
Degree	0.258***		0.329***	0.070***
	(0.026)		(0.026)	(0.021)
A-Levels		0.583***		0.570***
		(0.011)		(0.011)
Constant	0.065***	-0.000***	0.118***	-0.002
	(0.015)	(0.000)	(0.018)	(0.009)
Observations	6,870	6,870	6,870	6,870
R^2	0.077	0.487	0.114	0.490
Adjusted R^2	0.076	0.487	0.113	0.489
Residual Std. Error	0.461 (df = 6862)	0.344 (df = 6868)	0.541 (df = 6862)	0.343 (df = 6861)
F Statistic	82.033*** (df = 7; 6862)	6,518.529*** (df = 1; 6868)	126.012*** (df = 7; 6862)	822.521*** (df = 8; 6861)
Note:				*p**p***p<0.01

Table 8. Inverse probability weighted mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Levels for 1970 cohort. The regression baseline is Unskilled and Below Lower Secondary. (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1970

	Degree		A-Levels	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
Partly Skilled	0.063***		0.085***	0.001
	(0.020)		(0.023)	(0.011)
Skilled	0.073***		0.113***	-0.009
	(0.017)		(0.020)	(0.010)
Managerial/Technical	0.187***		0.228***	0.021
	(0.023)		(0.025)	(0.013)
Professional	0.262***		0.297***	0.046**
	(0.030)		(0.031)	(0.018)
Lower Secondary	0.063***		0.083***	0.003
	(0.011)		(0.012)	(0.007)
Upper Secondary	0.202***		0.232***	0.034***
	(0.018)		(0.019)	(0.011)
Degree	0.329***		0.367***	0.062***
	(0.019)		(0.019)	(0.012)
A-Levels		0.749***		0.727***
		(0.009)		(0.010)
Constant	0.061***	-0.000***	0.092***	-0.007
	(0.016)	(0.000)	(0.019)	(0.009)
Observations	7,322	7,322	7,322	7,322
R^2	0.122	0.665	0.124	0.670
Adjusted R^2	0.121	0.665	0.123	0.669
Residual Std. Error	0.486 (df = 7314)	0.300 (df = 7320)	0.528 (df = 7314)	0.298 (df = 7313)
F Statistic	144.890*** (df = 7; 7314)	14,514.290*** (df = 1; 7320)	147.862*** (df = 7; 7314)	1,852.850*** (df = 8; 7313)
Note:				*p**p***p<0.01

Table 9. Inverse probability weighted mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Levels for 1990 cohort. The regression baseline is Unskilled and Below Lower Secondary. (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1990

	Degree		A-Levels	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
Partly Skilled	-0.049		-0.036	-0.035
	(0.036)		(0.036)	(0.034)
Skilled	0.059*		0.090***	0.025
	(0.033)		(0.032)	(0.031)
Managerial/Technical	0.045		0.075**	0.017
	(0.033)		(0.033)	(0.032)
Professional	0.089**		0.103***	0.050
	(0.041)		(0.040)	(0.040)
Lower Secondary	-0.074**		0.026	-0.083***
	(0.033)		(0.033)	(0.031)
Upper Secondary	0.034		0.064**	0.010
	(0.030)		(0.030)	(0.028)
Degree	0.152***		0.193***	0.081***
	(0.029)		(0.028)	(0.028)
A-Levels		0.399***		0.373***
		(0.020)		(0.021)
Constant	0.560***	0.358***	0.559***	0.352***
	(0.030)	(0.017)	(0.030)	(0.031)
Observations	2,736	2,736	2,736	2,736
R^2	0.042	0.146	0.049	0.163
Adjusted R^2	0.039	0.145	0.046	0.160
Residual Std. Error	0.960 (df = 2728)	0.906 (df = 2734)	0.914 (df = 2728)	0.898 (df = 2727)
F Statistic	16.938*** (df = 7; 2728)	465.957*** (df = 1; 2734)	20.000*** (df = 7; 2728)	66.262*** (df = 8; 2727)
Note:				*p**p***p<0.01

Table 10. Summary of mediation effects.

	1958	1970	1990
Parent Education			
Lower Secondary	Full mediation	Full mediation	No mediation
Upper Secondary	Full mediation	Partial mediation	Insignificant
Degree	Partial mediation	Partial mediation	Partial mediation
Social Class			
Partly Skilled	Insignificant	Full mediation	Insignificant
Skilled	Full mediation	Full mediation	Full mediation
Managerial/Technical	Full mediation	Full mediation	Insignificant
Professional	Full mediation	Partial mediation	Full mediation

Note: Authors' summary.

Table 11. (1) LPM regression of Russell Group Degree Attainment on Social Class and Parental Education with 1970 cohort group, Unskilled and Below Lower Secondary as baseline (2) Male subgroup (3) Female subgroup.

Russell Group Results 1990

	Dependent Variable: Undegraduate Degree from Russell Group University		
	All	Males	Females
	(1)	(2)	(3)
Partly Skilled	0.004	0.020	-0.015
	(0.041)	(0.071)	(0.047)
Skilled	0.036	0.025	0.046
	(0.037)	(0.063)	(0.043)
Managerial/Technical	0.089**	0.080	0.093*
	(0.040)	(0.066)	(0.048)
Professional	0.164***	0.134	0.194***
	(0.054)	(0.087)	(0.066)
Lower Secondary	-0.0001	-0.005	0.007
	(0.039)	(0.064)	(0.046)
Upper Secondary	0.087**	0.133**	0.046
	(0.036)	(0.060)	(0.043)
Degree	0.166***	0.145**	0.192***
	(0.036)	(0.058)	(0.046)
Constant	0.119***	0.136**	0.102***
	(0.033)	(0.054)	(0.038)
Observations	1,408	600	808
R^2	0.052	0.042	0.072
Adjusted R^2	0.047	0.030	0.063
Residual Std. Error	1.227 (df = 1400)	1.381 (df = 592)	1.098 (df = 800)
F Statistic	10.940*** (df = 7; 1400)	3.666*** (df = 7; 592)	8.807*** (df = 7; 800)
Note:			*p**p***p<0.01

References

American Economic Association N.D., N.D.-last update, *What is economics?*. Available: https://www.aeaweb.org/resources/students/what-is-economics [2023.04.17].

Bell, A., Chetty, R., Jaravel, X., Petkova, N. & Van Reenen, J. 2019, ”Who Becomes an Inventor in America? The Importance of Exposure to Innovation”, *The Quarterly journal of economics*, vol. 134, no.2, pp. 647-713.

Bowman, M.J. 1991, "Educational Inequalities and Opportunity in Economic Perspective", *Oxford review of education*, vol. 17, no. 2, pp. 189-209.

Breen, R. 2022, *The stubborn persistence of educational inequality*, IFS Deaton Review of Inequalities.

Britton, J. 2020, *The impact of undergraduate degrees on lifetime earnings*, Institute for Fiscal Studies.

Britton, J., Drayton, E. & Van Der Erve, L. 2021, *Which university degrees are best for intergenerational mobility?*, Institute for Fiscal Studies.

Britton, J., Van Der Erve, L., Belfield, C., Vignoles, A., Dickson, M., Zhu, Y., Walker, I., Dearden, L., Sibieta, L. & Buscha, F. 2021, *How much does degree choice matter?*, Institute for Fiscal Studies.

Bukodi, E., Bourne, M. & Betthäuser, B. 2017, "Wastage of talent?Social origins, cognitive ability and educational attainment in Britain", *Advances in Life Course Research*, vol. 34, pp. 34-42.

Bukodi, E., Erikson, R. & Goldthorpe, J.H. 2014, "The effects of social origins and cognitive ability on educational attainment: Evidence from Britain and Sweden", *Acta sociologica*, vol. 57, no. 4, pp. 293-310.

Bukodi, E. & Goldthorpe, J.H. 2013, "Decomposing 'Social Origins': The Effects of Parents' Class, Status, and Education on the Educational Attainment of Their Children", *European sociological review*, vol. 29, no. 5, pp. 1024-1039.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. & Vignoles, A. 2013, "Widening participation in higher education: analysis using linked administrative data", *Journal of the Royal Statistical Society. Series A, Statistics in society,* vol. 176, no. 2, pp. 431-457.

Denning, J.T., Eide, E.R., Mumford, K.J., Patterson, R.W. & Warnick, M. 2022, "Why Have College Completion Rates Increased?", *American Economic Journal: Applied Economics*, vol. 14, no. 3, pp. 1-29.

Farquharson, C., Mcnally, S. & Tahir, I. 2022, *Education inequalities*, IFS Deaton Review of Inequalities.

Galindo-Rueda, F. & Vignoles, A. 2005, "The declining relative importance of ability in predicting educational attainment", *The Journal of human resources,* vol. 40, no. 2, pp. 335-353.

Metten, M., Costet, N., Multigner, L., Viel, J. & Chauvet, G. 2022, "Inverse probability weighting to handle attrition in cohort studies: some guidance and a call for caution", *BMC medical research methodology,* vol. 22, no. 1, pp. 45.

Michaillat, P. 2012, "Do Matching Frictions Explain Unemployment? Not in Bad Times", *American Economic Review,* vol. 102, no. 4, pp. 1721-1750.

OECD 2022, *Review education policies*.
Available: https://gpseducation.oecd.org/revieweducationpolicies/#!node=41761&filter=all [2023.04.09].

Office for National Statistics N.D., N.D.-last update, *The National Statistics Socio-economic classification (NS-SEC)*.
Available: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 [2023.03.13].

Pearson N.D., N.D.-last update, *About O levels*.
Available: https://qualifications.pearson.com/en/support/support-topics/understanding-our-qualifications/our-qualifications-explained/about-o-levels.html [2023.03.22].

Rose, D. & Pevalin, D.J. 2001, *The National Statistics Socio-economic Classification: unifying official and sociological approaches to the conceptualisation and measurement of social class*, University of Essex, Institute for Social and Economic Research (ISER), Colchester.

Russell Group N.D., N.D.-last update, *About*. Available: https://russellgroup.ac.uk/about/ [2023.04.19].

Study in UK N.D., N.D.-last update, *Russell Group Universities*. Available: https://www.studyin-uk.com/uk-study-info/russell-group-universities/ [2023.04.19].

Sullivan, A., Parsons, S., Wiggins, R., Heath, A. & Green, F. 2014, "Social origins, school type and higher education destinations", *Oxford Review of Education*, vol. 40, no. 6, pp. 739-763.

The International Labour Organization N.D., N.D.-last update, *Q&As on business, discrimination and equality*. Available: https://www.ilo.org/empent/areas/business-helpdesk/WCMS_DOC_ENT_HLP_BDE_FAQ_EN/lang--en/index.htm [2023.04.09].

Data sources

Bukodi, E. (2017). National Child Development Study and 1970 British Cohort Study Educational Qualifications Histories, 1981-2009. [data collection]. University of London, Institute of Education, Centre for Longitudinal Studies, [original data producer(s)]. University of London, Institute of Education, Centre for Longitudinal Studies. SN: 8127, DOI: http://doi.org/10.5255/UKDA-SN-8127-1

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2023). National Child Development Study. [data series]. 12th Release. UK Data Service. SN: 2000032, DOI: http://doi.org/10.5255/UKDA-Series-2000032

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2023). Next Steps: Sweeps 1-8, 2004-2016. [data collection]. 16th Edition. UK Data Service. SN: 5545, DOI: http://doi.org/10.5255/UKDA-SN-5545-8

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2023). 1970 British Cohort Study. [data series]. 9th Release. UK Data Service. SN: 200001, DOI: http://doi.org/10.5255/UKDA-Series-200001

Appendices

Appendix 1: Additional Summary Statistics and Graphs

In this appendix, additional statistics of key independent variables by male and female subgroups are presented.

Table 12. Distribution of key variables within the female cohort group.

Variables	**1958**	**1970**	**1990**
Undegraduate Degree Holders	17.09%	25.73%	60.18%
Social Class			
Unskilled	6.92%	4.31%	14.85%
Partly Skilled	11.14%	12.81%	16.98%
Skilled	60.91%	61.72%	30.54%
Managerial/Technical	15.79%	14.57%	30.04%
Professional	5.24%	6.59%	7.6%
Parental Education			
Below Lower Secondary	36.62%	28.53%	19.27%
Lower Secondary	50.31%	41.06%	17.78%
Upper Secondary	6.72%	14.12%	28.7%
Parent Degree	6.35%	16.29%	34.26%

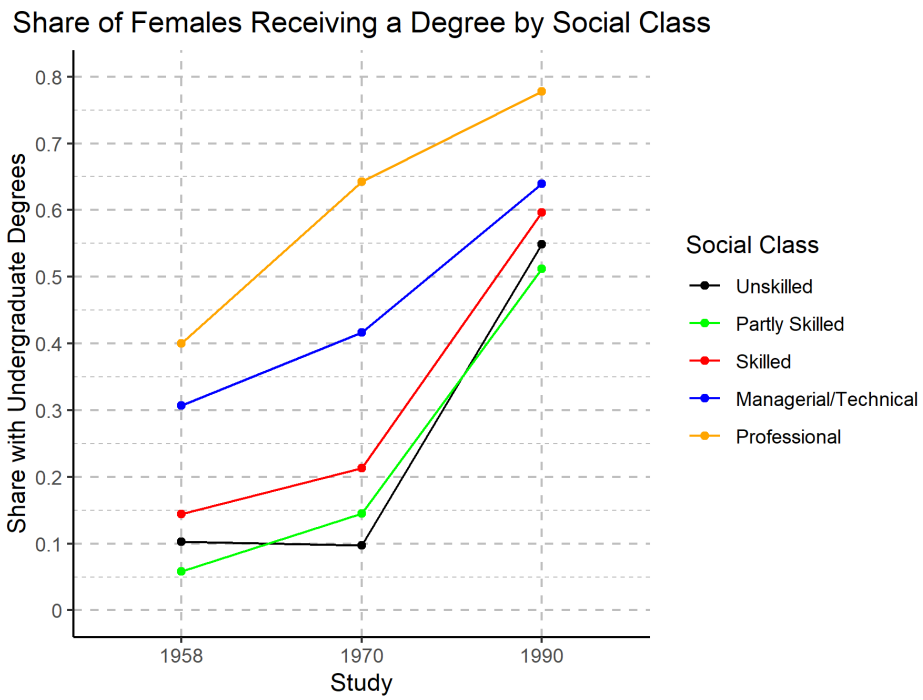
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 13. Distribution of key variables within the male cohort group.

Variables	**1958**	**1970**	**1990**
Undegraduate Degree Holders	20.83%	27.93%	57.96%
Social Class			
Unskilled	6.79%	4.52%	12.87%
Partly Skilled	10.59%	13.09%	17.23%
Skilled	61.43%	60.29%	29.1%
Managerial/Technical	15.62%	14.44%	31.41%
Professional	5.57%	7.66%	9.39%
Parental Education			
Below Lower Secondary	36.74%	29.42%	18.84%
Lower Secondary	49.88%	39.79%	17.48%
Upper Secondary	6.85%	12.58%	26.24%
Parent Degree	6.52%	18.21%	37.44%

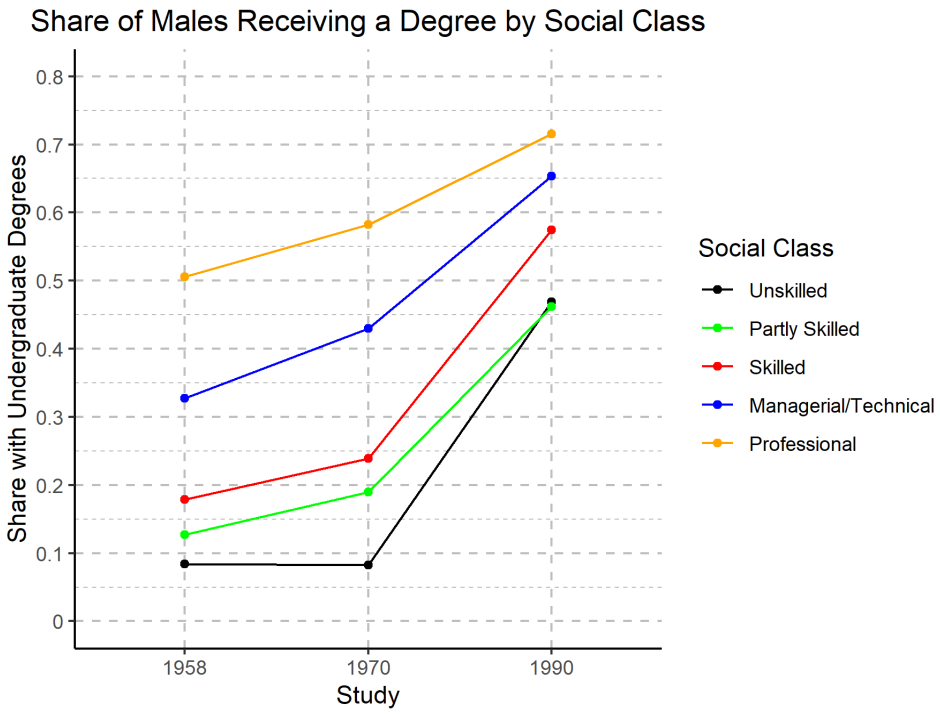
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 7. Share of female cohort members receiving an undergraduate degree by Social Class over time.



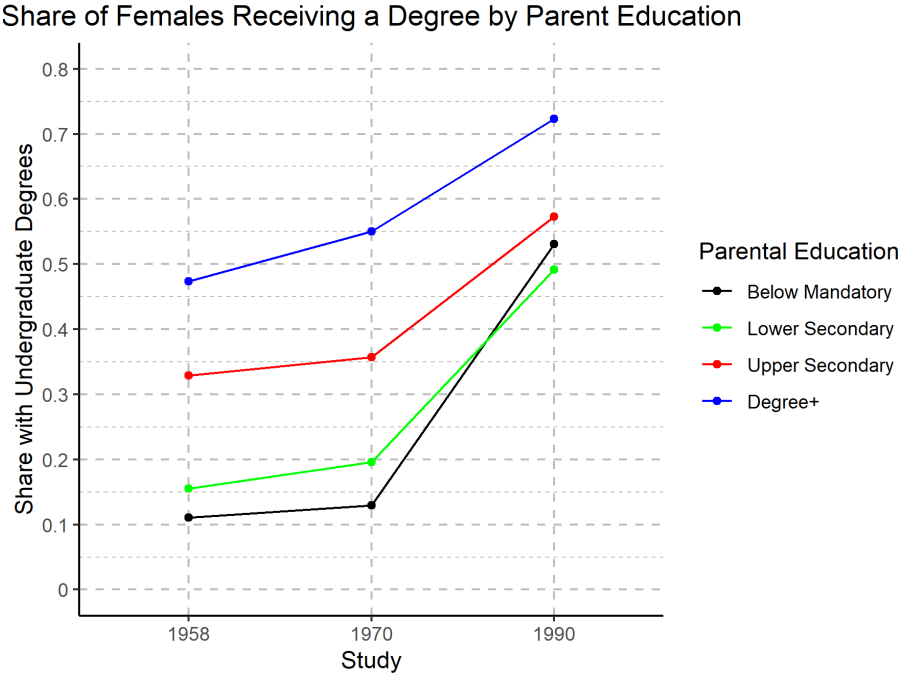
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 8. Share of male cohort members receiving an undergraduate degree by Social Class over time.



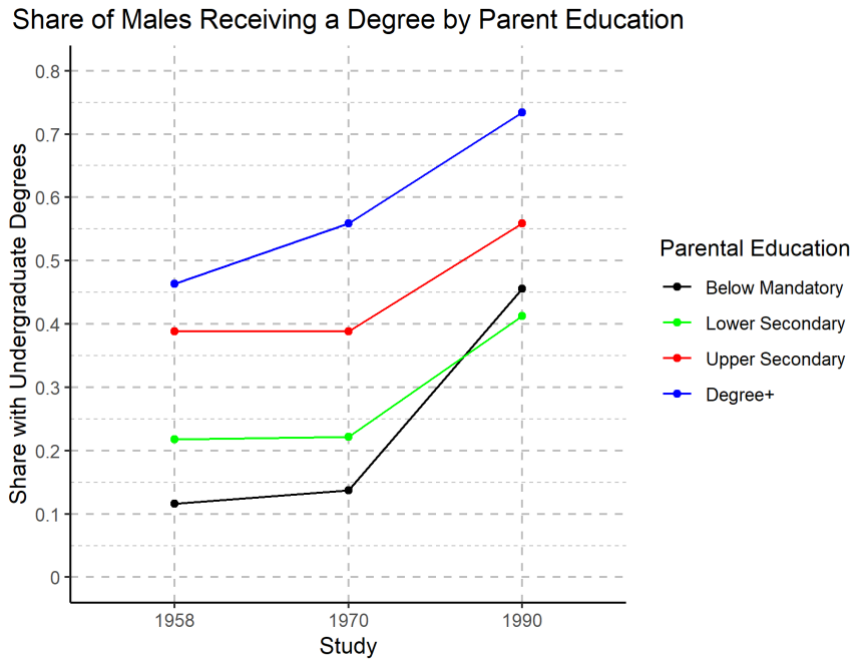
Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 9. Share of female cohort members receiving an undergraduate degree by Parental Education over time.



Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Figure 10. Share of male cohort members receiving an undergraduate degree by Parental Education over time.



Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Appendix 2: Construction of Dependent Variable

Undergraduate Degree Attainment is retrieved from individuals' answers on their highest educational qualification attained. One assumption made in our research is that those who have obtained educational qualifications higher than undergraduate level have obtained an undergraduate degree.

When constructing our dependent variable, several approaches were adopted. For the two earliest cohorts (1958 and 1970) we had both a variable that showed the age when an individual variable left full-time education, and a variable for the highest educational level they have achieved. Even though *Age Left Education* had a larger sample size, we have chosen to only use *Highest Educational Qualification Attained*, as we believe that the age question was misinterpreted by a large number of survey participants. As shown in **Table 14** below, about 40 % of people who do have a degree claim they left education before the age of 19. We acknowledge that it is possible that certain individuals are highly talented and able to achieve advanced educational qualifications at a younger age, and that individuals might have left education for a while before

continuing their studies. However, we do not believe that this is common enough that it affects almost half of the sample. Instead, we believe that a more plausible interpretation is that a large share of the sample interpreted the question at the time they left secondary education, and not HE. Therefore, although using *Highest Educational Qualification Attained* causes more missing data than using *Age Left Education*, we believe that it would be a more reliable measure.

Table 14. Distribution of the age when degree holders left education.

Under 18	18	19	20	21	Over 21
16.6%	19.4%	3%	1.7%	18.2%	41.2%

Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

In the 1990 cohort, the only variable we could use was one that indicated the highest educational qualifications of an individual. As discussed in **Missingness Analysis of Key Variables**, this led to a large drop in the sample size, and suspicions about degree holders being overrepresented:

Out of the 7707 people who participated in this particular wave, only 4769 people answered questions that laid ground for the construction of our undergraduate degree variable. If we assume that all of the participants who did not answer the question lacked degrees, our results would be more in line with the available data, but we have no proof of the viability of such an assumption, and have not used the same approach when constructing other variables.

Furthermore, there is a variable for whether an individual has attended university or not that could possibly be used to solve this problem, but we have serious concerns regarding its reliability. For instance, it shows how about 15 % of our degree holders did not attend university at all, and that 40 % of our non-degree holders did. Using such a variable to approximate the result for our missing individuals was therefore deemed inappropriate.

Appendix 3: Construction of *Parental Education*

Parental education is measured differently across cohorts. In the 1958 study, the best proxy for parental education is *Parent Age Left Education* which indicates the age at which they left full time education. On the other hand, the 1970 and the 1990 studies have variables for both *Parent*

Age Left Education and *Parent Highest Educational Qualification Attained*. For these two cohorts, we have chosen to use *Parent Highest Educational Qualification Attained* as our base measure for *Parental Education*. The reasoning behind this choice is similar to choosing the base measure for the dependent variable, namely the risk of misinterpretation.

In individual cases where information on *Parent Highest Educational Qualification Attained* is missing, we have used available information on *Parent Age Left Education* to estimate *Parental Education* which gives a combined measure for *Parental Education*. We acknowledge that this might lead to the same problems as described in **Appendix 2**, but since we have to use Parent Age Left Education in the 1958 study, we judged the benefits of getting a larger sample to be higher in this case.

In addition to this, the data contains educational information for both fathers and mothers. We have chosen to use a dominance approach where the parent with the highest educational qualification is chosen and applied. A strength of such an approach is that it captures cases where the father's educational level is unknown. Alternatively, one could use the father's educational level as a base measurement and only use the mother's educational level if the father's is missing from the data. We have not done this as we do not see why the father's educational level should have a dominant effect on the child's education level, even when the mother has a much higher educational level. Nevertheless, the result is almost identical.

Another approach would be to group individuals based on if one or both parents have a certain educational qualification (Bukodi, Bourne, & Betthäuser, 2017). We have chosen not to do this, as the educational level of mothers in 1958 is not the same type of indicator as it is in 1990. Since very few mothers were educated in 1958, we believe that such an approach would bias the parent educational level of that cohort downwards.

Finally, it could be argued that the variable *Parental Education* could have been treated in relative terms, meaning that we look at how educated a parent is compared to the other parents in a sample. The BG paper has stated that a relative treatment would facilitate comparisons across cohorts as the UK educational system has undergone substantial changes. While we agree that

such a treatment would be more suitable when comparing different education levels over time, we believe that an absolute treatment would yield additional insights on how parental education comes into effect via the channel of formal educational classifications.

The conversion from *Parent Age left Education* to *Parental Education* is done as shown in **Table 15**:

Table 15. Conversion of Parent Age Left Education to Parental Education.

Retrieved from data		Constructed
Parent Age Left Education	***Parent Highest Educational Qualification Attained***	***Parental Education***
1958		
14 years of age or younger	-	Below secondary
15 to 17 years of age	-	Lower Secondary
18 to 20 years of age	-	Upper Secondary
21 years of age or older	-	Degree
1970		
14 years of age or younger	No qualifications or other qualifications	Below secondary
15 to 17 years of age	Vocational qualifications or O-Level	Lower Secondary
18 to 20 years of age	A-level, SRN or Certification of Education	Upper Secondary
21 years of age or older	Degree or higher	Degree
1990		
15 years of age or younger	Level 1 or below, other qualifications or no qualifications	Below secondary
16 to 17 years of age	GCSE grades A-C or equal	Lower Secondary
18 to 20 years of age	GCE A level or equivalent or higher education below degree level	Upper Secondary
21 years of age or older	Degree or equivalent	Degree

Note: Authors' summary.

Regarding Mandatory Schooling

Since the parents of members of both the 1958 and 1970 cohorts went to school before the lowest school leaving age increased from 15 to 16 in 1973 (Sullivan, 2014), we have chosen to code our *Below Lower Secondary* variable as below 15 instead of below 16. This is because we have defined the first level as leaving before or at the lowest mandatory level, and believe that it is appropriate to put people in the second category if they have made the active choice of not leaving school. Since the school system in the UK changed quite a bit during the course of the 30 years that our studies cover, and our *Parental Education* variable therefore should not be seen as directly comparable between cohorts, we believe that such a choice is less problematic. Nevertheless, it does have large consequences on the distribution of the first two social classes in both cohorts.

Table 16. Comparison between setting the age limit at 14 or 15 for Below Lower Secondary for the 1958 cohort.

Parental Education	With 14 Years Mandatory Education	With 15 Years Mandatory Education
Below Lower Secondary	36.68%	63.74%
Lower Secondary	50.1%	23.04%
Upper Secondary	6.78%	6.78%
Degree	6.43%	6.43%

Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

Table 17. Comparison between setting the age limit at 14 or 15 for Below Lower Secondary for the 1970 cohort.

Parental Education	With 14 Years Mandatory Education	With 15 Years Mandatory Education
Below Lower Secondary	34.72%	28.95%
Lower Secondary	34.69%	40.45%
Upper Secondary	13.38%	13.38%
Degree	17.21%	17.21%

Note: Data from The Centre for Longitudinal Studies. NCDS survey (2023), BCS survey (2023) and NS survey (2023).

As shown in **Table 16** and **Table 17**, this has the effect of increasing the share of *Lower Secondary* and decreases the share of the *Below Lower Secondary*.

Appendix 4: Construction of *Social Class*

Similarly, *Parental Educational*, *Social Class* has also been measured by different scales across cohorts. Both 1958 and 1970 cohort studies have used membership of social classes based on Occupation (SC) to classify the *Social Class*. SC consists of 5 different levels that are based on the occupational status of the parent:

1. Professional, etc occupations
2. Managerial and Technical occupations
3. Skilled occupations, both non-manual and manual
4. Partly skilled occupations
5. Unskilled occupations

For the 1990 cohort, *Parent Social Class* is measured by National Statistics Social-economic Classification (NS-SEC). NS-SEC is a new system for classifying socio-economic status which replaces two previous socio-economic classifications: SC and Socio-economic Groups SEG (Office for National Statistics, N.D.). This classification divides individuals into 8 different classes based on the characteristics of their occupations:

1. Higher managerial and professional occupations
2. Lower managerial and professional occupations
3. Intermediate occupations (clerical, sales, service)
4. Small employers and own account workers
5. Lower supervisory and technical occupations
6. Semi-routine occupations
7. Routine occupations
8. Never worked or long-term unemployed

We have chosen to convert NS-SEC classes to the classification used in 1958 and 1970 cohorts to make the 1990 cohort comparable. The conversion is based on information retrieved from Rose & Pevalin (2001) and we have chosen to assign NS-SEC Class 8 to SC Class 5, as shown in **Table 18**:

Table 18. Conversion of NS-SEC Class 8 to SC Class 5.

1990 cohort (NS-SEC)	**1958 and 1970 cohorts (SC)**
Class 1	Class 1 (Unskilled)
Class 2	Class 2 (Partly Skilled)
Class 3, 4 and 5	Class 3 (Skilled)
Class 6	Class 4 (Managerial/Technical)
Class 7, 8	Class 5 (Professional)

Note: Authors' summary.

After doing the conversion as shown above, we were able to construct our five Social Class levels:

1. Unskilled
2. Partly Skilled
3. Skilled
4. Managerial/Technical
5. Professional

It is also worth mentioning that the 1990 study starts following its cohort members at cohort member age 14, compared to at birth for the two previous cohort studies. Therefore, we have imposed an assumption that *Social Class* is a time-constant variable in our study. Additionally, we have used the father's social class in this instance instead of the mother's social class, as family members usually have the same social class and many mothers were not working during the earlier studies.

Appendix 5: Main Regression Results without IPW

We have also performed regressions on the data without using Inverse Probability Weighting. The result is shown below, but few notable changes in direction occur. Our conclusions therefore do not seem to rely on whether we use IPW or not.

Table 19. (1) LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education with 1970 cohort group, Unskilled and Below Lower Secondary as baseline (2) Male subgroup (3) Female subgroup.

LPM Results

	Dependent variable:		
	Degree		
	LPM	Males	Females
	(1)	(2)	(3)
1958	-0.040***	-0.028***	-0.051***
	(0.007)	(0.010)	(0.009)
1990	0.249***	0.220***	0.271***
	(0.010)	(0.015)	(0.014)
Class 2	-0.008	0.031	-0.039**
	(0.013)	(0.020)	(0.018)
Class 3	0.035***	0.065***	0.012
	(0.012)	(0.016)	(0.016)
Class 4	0.120***	0.149***	0.098***
	(0.014)	(0.020)	(0.019)
Class 5	0.209***	0.214***	0.207***
	(0.019)	(0.027)	(0.026)
Parental Education 2	0.043***	0.063***	0.025***
	(0.007)	(0.010)	(0.009)
Parental Education 3	0.161***	0.186***	0.139***
	(0.012)	(0.018)	(0.016)
Parental Education 4	0.280***	0.304***	0.261***
	(0.012)	(0.018)	(0.017)
Constant	0.128***	0.094***	0.155***
	(0.012)	(0.017)	(0.017)
Observations	17,850	8,448	9,366
R^2	0.187	0.172	0.203
Adjusted R^2	0.186	0.171	0.202
Residual Std. Error	0.415 (df = 17840)	0.421 (df = 8438)	0.409 (df = 9356)
F Statistic	455.242*** (df = 9; 17840)	194.206*** (df = 9; 8438)	265.041*** (df = 9; 9356)
Note:			*p**p***p<0.01

Table 20. (1) LPM regression of Undergraduate Degree Attainment on Parent Social Class and Parental Education with 1970 cohort group as baseline and cohort interaction terms (2) Male subgroup (3) Female subgroup.

LPM With Interaction Term Results

	Dependent variable:		
		Degree	
	LPM	Males	Females
	(1)	(2)	(3)
1958	0.002	-0.008	0.014
	(0.022)	(0.030)	(0.031)
1990	0.430***	0.379***	0.459***
	(0.030)	(0.045)	(0.040)
Class 2	0.061***	0.088***	0.036
	(0.020)	(0.029)	(0.028)
Class 3	0.076***	0.093***	0.061**
	(0.017)	(0.024)	(0.025)
Class 4	0.194***	0.201***	0.189***
	(0.023)	(0.032)	(0.032)
Class 5	0.269***	0.235***	0.307***
	(0.030)	(0.042)	(0.043)
Parental Education 2	0.061***	0.072***	0.051***
	(0.011)	(0.016)	(0.015)
Parental Education 3	0.200***	0.217***	0.186***
	(0.018)	(0.026)	(0.024)
Parental Education 4	0.324***	0.345***	0.303***
	(0.019)	(0.027)	(0.027)
1958 *Class 2	-0.064**	-0.044	-0.082**
	(0.026)	(0.039)	(0.036)
1958 *Class 3	-0.018	-0.010	-0.029
	(0.023)	(0.032)	(0.033)
1958 *Class 4	-0.033	-0.021	-0.046
	(0.030)	(0.043)	(0.043)
1958 *Class 5	-0.028	0.069	-0.119**
	(0.043)	(0.061)	(0.060)
1958 * Parental Education 2	-0.009	0.010	-0.027
	(0.014)	(0.021)	(0.019)
1958 *Parental Education 3	-0.027	-0.017	-0.038
	(0.030)	(0.043)	(0.041)

1958 *Parental Education 4	-0.068**	-0.116**	-0.027
	(0.032)	(0.046)	(0.045)
1990 *Class 2	-0.102***	-0.106*	-0.087*
	(0.036)	(0.054)	(0.048)
1990 *Class 3	-0.034	-0.018	-0.037
	(0.032)	(0.049)	(0.044)
1990 *Class 4	-0.146***	-0.108**	-0.166***
	(0.036)	(0.054)	(0.049)
1990 *Class 5	-0.138***	-0.111*	-0.164***
	(0.046)	(0.068)	(0.063)
1990 * Parental Education 2	-0.117***	-0.136***	-0.099**
	(0.030)	(0.044)	(0.040)
1990 *Parental Education 3	-0.148***	-0.141***	-0.155***
	(0.031)	(0.047)	(0.041)
1990 *Parental Education 4	-0.134***	-0.113**	-0.141***
	(0.031)	(0.046)	(0.043)
Constant	0.059***	0.046**	0.071***
	(0.016)	(0.023)	(0.023)
Observations	17,850	8,448	9,366
R^2	0.192	0.176	0.209
Adjusted R^2	0.191	0.174	0.208
Residual Std. Error	0.414 (df = 17826)	0.420 (df = 8424)	0.408 (df = 9342)
F Statistic	183.784*** (df = 23; 17826)	78.448*** (df = 23; 8424)	107.631*** (df = 23; 9342)
Note:			*p**p***p<0.01

Table 21. Mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Level for 1958 cohort (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1958

	Dependent variable:			
	Degree		A_level	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
ParentalEducation2	0.052***		0.082***	0.005
	(0.009)		(0.011)	(0.007)
ParentalEducation3	0.173***		0.274***	0.017
	(0.024)		(0.025)	(0.019)
ParentalEducation4	0.257***		0.329***	0.070***
	(0.026)		(0.026)	(0.020)
class2	-0.002		0.032	-0.020*
	(0.017)		(0.022)	(0.012)
class3	0.058***		0.106***	-0.003
	(0.015)		(0.019)	(0.010)
class4	0.161***		0.269***	0.009
	(0.020)		(0.023)	(0.014)
class5	0.241***		0.384***	0.023
	(0.031)		(0.031)	(0.024)
A_level		0.582***		0.568***
		(0.010)		(0.011)
Constant	0.061***	-0.000***	0.113***	-0.003
	(0.014)	(0.000)	(0.018)	(0.010)
Observations	6,870	6,870	6,870	6,870
R^2	0.080	0.484	0.117	0.487
Adjusted R^2	0.079	0.484	0.116	0.486
Residual Std. Error	0.376 (df = 6862)	0.281 (df = 6868)	0.440 (df = 6862)	0.281 (df = 6861)
F Statistic	84.824*** (df = 7; 6862)	6,447.226*** (df = 1; 6868)	129.871*** (df = 7; 6862)	814.224*** (df = 8; 6861)
Note:				*p**p***p<0.01

Table 22. Mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Level for 1970 cohort (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1970

	Dependent variable:			
	Degree		A_level	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
ParentalEducation2	0.061***		0.081***	0.002
	(0.011)		(0.012)	(0.007)
ParentalEducation3	0.200***		0.230***	0.032***
	(0.018)		(0.019)	(0.011)
ParentalEducation4	0.324***		0.363***	0.059***
	(0.019)		(0.019)	(0.012)
class2	0.061***		0.084***	-0.0002
	(0.020)		(0.023)	(0.012)
class3	0.076***		0.116***	-0.009
	(0.017)		(0.020)	(0.010)
class4	0.194***		0.235***	0.022
	(0.023)		(0.025)	(0.013)
class5	0.269***		0.303***	0.047**
	(0.030)		(0.031)	(0.019)
A_level		0.756***		0.732***
		(0.008)		(0.009)
Constant	0.059***	0.000***	0.091***	-0.008
	(0.016)	(0.000)	(0.019)	(0.010)
Observations	7,322	7,322	7,322	7,322
R^2	0.130	0.667	0.133	0.672
Adjusted R^2	0.129	0.667	0.132	0.671
Residual Std. Error	0.413 (df = 7314)	0.256 (df = 7320)	0.446 (df = 7314)	0.254 (df = 7313)
F Statistic	155.912*** (df = 7; 7314)	14,643.090*** (df = 1; 7320)	159.849*** (df = 7; 7314)	1,870.763*** (df = 8; 7313)
Note:				*p**p***p<0.01

Table 23. Mediation analysis of Undergraduate Degree Attainment on Social Class, Parental Education and A Level for 1990 cohort (1) Model c, (2) Model a, (3) Model b, (4) Model c2.

Mediation Analysis 1990

	Dependent variable:			
	Degree		StudiedALevelsW5	Degree
	c	a	b	c2
	(1)	(2)	(3)	(4)
ParentalEducation2	-0.071**		0.025	-0.080***
	(0.033)		(0.032)	(0.031)
ParentalEducation3	0.034		0.063**	0.011
	(0.030)		(0.029)	(0.028)
ParentalEducation4	0.150***		0.193***	0.077***
	(0.029)		(0.028)	(0.027)
class2	-0.049		-0.019	-0.042
	(0.035)		(0.035)	(0.033)
class3	0.053*		0.098***	0.016
	(0.032)		(0.031)	(0.030)
class4	0.039		0.077**	0.011
	(0.032)		(0.032)	(0.031)
class5	0.094**		0.109***	0.053
	(0.039)		(0.038)	(0.038)
StudiedALevelsW5		0.401***		0.373***
		(0.020)		(0.020)
Constant	0.566***	0.364***	0.556***	0.358***
	(0.029)	(0.017)	(0.029)	(0.031)
Observations	2,736	2,736	2,736	2,736
R^2	0.043	0.145	0.051	0.162
Adjusted R^2	0.040	0.144	0.048	0.159
Residual Std. Error	0.468 (df = 2728)	0.441 (df = 2734)	0.442 (df = 2728)	0.438 (df = 2727)
F Statistic	17.335*** (df = 7; 2728)	462.194*** (df = 1; 2734)	20.906*** (df = 7; 2728)	65.805*** (df = 8; 2727)
Note:				*p**p***p<0.01

Table 24. (1) LPM regression of Russell Group Degree Attainment on Social Class and Parental Education with 1970 cohort group, Unskilled and Below Lower Secondary as baseline (2) Male subgroup (3) Female subgroup.

Russelll Group Results 1990

	Dependent variable:		
	RussellGroup		
	LPM	Males	Females
	(1)	(2)	(3)
Class 2	0.011	0.042	-0.014
	(0.040)	(0.072)	(0.048)
Class 3	0.041	0.032	0.048
	(0.036)	(0.062)	(0.044)
Class 4	0.103***	0.096	0.104**
	(0.038)	(0.064)	(0.048)
Class 5	0.189***	0.144*	0.223***
	(0.052)	(0.084)	(0.066)
Parental Education 2	-0.005	-0.006	-0.004
	(0.036)	(0.063)	(0.045)
Parental Education 3	0.078**	0.128**	0.045
	(0.034)	(0.058)	(0.042)
Parental Education 4	0.161***	0.139**	0.180***
	(0.034)	(0.056)	(0.044)
Constant	0.112***	0.129**	0.102***
	(0.032)	(0.053)	(0.039)
Observations	1,408	600	808
R^2	0.055	0.036	0.078
Adjusted R^2	0.050	0.024	0.070
Residual Std. Error	0.434 (df = 1400)	0.450 (df = 592)	0.421 (df = 800)
F Statistic	11.614*** (df = 7; 1400)	3.133*** (df = 7; 592)	9.632*** (df = 7; 800)
Note:			*p**p***p<0.01

Appendix 6: Diagnostic Checkings

Heteroskedasticity

In all our Linear Probability models, heteroskedasticity robust standard errors (HC1) have been used to avoid problems in this regard. In our logistic models, heteroskedasticity is not a solvable problem.

Multicollinearity

We have tested for the presence of multicollinearity in our IPW LPM model in **Table 5**. As can be seen in **Table 25** below, the scores indicate that this is probably not a serious issue.

Table 25. VIF analysis for regression in Table 5.

Variables	**GVIF Score**
Study	1.34
Partly Skilled	2.03
Skilled	3.27
Managerial/Technical	2.69
Professional	1.79
Lower Secondary	1.42
Upper Secondary	1.46
Degree	1.73

Appendix 7: Further Regression Results

Table 26. (1) Inverse probability weighted logistic regression of Undergraduate Degree Attainment on Parent Social Class and Parental Education with 1970 cohort group as baseline (2) Male subgroup (3) Female subgroup.

Logistic Results

	Dependent Variable: Undegraduate Degree Attainment		
	All	Males	Females
	(1)	(2)	(3)
1958	-0.257***	-0.180***	-0.347***
	(0.035)	(0.049)	(0.051)
1990	1.195***	1.075***	1.317***
	(0.033)	(0.046)	(0.047)
Partly Skilled	-0.148***	0.031	-0.310***
	(0.056)	(0.079)	(0.079)
Skilled	0.116**	0.249***	0.0002
	(0.047)	(0.068)	(0.066)
Managerial/Technical	0.460***	0.586***	0.353***
	(0.052)	(0.073)	(0.073)
Professional	0.882***	0.878***	0.929***
	(0.068)	(0.094)	(0.098)
Lower Secondary	0.207***	0.292***	0.113**
	(0.034)	(0.048)	(0.050)
Upper Secondary	0.677***	0.778***	0.571***
	(0.041)	(0.058)	(0.059)
Degree	1.244***	1.358***	1.129***
	(0.043)	(0.060)	(0.062)
Constant	-1.669***	-1.812***	-1.540***
	(0.050)	(0.072)	(0.070)
Observations	17,814	8,448	9,366
Log Likelihood	-16,458.650	-8,432.456	-8,000.684
Akaike Inf. Crit.	32,937.300	16,884.910	16,021.370
Note:			*p**p***p<0.01

Table 27. (1) Inverse probability weighted LPM regression of Undergraduate Degree Attainment on Social Class and Parental Education with 1970 cohort, Unskilled and Below Lower Secondary as baseline, as well as cohort interaction terms (2) Male subgroup (3) Female subgroup.

LPM With Interaction Term Results

	Dependent Variable: Undegraduate Degree Attainment		
	All	Males	Females
	(1)	(2)	(3)
1958	0.005	-0.006	0.017
	(0.022)	(0.030)	(0.031)
1990	0.422***	0.381***	0.460***
	(0.030)	(0.045)	(0.040)
Partly Skilled	0.063***	0.089***	0.035
	(0.020)	(0.029)	(0.028)
Skilled	0.073***	0.087***	0.059**
	(0.017)	(0.024)	(0.025)
Managerial/Technical	0.187***	0.191***	0.183***
	(0.023)	(0.032)	(0.032)
Professional	0.262***	0.226***	0.302***
	(0.030)	(0.042)	(0.043)
Lower Secondary	0.063***	0.075***	0.052***
	(0.011)	(0.016)	(0.015)
Upper Secondary	0.202***	0.220***	0.187***
	(0.018)	(0.026)	(0.024)
Degree	0.329***	0.350***	0.306***
	(0.019)	(0.027)	(0.027)
1958 *Partly Skilled	-0.067**	-0.049	-0.084**
	(0.026)	(0.039)	(0.036)
1958 *Skilled	-0.019	-0.010	-0.032
	(0.023)	(0.031)	(0.032)
1958 *Managerial/Technical	-0.032	-0.023	-0.043
	(0.030)	(0.042)	(0.043)

1958 *Professional	-0.101	0.306	-0.511
	(0.239)	(0.341)	(0.336)
1958 * Lower Secondary	-0.020	0.109	-0.184
	(0.089)	(0.122)	(0.130)
1958 *Upper Secondary	-0.106	-0.022	-0.204
	(0.131)	(0.182)	(0.189)
1958 *Degree	-0.271**	-0.427**	-0.136
	(0.133)	(0.188)	(0.191)
1990 *Partly Skilled	-0.796***	-0.952***	-0.595**
	(0.185)	(0.266)	(0.260)
1990 *Skilled	-0.529***	-0.561**	-0.477**
	(0.171)	(0.248)	(0.237)
1990 *Managerial/Technical	-1.062***	-0.993***	-1.110***
	(0.180)	(0.259)	(0.249)
1990 *Professional	-0.983***	-0.955***	-0.959***
	(0.207)	(0.292)	(0.296)
1990 * Lower Secondary	-0.693***	-0.805***	-0.597***
	(0.088)	(0.124)	(0.126)
1990 *Upper Secondary	-0.944***	-0.934***	-0.971***
	(0.098)	(0.139)	(0.138)
1990 *Degree	-0.795***	-0.740***	-0.856***
	(0.102)	(0.142)	(0.147)
Constant	-2.565***	-2.689***	-2.446***
	(0.158)	(0.229)	(0.217)
Observations	17,814	8,448	9,366
Log Likelihood	-16,288.260	-8,337.612	-7,902.697
Akaike Inf. Crit.	32,624.510	16,723.220	15,853.400
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 28. (1) Inverse probability weighted logistic regression of Undergraduate Degree Attainment on Social Class and Parental Education with 1970 cohort group, Unskilled and Below Lower Secondary as baseline, as well as cohort interaction terms (2) Male subgroup (3) Female subgroup.

Logistic With Interaction Term Results

	Dependent Variable: Undegraduate Degree Attainment		
	All	Males	Females
	(1)	(2)	(3)
1958	0.041	-0.056	0.156
	(0.199)	(0.290)	(0.273)
1990	2.493***	2.403***	2.573***
	(0.166)	(0.241)	(0.230)
Partly Skilled	0.625***	0.863***	0.372
	(0.172)	(0.247)	(0.241)
Skilled	0.708***	0.854***	0.570***
	(0.159)	(0.231)	(0.220)
Managerial/Technical	1.275***	1.368***	1.189***
	(0.168)	(0.242)	(0.232)
Professional	1.549***	1.476***	1.650***
	(0.185)	(0.264)	(0.260)
Lower Secondary	0.464***	0.527***	0.400***
	(0.064)	(0.089)	(0.092)
Upper Secondary	1.153***	1.220***	1.093***
	(0.080)	(0.114)	(0.113)
Degree	1.656***	1.737***	1.568***
	(0.082)	(0.114)	(0.119)
1958 *Partly Skilled	-0.664***	-0.413	-1.042***
	(0.236)	(0.330)	(0.347)
1958 *Skilled	-0.186	-0.101	-0.295
	(0.201)	(0.292)	(0.278)
1958 *Managerial/Technical	-0.165	-0.114	-0.224
	(0.213)	(0.308)	(0.295)

1958 *Professional	-0.101	0.306	-0.511
	(0.239)	(0.341)	(0.336)
1958 * Lower Secondary	-0.020	0.109	-0.184
	(0.089)	(0.122)	(0.130)
1958 *Upper Secondary	-0.106	-0.022	-0.204
	(0.131)	(0.182)	(0.189)
1958 *Degree	-0.271**	-0.427**	-0.136
	(0.133)	(0.188)	(0.191)
1990 *Partly Skilled	-0.796***	-0.952***	-0.595**
	(0.185)	(0.266)	(0.260)
1990 *Skilled	-0.529***	-0.561**	-0.477**
	(0.171)	(0.248)	(0.237)
1990 *Managerial/Technical	-1.062***	-0.993***	-1.110***
	(0.180)	(0.259)	(0.249)
1990 *Professional	-0.983***	-0.955***	-0.959***
	(0.207)	(0.292)	(0.296)
1990 * Lower Secondary	-0.693***	-0.805***	-0.597***
	(0.088)	(0.124)	(0.126)
1990 *Upper Secondary	-0.944***	-0.934***	-0.971***
	(0.098)	(0.139)	(0.138)
1990 *Degree	-0.795***	-0.740***	-0.856***
	(0.102)	(0.142)	(0.147)
Constant	-2.565***	-2.689***	-2.446***
	(0.158)	(0.229)	(0.217)
Observations	17,814	8,448	9,366
Log Likelihood	-16,288.260	-8,337.612	-7,902.697
Akaike Inf. Crit.	32,624.510	16,723.220	15,853.400
Note:			*p<0.1; **p<0.05; ***p<0.01

The cohort baseline is the 1970 cohort which means that the coefficients of different discrete levels of *Parental Education* and *Social Class* should be interpreted as their effects on the 1970

cohort. Furthermore, the *Social Class* and *Parental Education* baselines are *Unskilled* and *Below Lower Secondary*, respectively. Therefore, the interaction terms between cohort variables and discrete level variables should be interpreted as a calibrator of the discrete levels' effects which are measured in 1970 cohort's terms. For example, the coefficient of *1958 * Lower Secondary* calibrates the effect of *Lower Secondary* for the 1958 cohort.

Appendix 8: Construction of *Russell Group Degree Attainment*

The indicator for Russell Group attendance only exists for a share of our degree holders in the 1990 cohort. This means that all other individuals are not divided into not answering and NA. We can therefore not know whether the 600 degree holders in our sample who lack a Russell Group indicator went to such a school or not. Therefore, regression where these individuals are removed but our non-degree holders are kept would be biased. Our sample size is therefore particularly small for the Russell Group recessions, but we have used the inverse probability weighting method to try to alleviate some of the issues associated with this.