

‘Persistence of the social’: The role of cognitive ability in mediating the effects of social origins on educational attainment in Britain

Mollie Bourne

Department of Social Policy and Intervention and Nuffield College,

University of Oxford

mollie.bourne@spi.ox.ac.uk

Erzsébet Bukodi^(a)

Department of Social Policy and Intervention and Nuffield College,

University of Oxford

erzsebet.bukodi@spi.ox.ac.uk

Bastian Betthäuser

Department of Social Policy and Intervention and Nuffield College,

University of Oxford

bastian.betthaeuser@spi.ox.ac.uk

John H. Goldthorpe

Nuffield College,

University of Oxford

john.goldthorpe@nuffield.ox.ac.uk

(a) Corresponding author

Accepted for publication in *‘Research in Social Stratification and Mobility’*

‘Persistence of the social’: The role of cognitive ability in mediating the effects of social origins on educational attainment in Britain

Abstract

In this paper, we address two research questions on the basis of the series of British birth cohort studies. First, how large is the part played by cognitive ability in mediating the association that exists between individuals’ educational attainment and their social origins, and is there evidence of any change in the importance of its mediating role over recent decades? Second, does the importance of its mediating role change over the course of individuals’ educational careers? As regards the first question, we find that only around half of the effects of individuals’ social origins on their educational attainment is mediated via their cognitive ability, as measured in early life. There has been some fluctuation in the mediation percentage over time, but no sustained increase. Moreover, this is the case in whatever way we measure social origins. As regards the second question, we find that the mediating role of cognitive ability changes little in importance as individuals’ educational careers progress, with the possible exception that it declines in the case of an educational threshold relating to upper secondary qualifications. In the light of our results, we call into question the idea that the intergenerational reproduction of educational inequalities is driven overwhelmingly via the intergeneration transmission of cognitive ability; and also claims such as those made by Marks (2014) of ‘the decline of the social’ in the determination of the educational attainment of children from more or less advantaged families. While the relative importance of different forms of parental resources for children’s educational success may be changing somewhat, our findings indicate a strong ‘persistence of the social’.

1. Introduction

The publication of Herrnstein and Murray's *The Bell Curve* (1994) served to reignite longstanding debates concerning the role of intelligence, or cognitive ability, in the intergenerational reproduction of social inequalities in educational and socioeconomic attainment. These authors saw cognitive ability as playing a dominant role in such reproduction on account of it being a trait with a high degree of heritability. *The Bell Curve* aroused considerable controversy, turning on empirical, theoretical and normative issues (for critiques see Jacoby & Glauber, 1995; Fischer *et al.*, 1996; Devlin *et al.*, 1997), and these issues remain ones on which divergent and indeed often sharply conflicting views are still apparent.

The present paper aims to contribute to the ongoing arguments, but in a specific way that we need to make clear. It is important to differentiate between the following two quite distinct questions: first, that concerning the extent to which cognitive ability is in fact associated with educational and socioeconomic attainment; and, second, that of the extent to which cognitive ability is heritable. Our focus is on the first of these questions and, in particular, on the role of cognitive ability in the educational attainment of children of differing socioeconomic origins. For our present purposes we therefore take variation in cognitive ability as a 'given', without entering into the second question of the relative importance of heredity and environment in determining this variation. If our position on this second question should be thought of any interest, it is the following. We would concur with the conclusions reached by Heckman and associates (Heckman, 2007, 2013; Cunha and Heckman, 2007; Heckman and Mosso, 2014) from their extensive reviews of the relevant literature that in the case of a phenotypical trait such as cognitive ability attempts to partition its variance additively into one part that is determined by genetic effects and another by environmental effects are by now scientifically outmoded: not only are complex gene–gene interactions involved but, further, yet more complex processes of gene–environment interaction from the womb onwards. However, we would stress that the

findings that we report and the conclusions that we draw from them stand quite free of this position.

To date, previous researchers – ourselves included (Authors, 2014; Authors, 2018) – have typically aimed to determine the relative importance of cognitive ability and social origins on educational attainment through establishing their net effects in this regard, or by observing the disparities that social origins create in individuals' educational attainment when early life cognitive ability is held constant. But these approaches do not adequately accommodate the fact that cognitive ability is itself quite strongly associated with social origins (for evidence from Britain, see Sullivan *et al.*, 2013; Jerrim *et al.*, 2013).

The approach we now take is to conceptualise the joint relationships that are here involved in the form of an 'origins – cognitive ability – educational attainment' (OCE) triangle. This kind of conceptualisation has proved useful in the field of social mobility research in examining the extent to which the effects of social origins on social destinations are mediated via educational attainment. But it has been little employed in examining the mediating potential of cognitive ability in regard to educational attainment and, insofar as it has been used, there is a good deal of variability in the details of the analyses undertaken and in findings produced. For example, evidence from Sweden suggests that only around one third of the effects of social origins on educational attainment, as measured in adulthood, is transmitted via cognitive ability (Erikson, 2016), while in the US, cognitive ability has been shown to mediate around 50% of the effect of social origins as indexed by mothers' education and more than 80% of the effect of permanent family income on children's eighth grade attainment (Hsin & Xie, 2017).¹

¹ These are the mediation percentages calculated by the authors when cognitive ability is measured in 5th grade (i.e. at age 10/11 of the children). Of the various ages they consider, this is most similar to ours.

In pursuing this approach, a particular substantive concern that we have is with claims made by Marks (2014) of ‘the decline of the social’. In the context of a re-evaluation of modernisation theory, and in particular of arguments such as those advanced by Blau and Duncan (1967) of the progressive prevalence of ‘achievement’ over ‘ascription’, Marks maintains that existing research would support the two following propositions: first, that in modern societies the effects of socioeconomic inequalities of origin on individuals’ educational attainment are tending to decrease over time (2014: 234) - i.e. in terms of the OCE triangle the direct OE association is weakening; and second, that there is a further tendency for the importance of individuals’ cognitive ability for their educational attainment to increase - i.e. the CE association is strengthening. This latter tendency, Marks recognises, may be obscured at the higher levels of modern educational systems since their expansion means that more individuals of relatively low ability are able to reach them; but it is still reflected in actual performance at these levels and in access to ‘elite’ institutions and courses (2014: 236, 240).

We would note that, in the British case at least, our own recent work previously referred to, and that of others, has produced results that throw some doubt on Marks’ claims. The role of cognitive ability in educational attainment has been found actually to decline across cohorts born between the 1940s and the 1970s (Galindo Rueda & Vignoles, 2005; Richards *et al.*, 2009); and, while for a cohort born in the early 1990s this trend would appear to have reversed, no corresponding decline shows up in the net effects of social origins (Authors, 2018). But, to repeat, the studies in question have not simultaneously taken into account the association that exists between social origins and early life cognitive ability, nor how far this association may itself have changed over time.

One further issue of relevance, but about which little is known, is that of how far the role of cognitive ability – and in particular, the extent of its mediating influence – changes across individuals’ educational careers. It is conceivable that early life cognitive ability mediates *less* of

the effects of social origins at the later stages of educational careers, while social origins, and specifically parental resources, count for more, and especially as regards the educational choices that young people make, independently of their previous performance. But, conversely, it could be the case that cognitive ability comes to play a *more* important role at later educational stages insofar as these are likely to be more cognitively demanding.

Against this background, we therefore pose the two following research questions, to be taken up in the British case. First, to what extent are the effects of social origins on educational attainment mediated by early-life cognitive ability and how, if at all, has this changed over recent decades? Second, does the mediating role of individuals' early-life cognitive ability change over the course of their educational careers and, if so, does its importance decrease or increase?

One possible reason why previous research has often led to inconsistent or otherwise inconclusive findings is that social origins have been inadequately conceptualised and measured. As shown in our earlier work (Authors, 2013; Authors, 2014), different aspects of social origins – or, more specifically, of parental resources – give rise to different forms of (dis)advantage, and must be regarded as having independent, and perhaps independently changing, effects on children's educational attainment, and effects which will, moreover, be in some degree cumulative. Thus, considering only one type of parental resource is likely to lead to an underestimation of the total effects of social origins, as well as to an overestimation of the effects of other individual characteristics, such as cognitive ability. In the present paper, we distinguish between three different components of social origins: parental class, parental status and parental education.

The structure of the paper is as follows. In Section 2, we describe the data on which we draw and our key variables. In Section 3, we explain the analytical methods we use and present the results for each of our research questions in turn. In Section 4 we summarise our conclusions and comment on some further questions that directly arise.

2. Data and key variables

We use the data of three separate birth cohort studies, from which we aim to construct highly comparable variables in order to examine over-time trends. The National Child Development Study (NCDS) and the British Cohort Study (BCS70) are longitudinal surveys, which each sampled over 17,000 children born in Britain in one week in 1958 and 1970 respectively (see further Elliott & Shepherd, 2006; Power & Elliott, 2006).² The Avon Longitudinal Study of Parents and Children (ALSPAC) follows throughout their lives a sample of children born in the former region of Avon to over 14,000 pregnant women whose expected delivery date fell between April 1991 and December 1992 (see Boyd *et al.*, 2013).³

We use ALSPAC since it is the only dataset available for a 1990s born cohort that contains reliable and well validated measures of cognitive ability, which are essential for our purposes. But in doing so, we met two limitations that we here acknowledge and address.

First, the sample is not drawn from the national population. It has been shown that the demographic composition of respondents is broadly similar to that of the country as a whole (Boyd *et al.* 2012) but we have undertaken further checks by comparing the distribution of respondents on our key variables, and the associations between these variables and those obtained using identically constructed variables from the Longitudinal Study of Young People in England (LSYPE).⁴ *Appendix B* of the online supplementary material shows that we find only minor inconsistencies in the distributions for ALSPAC and LSYPE. Examination of the associations between variables indicates that the only difference likely to have implications for

² In both surveys, the target sample was later augmented to include immigrants born in the same week.

³ The ALSPAC website contains details of all data available through a fully searchable data dictionary (<http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>). Our use of the data received approval from the ALSPAC Ethics and Law Committee and the Local Research Ethics committees.

⁴ The LSYPE is an appropriate comparator since it is a nationally representative sample of roughly the same birth cohort – i.e. children born in 1989/90.

our findings is that in ALSPAC a somewhat stronger influence of parental status on children's educational attainment is apparent. We bear this in mind in interpreting our results.

Second, ALSPAC suffers from a large amount of missing data (see *Appendix A*, Table A1), and especially on the measure of cognitive ability (57%). This would appear to result from the high frequency of successive interviews (Fraser *et al.*, 2012; Crawford *et al.*, 2013) and, as regards cognitive ability, from the fact that tests were administered in clinics which cohort members had to attend.

To address the problem of missing data in ALSPAC – and likewise in NCDS and BCS70 where it is also present, although less severe – we apply multiple imputation (MI) methods, as described in detail in *Appendix C*. Following MI, our analytical samples include all cohort members surveyed in at least one wave of each study (NCDS N=18,558; BCS70 N=18,737; ALSPAC N=15,445).

As regards the approach we have taken to harmonise our key variables across studies, we provide a detailed account in *Appendix A*. The variables are as follows.

Education. Educational level, as our dependent variable, is measured by a 7 category ordered scale of educational qualifications, developed in earlier work (Authors, 2013), which incorporates the different qualifications that were available during the periods in which members of each of our birth cohorts were in education. This scale is shown in the left hand column of Table 1.

Distributions on the scale across our cohorts are given in *Appendix A*, Tables A8 and A9.

For our first research question, we focus on educational attainment at *age 20*, since this is the latest age at which we have information for our youngest cohort. We thus use only categories 3–7 of the scale, and treat these as forming a continuous variable. Insofar as our inability to include tertiary level qualifications has any implications for our findings, this would, if anything, be likely to lead to an underestimation of the effects of social origins (see Authors, 2017).

For our second research question, concerned with change across educational careers, we take as our dependent variables a series of binary measures indicating whether cohort members exceeded, or failed to exceed, each of four successive educational thresholds by *age 38*. In treating this question, we work only with our two older cohorts, since only for them do we have information on educational attainment in later life. In this case, we use all seven categories of our educational scale. As shown in Table 1, threshold 1 distinguishes between cohort members who achieved at least lower secondary qualifications (five or more O levels or GCSE passes) and those whose attainment is below this level; threshold 2, between those who achieved at least two academic qualifications at the higher secondary level (A levels) and the rest; threshold 3, between those with sub-degree- or degree-level qualifications and the rest; and threshold 4, between those with degrees and the rest.⁵

Since we are concerned with educational thresholds and not educational transitions, we work with our full samples throughout, and so do not come up against problems of selectivity at higher educational levels. Table 2 provides relevant descriptive statistics.

[Tables 1 and 2]

Early life cognitive ability. In NCDS, cognitive ability was measured at respondents' age 11 via a general ability test. In BCS70, four subtests of the British Ability Scales were administered at respondents' age 10. In ALSPAC, cognitive ability was measured at age 8, via 11 subtests of the Wechsler Intelligence Scale for Children (Kamphaus, 1993; Kaufman, 1994). We comment on the potential implications of cognitive ability being measured at an earlier age in ALSPAC in our results sections. Following what has now become a standard approach in measuring cognitive

⁵ It should be added here that there are many different vocational qualifications in Britain which have different values in the labour market (Wolf *et al.*, 2006). These either fall within the framework of the National Vocational Qualifications (NVQ) system or can be represented as NVQ equivalents. As indicated in Table 1, these NVQs or their equivalents are also included in each level of our educational scale and thus, in turn, in each of the educational thresholds that we define.

ability (Deary, 2001; Colom *et al.*, 2002), we aim to capture the common variance across the different items in the tests administered: i.e. the latent construct of general intelligence, g (Spearman, 1904; Jensen, 1998). For this purpose we use principal components analysis, saving scores from the first component extracted.⁶ From these scores we then derive cohort specific quintiles of cognitive ability: i.e. we create a relative measure, thus controlling for ‘Flynn effects’ (Flynn, 1987), and also allowing us to uncover any potential nonlinear effects on educational attainment. See online *Appendix A* for further information.

Social origins. As in our previous work, we endeavour to measure social origins in a way which is as comprehensive as possible. We include measures of parental class, parental status and parental education, each being taken at around cohort members’ age 10.

We see each component of social origins as capturing distinctive resources that parents can draw on in support of their children’s education. Parental class is taken to represent family economic resources. Class provides a more comprehensive indicator in this regard than does current income level, since it can be shown to be associated with income security, short-term income stability and longer-term income prospect, and further with wealth (Goldthorpe & McKnight, 2006; Authors, forthcoming). Further, in Britain, over time increases in income inequality have occurred to a greater extent between, rather than within, social classes (Williams, 2012) and the intergenerational association between class position has been shown to be stronger than that between income level (Goldthorpe, 2013). We begin with the 7 category National Statistics Socio-Economic Classification (NS-SEC) in which class is based on occupation and employment status, and which we derive using the ‘reduced’ method (Office for National Statistics, 2005: Table 17). We then collapse the categories in various ways that we describe below. Where both

⁶ Previous studies based on the birth cohorts have taken this same approach to obtaining a comparable cross-cohort measure (Galindo-Rueda and Vignoles, 2005; Schoon, 2008, 2010; Authors, 2014). The validity of the g concept has been much discussed, but there is now emerging evidence of its having a neurological basis (see Burnett, 2016, esp. pp. 135-143).

parents are in employment, we use the dominance method of determining parental class (Erikson, 1984).

Parental status is taken to indicate the extent of parents' socio-cultural resources: i.e. of their social networks and contacts and forms of cultural taste and participation (Chan *et al.*, 2010). We use parents' occupations to assign them to the 31 categories of the Chan–Goldthorpe (2004) status scale, which is based on the occupational structure of relations of close friendship, and we again use the dominance method in cases where both parents are in employment.⁷

Given that parental education is included in our analyses together with parental class and parental status, we take it as indicating the extent of the specifically educational resources that parents have available to help their children: for example, their ability to create a favourable home learning environment, to give assistance with homework or preparation for exams, and to provide their children with informed guidance through the educational system in regard to choice of schools, subjects, courses and examinations to take (cf. Schütz *et al.* 2005). Parents are allocated a value on a seven category scale developed in previous work and shown in *Appendix A, Table A4*, which takes into account the qualifications of both parents. The categories range from the lowest, in which neither parent has any qualification, to the highest, in which both have degree level qualifications. In the context of over time analyses in which the distribution of education has changed substantially, we see it as important to treat parental education in *relative* terms. We therefore measure parents' education according to the proportion of all parents whose level of qualification is below their own within the cumulative distribution for their child's cohort.

⁷ While our class and status measures are both based on occupation, they capture different aspects of occupations, and the association between them is sufficiently low to avoid any problems of multicollinearity (NCDS $\rho=0.60$, $p<0.05$; BCS70 $\rho=0.63$, $p<0.05$; ALSPAC $\rho=0.70$, $p<0.05$).

Table 3 gives descriptive statistics – distributions, means and standard deviations – for the three components of social origins.

[Table 3]

In treating our first research question, we take the five-fold hierarchical collapse of the parental class variable – as indicated by the dotted lines in *Table 3* – along with within cohort approximations to quintiles for parental status and education, and then rescale each of these measures to take values within the range 0–1 so as to allow comparability in the mediation percentages across the three components of parental background.

In previous work, we have shown that although the three components of social origins are correlated, the correlations are not perfect, and cumulative effects may therefore be expected in some degree to arise (see Authors, 2013). For this reason, we also use a combined measure of social origins, taking all three indicators of parental resources into account. Specifically, we work with a three-fold collapse of parental class and an approximation to tertiles within each cohort for parental status and parental education. We then combine these as shown in *Appendix D, Table D1*, so as to create consistently advantaged, consistently disadvantaged and intermediate parental categories. Distributions of cohort members across these parental categories are shown in *Table 3*.

Our interest in this paper is not in gender differences and, while we include gender as a control in all of our models, we do not report separate findings for men and women. Further analyses (available on request) reveal that no significant differences by gender do in fact arise, except in one instance that we later note.

3. Methods and results

3.1 Research Question 1

We seek to determine the extent to which early life cognitive ability mediates the effects of social origins on individuals' educational attainment at age 20, and to examine whether there has been any change in this regard over recent decades. For this purpose, we simultaneously estimate two sequential multivariate linear regression models for each cohort. Cognitive ability is treated as an exogenous variable in one of the equations and as an endogenous variable in the other. The procedure is analogous to the method of classical path analysis. Taking the three separate components of parental class, parental status and parental education to comprise social origins, our conceptual model has the form depicted in Figure 1.⁸

[Figure 1]

This conceptual model can then be formally represented by the pair of linked regression equations:

$$\chi_4 = \beta_{41}\chi_1 + \beta_{42}\chi_2 + \beta_{43}\chi_3 + \varepsilon_Z,$$

$$\chi_5 = \beta_{51}\chi_1 + \beta_{52}\chi_2 + \beta_{53}\chi_3 + \beta_{54}\chi_4 + \varepsilon_Y,$$

where β represents the coefficient for each exogenous variable, with the first subscript referring to the dependent variable it is predicting and the second subscript referring to the exogenous variable in question. Error terms, represented by ε , are assumed to be stochastically independent. Since our interest is in over-time change, which involves comparisons across populations with differing distributions of explanatory variables, we work with unstandardized coefficients and

⁸ We have, in previous work, investigated the possibility of an interaction between social origins and cognitive ability, but found no significant effects.

partition the total effect of each social origin component into its direct part – i.e. the part of the total association that occurs net of the indirect path through cognitive ability – and its indirect part – i.e. the part of the total association that is mediated through cognitive ability. This latter is computed as the product of regression coefficients for the paths that link the relevant variables.

Taking parental class as an example, the total effect of parental class (χ_1) on educational attainment is given by:

$$\beta_{51} + (\beta_{54} * \beta_{41})$$

where β_{51} quantifies the direct effect and $\beta_{54} * \beta_{41}$ quantifies the indirect effect.

From this, the proportion of the total effect of parental class that is mediated through cognitive ability can be calculated by:

$$\beta_{54} * \beta_{41} / \beta_{51} + (\beta_{54} * \beta_{41}).$$

In *Figure 2*, we summarise the results of our decomposition analyses in percentage terms. We show how much of the total effect of each social origin component is mediated by cognitive ability and also the percentage of the total effect of our combined measure of social origins that is mediated by cognitive ability. In this latter case, we work with the three category parental variable previously described, and our results represent the contrast of being in the consistently advantaged as opposed to the consistently disadvantaged parental group. In *Table 4* we present the coefficients from each of the linked regression equations which we use for the calculation of the mediation percentages.^{9 10}

⁹ As a check, in *Appendix D Table D2* and *Figure D1* we present the results of analyses and mediation percentages that take the intermediate parental group as comparator, in the case of our combined origins variable. The pattern of these results is essentially the same as that shown in *Figure 2*, except that there is no change in the mediation percentage between the 1970 and 1990s cohorts for the consistently disadvantaged.

¹⁰ In *Appendix D Table D3* we also report the standardised regression coefficients for the linked regression equations, when taking each component of social origins separately. And, in *Table D4* we present the

[Figure 2 & Table 4]

As is apparent from Figure 2, between 30% and 50% of social origin effects are mediated by early life cognitive ability, irrespectively of whether we consider the components of social origins separately or in combination.¹¹ In other words, we can say that at least half of the effects of social origins is direct – that is, occurs *other than through* early life cognitive ability.

Turning to changes over time, we find no evidence of any sustained increase in the mediating role of cognitive ability across our cohorts, nor of any sustained decrease. However, what we do see is some evidence of a ‘dip’ in the role of cognitive ability with the 1970 cohort. This is most apparent when the components of social origins are taken in combination via our parental groups, and it may be added that the dip is somewhat more marked in the case of women than of men.

Galindo-Rueda and Vignoles (2005) have suggested that a decline in the importance of cognitive ability for educational attainment, which they also show for the 1970 cohort, could result from the move from selective to comprehensive secondary education during the 1960s, benefiting children who might otherwise have lost out due to early selection based on cognitive ability. Such an effect could then also underlie our own finding. And, conversely, the subsequent increase in the mediating role of cognitive ability could reflect the growing governmental pressure on schools, from around the mid-1990s, to replace mixed ability teaching with setting and streaming.

mediation percentages that are based on these standardised coefficients. As is apparent from the table, the proportion of the total effect of each social origin component that is mediated by cognitive ability is the same as in Figure 2.

¹¹ This is more than the approximately 30% found by Erikson (2016) for Sweden, but around the same as was found by Hsin and Xie (2017) for the US, although when indexing social origins only by mother’s education. But neither of these studies is directly comparable to ours as regards the dependent variable. Erikson measures educational attainment at around age 32–40, thus taking account of tertiary-level qualifications, and Hsin and Xie measure educational attainment at around age 14–15. Our findings are, however, similar to Erikson’s in showing no great difference in mediation percentages for different components of social origins.

To gain greater insight into the over-time patterns in the mediation percentages revealed in Figure 2, we go on to compare the unstandardized regression coefficients in the OCE triangle across cohorts. We summarise the results in a stylised form in *Figure 3* and full details are again given in Table 4.

[Figure 3]

Figure 3 and Table 4 bring out three main findings. First, we see that the association between cognitive ability and educational attainment – the CE association – first weakens between the 1958 and 1970 cohorts and then strengthens between the 1970 and 1990s cohorts.¹² This is consistent with findings we have presented in other work (Authors, 2018), and with the dip in the mediation percentages suggested in Figure 2.

Second, the direction of change in the strength of the association between social origins and cognitive ability – the OC association – depends on which component of social origins we consider. There is a sustained weakening in the association in the case of parental class: i.e. a declining effect of parental economic resources on children’s cognitive ability – a finding in line with recent work by Connelly and Gayle (2017). But we see some strengthening in the association in the case of parental status and one of a more sustained kind in the case of parental education: i.e. increasing effects of parents’ socio-cultural and educational resources on children’s cognitive ability. In turn, the effect on children’s cognitive ability of social origins taken in combination

¹² We are aware of the possibility that the strengthening CE association found for the 1990s cohort could be an artefact of better measurement of cognitive ability in ALSPAC than in the earlier studies. We cannot, unequivocally, assert that this is not the case. However, we have carried out further checks that imply that our results are robust and reliable (Authors, 2018). Also, we know that cognitive ability in ALSPAC is measured at an earlier age than in the two other studies (age 8, compared with ages 11 and 10), and one might argue that, since it is observed further in time from the dependent variable, it would be less predictive of educational attainment. We would therefore consider our analyses to be providing, if anything, a conservative estimate of the increasing effect of cognitive ability for the youngest cohort.

increases between the 1958 and 1970 cohorts and then remains unchanged between the 1970 and 1990s cohorts.

Third, so far as the direct association between social origins and educational attainment – the OE association – is concerned, this weakens only in the case of parental class.¹³ With the other two components of social origins, we see stability in the case of parental education and some strengthening in the case of parental status. When the components of social origins are considered in combination, there is some strengthening between the 1958 and 1970 cohorts and no weakening subsequently. This then implies that the dip in the mediation percentage that we report in Figure 2 is due *solely* to the weakening CE association between the 1958 and 1970 cohorts. In other words, there is, in the British case at least, little indication of ‘the decline of the social’: the effects of social origins not mediated via cognitive ability would appear to be quite persistent.

3.2 Research Question 2

We now turn to our second research question of whether cognitive ability plays a stronger or weaker mediating role as individuals’ educational careers progress. Because of space limitations, we present here only the results of analyses using our combined measure of social origins as represented by the three parental groups earlier described. We have, however, repeated the analyses taking our three components of social origins separately, and the results obtained (see *Appendix D Figure D2*) are in general in line with those reported in the text below.

¹³ This result would merit further investigation. Due to educational expansion and generally rising living standards, members of our youngest cohort may have faced fewer economic barriers in education, at all events up to tertiary level (cf. Buis, 2013). But economic resources could still be thought to play some – and possibly increasing – part in social inequalities in educational attainment through parents’ resort to the private school sector, the employment of private tutors (Kirby, 2016), and the purchase of houses in the catchment areas of a good state schools.

We focus on the two older cohorts – those born in 1958 and 1970 – for whose members we have information about their tertiary level education. Educational attainment is measured by highest qualification achieved by age 38 according to our four educational thresholds, as shown in Table 1, and with the distributions as shown in Table 2.

Since our dependent variables are binary, we apply nonlinear probability models. With such models, partitioning the total effects of social origins into direct and indirect components is not as straightforward as in the linear case. Coefficients for the explanatory variables of interest are affected not only by the inclusion of other variables in the model with which they are correlated, but also by the residual variation which changes in response to alterations in model specification (see Winship & Mare, 1984; Mood, 2010; Kohler *et al.*, 2011). We thus make use of an approach developed by Erikson *et al.* (2005) and generalised by Buis (2010) which seeks to overcome these problems through the counterfactual analysis of log odds ratios.

In this case, calculation of the indirect effect involves comparing the observed proportion of success at a particular threshold for individuals from one parental group with the proportion of success that these individuals would have achieved if they had the cognitive ability distribution of individuals from a second parental group. This comparison thus allows one to see how far the effect of social origins on success is due to differences in the distribution of cognitive ability.

Calculation of the direct effect then involves comparing the observed proportions of success for individuals in the second group with the proportion of success that individuals in the first group would have achieved if they had the cognitive ability distribution of individuals in the second group – that is, the distribution of cognitive ability, the mediator, is now held constant.

Alternatively, one could follow the same approach but use the counterfactual proportions of success of individuals in the second group assuming they had the distribution of cognitive ability of individuals in the first group. These alternatives result in estimates that are usually very similar,

although not identical. We take the average of the indirect effects relative to the total effect across the two alternative specifications (see further Erikson *et al.*, 2005; Jackson *et al.*, 2007).

Figure 4 shows the proportion of the total effect of combined social origins at our four educational thresholds that is mediated by cognitive ability. Our results are, again, based on the contrast between being in the consistently advantaged as opposed to the consistently disadvantaged parental group. On the left hand side of the figure we show the gross mediation percentages. On the right hand side, we show the mediation percentages from models which include controls for prior performance at threshold 1 – i.e. number of GCSEs or O-levels obtained – in the case of threshold 2; and controls for prior performance at threshold 2 – i.e. number of A-levels obtained – in the case of thresholds 3 and 4. Beyond threshold 1, individuals have to decide whether to continue to higher levels of education prior to taking the relevant examinations, so that as regards exceeding thresholds 2, 3 and 4, elements of both choice and performance are involved.¹⁴

[Figure 4]

From Figure 4, it can be seen that, in line with our results in regard to our first research question, there is some suggestion, and especially when we control for prior educational performance, that cognitive ability mediates less of social origin effects in the 1970 cohort than in the 1958 cohort. The point estimates are consistently lower for the 1970 cohort, although the confidence intervals around the point estimates for the two cohorts regularly overlap.

However, specifically as regards our second research question, the important outcome is that no sustained increase or decrease shows up in the mediation percentages of cognitive ability across

¹⁴ These choices are of course made under some constraints: i.e. individuals must demonstrate a degree of academic proficiency before they have a choice to make. For example, to have the opportunity to choose whether or not to continue to A-levels, students must typically achieve a minimum of five A*–C grade GCSEs.

the successive educational thresholds that we distinguish. Roughly 40% of the combined social origin effect is mediated through cognitive ability at the first and at the third and fourth thresholds. There is, however, an indication of a drop in the mediation percentage for threshold 2, the upper secondary threshold, when we control for prior academic performance.¹⁵ This could be because whether or not to proceed to the upper secondary level is, in the British case, a rather critical decision in individuals' school careers (Jackson, 2013), and it is perhaps through differences in choice in this regard that social origins show increased direct effects. Following the distinction made by Boudon (1974) between 'primary effects' in educational inequalities – differences in educational performance – and 'secondary effects' – differences in educational choices given performance – attention has focussed (see Breen & Goldthorpe, 1997 and subsequent literature) on how, when faced with an educational branching point, individuals consider not only how well they have done previously but also the resources they, or their families, have at their disposal, and in turn the risks and the relative costs and benefits of taking a more or a less ambitious option.

4. Conclusions

In this paper we have addressed two research questions on the basis of the rich data of the British birth cohort studies. First, how important is cognitive ability, as measured in early life, in mediating the association that exists between individuals' social origins and their educational

¹⁵ Differences across the thresholds are not significant at the 0.05 level, but the drop at threshold 2 when controlling for academic performance is significant at $p < 0.10$.

As a general robustness check, we have also used another method of decomposition, the so-called KHB method (Karlson *et al.*, 2012; Kohler *et al.*, 2011). We find that for the most part, this method gives very similar results to those we report. The only notable difference is that with the KHB method the drop in the mediation percentage at threshold 2 is somewhat more pronounced (*Appendix D Figure D3*).

attainment, and is there evidence of any change in its importance over recent decades? Second, how far does this mediating effect change over the course of individuals' educational careers?

As regards the first question, our main finding is that while cognitive ability does clearly play a significant mediating role, the extent of this should not be exaggerated: at most, only something approaching a half of the effects of individuals' social origins, when measured in an adequately comprehensive manner, appears to be mediated through such ability. There has been some fluctuation in the mediation percentage, with a decrease between the 1958 and 1970 cohorts being reversed with the 1990s cohort. But there has been no sustained increase. And, conversely, there has been no 'decline of the social' in the way that Marks (2014) would claim: i.e. no decline in the direct association between individuals' social origins and their educational attainment. This association weakens only if the focus is on parental class; in the case of parental education it is stable and in that of parental status it strengthens. With our combined measure of social origins there is some strengthening of the direct association between the 1958 and 1970 cohorts and no change between the 1970 and 1990s cohorts. It is of some interest that our results on the absence of any 'decline of the social' are in line with those of a US study of very different design (Conley and Dominigue, 2016), in which it has been possible to compare the effect of genotype on educational attainment across mid-twentieth century birth cohorts.

As regards the second question, we show, using our combined measure of social origins, that the mediating role of cognitive ability changes little in importance as individuals' educational careers progress, with the possible exception that it declines somewhat – and the direct effect of social origins strengthens – in the case of the educational threshold relating to upper secondary qualifications.

As we stressed at the outset, in the analyses that we have carried out we take cognitive ability simply as a 'given' and do not enter into the question of how far variation in such ability is genetically or environmentally determined. We would, however, observe that the results that we

have reported are unaffected by whatever position might be taken on this question, and two implications then follow. First, even if cognitive ability is regarded as having a high degree of heritability, our results would indicate that the intergenerational reproduction of inequalities in educational attainment still cannot be seen as resulting from the intergenerational transmission of innate ability to the dominant extent that authors such as Herrnstein and Murray would suggest. Second, insofar as environmental factors associated with individuals' social origins do play a part in the formation of their cognitive ability – as, say, in interaction with genetic factors – then our estimates of the effects of social origins on individuals' educational attainment will be conservative ones. In this regard, some interest may be attached to our further finding that while the association of children's cognitive ability with their parents' class appears to be weakening over time, that with their parents' status and education is strengthening.

Finally, we would note that if cognitive ability mediates no more than half of the association between social origins and educational attainment, the question obviously arises of how the rest of the association is mediated or, in other words, of how the direct effects of social origin actually operate other than through cognitive ability. Two different lines of inquiry of this question might be pursued in future research.

On the one hand, other *individual attributes*, apart from cognitive ability, may be identified that are associated with both educational attainment and social origins. There is in fact now a growing body of research into various non-cognitive attributes – for example, self-control, self-efficacy and 'social skills' – which suggests that these may be of comparable importance for educational outcomes to cognitive ability (for an overview, see Gutman & Schoon, 2013) and also showing a degree of association with parental characteristics (e.g. Osborne-Groves, 2005; Carneiro *et al.*, 2006; Blanden *et al.*, 2007; Carneiro *et al.*, 2007).

On the other hand, *social processes* may be identified through which different parental resources are actually brought to bear on children's chances of educational success. For example, our more

detailed analyses of the direct effects of different components of social origins indicate that while those of parental class, status and education are all significant, the importance of the two latter components are, to repeat, increasing. And there is indeed evidence that, from the 1990s, with the greater official emphasis on school choice and school competition in regard to standards (Gibbons *et al.*, 2006), parents may be drawing on their sociocultural and specifically educational resources, yet more than on their economic resources, in order to defend and further their children's interests. Hansen and Vignoles (2005) suggest that parents are now acting as 'consumers' with respect to schools and that those with higher levels of education are better placed to understand school performance and quality. And Wespieser *et al.* (2015) find that parents of higher social status are advantaged in being able to draw on wider social networks and more useful connections to learn about school characteristics, available options and how to realise them. More generally, Buis (2013) has argued that parents' superior sociocultural and educational resources play a major part in offsetting policy and institutional changes aimed at reducing social inequalities in educational attainment and in thus maintaining the *status quo*. It is then possible that further research on the lines in question will show that, rather than there being any decline of the social, in the linkage between individuals' social origins and their educational attainment, socially contextualised flows of information and influence are of growing importance.

References

- Blanden, J., Gregg, P. & Macmillan, L. (2007) Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education, *The Economic Journal*, 117, C43–C60.
- Blau, P.M. and Duncan, O.D. (1967) *The American Occupational Structure*. New York: John Wiley & Sons.
- Boudon, R. (1974). *Education, Opportunity and Social Inequality*. New York: Wiley.
- Boyd, A., Goulding, J., Macleod, J. et al. (2013). Cohort Profile: The ‘Children of the 90s’ – the Index Offspring of the Avon Longitudinal Study of Parents and Children, *International Journal of Epidemiology*, 42, 111–1277.
- Breen, R. & Goldthorpe, J. (1997) Explaining Educational Differentials: Towards a Formal Rational Action Theory, *Rationality and Society*, 9, 275–305.
- Buis, M. (2013) The Composition of Family Background: The Influence of the Economic and Cultural Resources of both Parents on the Offspring’s Educational Attainment in the Netherlands between 1939 and 1991. *European Sociological Review*, 29, 593–602.
- Buis, M. (2010) Direct and Indirect Effects in a Logit Model. *The Stata Journal*, 10, 11–29.
- Burnett, D. (2016) *The Idiot Brain. A Neuroscientist Explains What Your Head Is Really Up To*. London: Guardian Books.
- Carneiro, P, Crawford, C. & Goodman, A. (2007) *The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes*. Working Paper CEE DP 92. London: Centre for the Economics of Education.
- Carneiro, P., Crawford, C. & Goodman, A. (2006) *Which Skills Matter?* Working Paper CEE DP 59. London: Centre for the Economics of Education.
- Chan, T.–W. (Ed.) (2010). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.
- Chan, T.–W. & Goldthorpe, J. (2004). Is There a Status Order in Contemporary British Society? Evidence from the Occupational Structure of Friendship, *European Sociological Review*, 20, 383–401.
- Colom, R., Abad, F. J., Garcia, L. F. & Juan-Espinosa, M. (2002). Education, Wechsler’s Full Scale and g, *Intelligence*, 30, 449–462.
- Conley, D. & Domingue, B. (2016) The Bell Curve Revisited: Testing Controversial Hypotheses with Molecular Genetic Data, *Sociological Science*, 3, 520–539.
- Connelly, R. & Gayle, V. (2017) An Investigation of Social Class Inequalities in General Cognitive Ability in Two British Birth Cohorts, *British Journal of Sociology*, DOI: 10.1111/1468-4446.12343.
- Crawford, C., Goodman, A. & Greaves, E. (2013) *Cohabitation, Marriage, Relationship Stability and Child Outcomes: Final Report*. London: Institute for Fiscal Studies.
- Cunha, F. & Heckman, J. (2007) The Technology of Skill Formation, *American Economic Review*, 97, 31–47.
- Deary, I. J. (2001). *Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.

- Devlin, B., Fienberg, S. E., Resnick, D. P. & Roeder, K. (1997) (eds) *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*. New York: Springer-Verlag New York Inc.
- Elliott, J. & Shepherd, P. (2006). Cohort Profile: 1970 British Birth Cohort (BCS70), *International Journal of Epidemiology*, 35, 836-843.
- Erikson, R. (2016) Is It Enough to be Bright? Parental Background, Cognitive Ability and Educational Attainment, *European Societies*, 18, 117–135.
- Erikson, R. (1984). Social Class of Men, Women and Families, *Sociology*, 18, 500-514.
- Erikson, R., Goldthorpe, J. H., Jackson, M., Yaish, M. & Cox, D. R. (2005). On Class Differentials in Educational Attainment. *Proceedings of the National Academy of Science*, 102, 9730–9733.
- Fischer, C. S., Hout, M. & Jankowski, M. S. (1996) *Cracking the Bell Curve Myth*. Princeton, NJ: Princeton University Press.
- Flynn, J. R. (1987). Massive IQ Gains in 14 Nations: What IQ Tests Really Measure, *Psychological Bulletin*, 101, 171-191.
- Galindo-Rueda, F. & Vignoles, A. (2005). The Declining Relative Importance of Ability in Predicting Educational Attainment, *Journal of Human Resources*, 40, 335-353.
- Goldthorpe, J. H. & McKnight, A. (2006). The Economic Basis of Social Class, in S. Morgan, D. B. Grusky and G. S. Fields (Eds.) *Mobility and Inequality: Frontiers of Research from Sociology and Economics*. Stanford University Press.
- Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M. & Lawlor, D. A. (2012). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers Cohort, *International Journal of Epidemiology* 42, 97-110.
- Gibbons, S., Machin, S. & Silva, O. (2006) *Competition, Choice and Pupil Achievement*. Centre for the Economics of Education Discussion Paper 56. London: London School of Economics.
- Gutman, L. M. & Schoon, I. (2013) *The Impact of Non-cognitive Skills on Outcomes for Young People: Literature Review*. London: Institute of Education.
- Hansen, K. & Vignoles, A. (2005) The United Kingdom Education System in a Comparative Context. In S. Machin and A. Vignoles (eds.) *What's the Good of Education? The Economics of Education in the UK*. Princeton: Princeton University Press.
- Heckman, J. J. (2013). *Giving Kids a Fair Chance*. Boston Review Books.
- Heckman, J. J. (2007). The Economics, Technology and Neuroscience of Human Capability Formation. *Proceedings of the National Academy of Sciences*, 104, 13250–13255.
- Heckman, J. J. & Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, 6, 689-733.
- Herrnstein, R. J. & Murray, C. (1994) *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.
- Hsin, A. & Xie, Y. (2017) Life-course changes in the mediation of cognitive and non-cognitive skills for parental effects on children's academic achievement. *Social Science Research*, 63, 150-165.

- Jackson, M. (2013) Social Background and Educational Transitions in England, in M. Jackson (eds) *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford, CA: Stanford University Press.
- Jackson, M., Erikson, J. H., Goldthorpe, J. H. & Yaish, M. (2007) Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-level Courses in England and Wales. *Acta Sociologica*, 50, 211–229.
- Jacoby, R. & Glauber, N. (1995) *The Bell Curve Debate: History, Documents, Opinions*. New York: Times Books.
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.
- Jerrim, J., Vignoles, A., Lingham, R. & Friend, A. (2013). *The Socio-economic Gradient in Children's Reading Skills and the Role of Genetics*. Department of Quantitative Social Science, Institute of Education, London
- Kamphaus, R. W. (1993). *Clinical Assessment of Children's Intelligence*. Boston, MA: Allyn & Bacon.
- Karlson, K. B., Holm, A. & Breen, R. (2012) Comparing Regression Coefficients between Same-sample Nested Models using Logit and Probit: A New Method. *Sociological Methodology*, 42, 286–313.
- Kaufman, A. S. (1994). *Intelligent Testing with the WISC-III*. New York: John Wiley & Sons.
- Kirby, P. (2016) *Shadow Schooling. Private Tuition and Social Mobility in the UK*. The Sutton Trust.
- Kohler, U., Karlson, K. B. & Holm, A. (2011) Computing Coefficients of Nested Nonlinear Probability Models. *The Stata Journal*, 11, 420–438.
- Marks, G. N. (2014). *Education, Social Background and Cognitive Ability: The Decline of the Social*. London: Routledge.
- Mood, C. (2010) Logistic Regression: Why We Cannot Do What We Think We Can Do and What We Can Do About It. *European Sociological Review*, 26, 67–82.
- Office for National Statistics (ONS) (2005). *The National Statistics Socio-economic Classification: User Manual*. London: Palgrave Macmillan.
- Osborne-Groves, M. (2005) Personality and the Intergenerational Transmission of Earnings. In S. Bowles, H. Gintis and M. Osborne-Groves (eds) *Unequal Chances: Family Background and Economic Success*. Princeton: Princeton University Press.
- Power, C. & Elliott, J. (2006). Cohort Profile: 1958 British Birth Cohort (National Child Development Study), *International Journal of Epidemiology*, 35, 34–41.
- Richards, M., Power, C. & Sacker, A. (2009). Paths to literacy and numeracy problems: evidence from two British birth cohorts. *Journal of Epidemiology and Community Health*, 63, 239–44.
- Schoon, I. (2008). A Transgenerational Model of Status Attainment: The Potential Mediating Role of School Motivation and Education, *National Institute Economic Review*, 205, 72–82.
- Schoon, I. (2010). Childhood Cognitive Ability and Adult Academic Attainment: Evidence from Three British Cohort Studies, *Longitudinal and Life Course Studies*, 1, 241–258.

- Schütz, G., H.W. Ursprung & L. Woessmann. 2005. *Education Policy and Equality of Opportunity*. Bonn.
- Spearman, C. (1904). “General Intelligence”, Objectively Determined and Measured, *American Journal of Psychology*, 15, 201-293.
- Sullivan, A., Ketende, S. & Joshi, H. (2013). Social Class Inequalities in Early Cognitive Scores, *Sociology*, 47, 1187-1206.
- Wespieser, K., Durbin, B. & Sims, D. (2015) *School Choice: The Parent View*. Slough: National Foundation for Educational Research.
- Williams, M. (2012). Occupations and British Wage Inequality, 1970s–2000s. *European Sociological Review*, 29, 841-857.
- Winship, C. & Mare, R. D. (1984) Regression Models with Ordinal Variables. *American Sociological Review*, 49, 512–525.
- Wolf, A., Jenkins, A. & Vignoles, A. (2006). Certifying the Workforce: Economic Imperative or Failed Social Policy?’, *Journal of Education Policy*, 21, 535–565.

Tables

Table 1. Educational qualifications scale and the derivation of four educational thresholds

Educational qualifications scale	Threshold 1 [lower secondary]	Threshold 2 [higher secondary]	Threshold 3 [lower tertiary]	Threshold 4 [higher tertiary]
1. Degree or higher degree, NVQ 5 or NVQ 6 [higher tertiary]				1 vs. 2–7
2. Tertiary level sub-degree qualification, NVQ 4 [lower tertiary]			1–2 vs. 3–7	
3. 2+ A-level passes [higher secondary]		1–3 vs. 4–7		
4. 5+ O-level passes or 1 A-level pass, NVQ 3 [lower secondary – high performance]	1–4 vs. 5–7			
5. 1–4 O-level passes, NVQ 2 [lower secondary – low performance]				
6. Below O-level, NVQ1 [sub-secondary]				
7. No qualifications				

Notes: All vocational qualifications are either National Vocational Qualifications (NVQ) or are given as their NVQ equivalents. Academic qualifications include the General Certificate of Secondary Education (GCSE) (formerly Ordinary-level CSE (O-level)), which are achieved via compulsory examinations taken by all pupils at the end of the lower secondary level (age 16), and Advanced-level (A-level) qualifications, which are taken by students who have continued in further academic study at the upper-secondary level (usually age 16–18). At the tertiary level, we differentiate sub-degrees (which are usually vocational) and academic degree-level qualifications.

Table 2. Educational attainment at age 20 and age 38

	1958 cohort	1970 cohort	1990s cohort
	Mean (SD)	Mean (SD)	Mean (SD)
Highest educational attainment at age 20 (continuous)	2.86 (1.50)	3.05 (1.45)	3.79 (1.22)
Highest educational attainment at age 38 (thresholds)	% exceeded threshold	% exceeded threshold	
Exceeded Threshold 1 (lower secondary)	46	57	
Exceeded Threshold 2 (higher secondary)	30	36	
Exceeded Threshold 3 (lower tertiary)	26	33	
Exceeded Threshold 4 (higher tertiary)	12	19	

Table 3. Descriptive statistics for variables of social origin

	1958 cohort	1970 cohort	1990s cohort
<i>Parental class (%)</i>			
Class 1: Higher managerial and professional	8	11	21
Class 2: Lower managerial and professional	12	20	26
Class 3: Intermediate	9	9	13
Class 4: Small employers and own account workers	13	13	5†
Class 5: Lower supervisory and technical	20	18	12
Class 6: Semi routine	16	14	14
Class 7: Routine	22	15	9
<i>Hierarchical component (Mean (SD))</i>			
Parental class collapsed and recoded 0–1	0.42 (0.29)	0.49 (0.28)	0.58 (0.30)
<i>Parental status (Mean SD))</i>			
Parental status quintiles recoded 0–1	0.47 (0.35)	0.47 (0.35)	0.48 (0.36)
<i>Parental education (Mean SD))</i>			
Parental education quintiles recoded 0–1	0.30 (0.32)	0.39 (0.32)	0.44 (0.30)
<i>Combined social origins (%)</i>			
Consistently advantaged	20	25	28
Intermediate	52	46	43
Consistently disadvantaged	28	29	29

Notes: † This estimate for the proportion of cohort members' parents that are in class 4 is smaller than expected. The estimate for the size of this class that we find using a nationally representative dataset (LSYPE), is 9% (see Online Appendix B). The potential implications of this are that the proportions in other classes are slightly overestimated.

Table 4. Effects of social origins on cognitive ability and effects of social origins and cognitive ability on educational attainment at age 20 - results from sequential linked regression analyses (unstandardized regression coefficients with standard errors)

	1958 cohort		1970 cohort		1990 cohort	
	DV: Cognitive ability	DV: Educational attainment	DV: Cognitive ability	DV: Educational attainment	DV: Cognitive ability	DV: Educational attainment
<i>Components of social origins</i>						
Parental class	0.82 ** [0.06]	0.47 ** [0.05]	0.52 ** [0.06]	0.37 ** [0.06]	0.29 ** [0.07]	0.18 ** [0.04]
Parental status	0.39 ** [0.04]	0.20 ** [0.05]	0.54 ** [0.05]	0.23 ** [0.05]	0.64 ** [0.05]	0.46 ** [0.04]
Parental education	0.74 ** [0.05]	0.51 ** [0.04]	0.97 ** [0.04]	0.55 ** [0.05]	1.44 ** [0.06]	0.63 ** [0.04]
Cognitive ability	-	0.47 ** [0.01]	-	0.32 ** [0.01]	-	0.41 ** [0.01]
R ²	12%	29%	15%	19%	21%	45%
<i>Combined measure of social origins</i>						
Consistently advantaged	1.24 ** [0.04]	0.69 ** [0.04]	1.46 ** [0.04]	0.78 ** [0.04]	1.50 ** [0.04]	0.79 ** [0.03]
Intermediate	0.50 ** [0.03]	0.25 ** [0.03]	0.65 ** [0.03]	0.31 ** [0.03]	0.68 ** [0.03]	0.42 ** [0.02]
Consistently disadvantaged (ref.)						
Cognitive ability	-	0.49 ** [0.01]	-	0.33 ** [0.01]	-	0.44 ** [0.01]
R ²	10%	28%	14%	19%	15%	44%
N	18558		18737		15242	

Note: Models control for gender and run separately by cohort.

** p<0.001

Figures

Figure 1. Conceptual path diagram

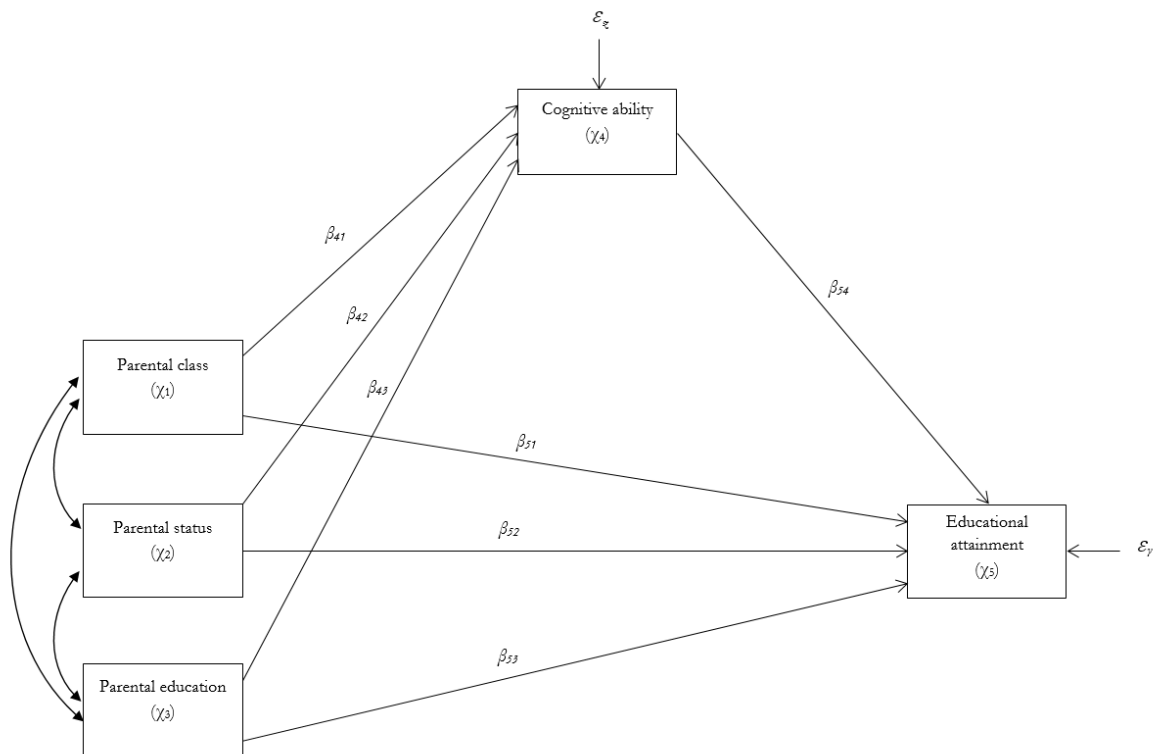
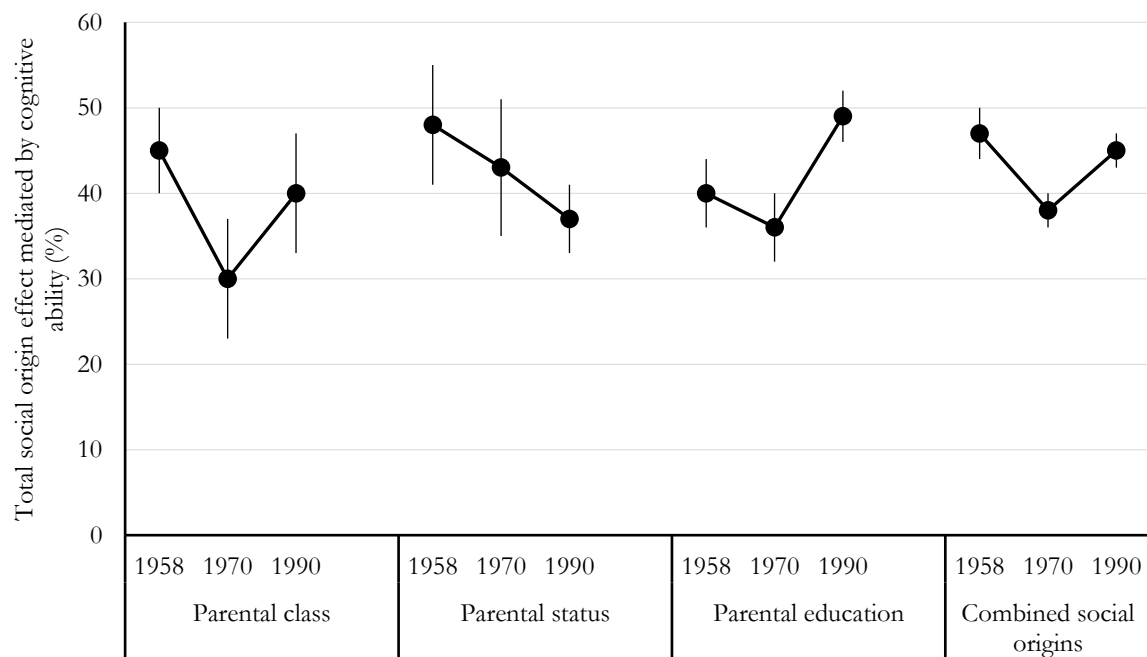
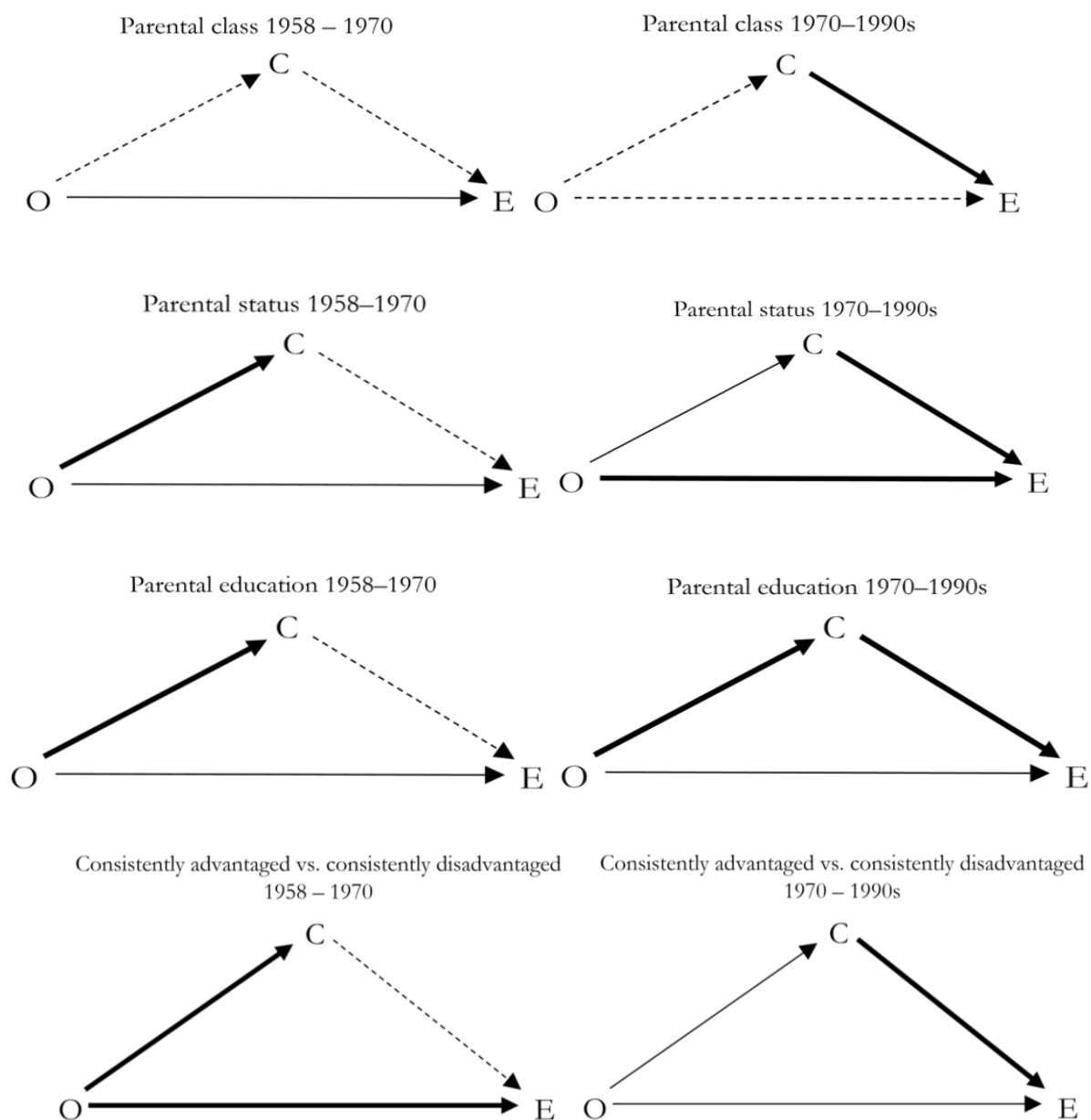


Figure 2. Proportion of the total effects of social origins on educational attainment at age 20 mediated by cognitive ability, by cohort (95% CIs)



Notes: All models control for gender, and are run separately by cohort. For combined social origins, estimates refer to the proportion of the total effect of belonging to the consistently advantaged group compared with belonging to the consistently disadvantaged group that is mediated by cognitive ability.

Figure 3. Cross-cohort change in the O–E, O–C and C–E associations



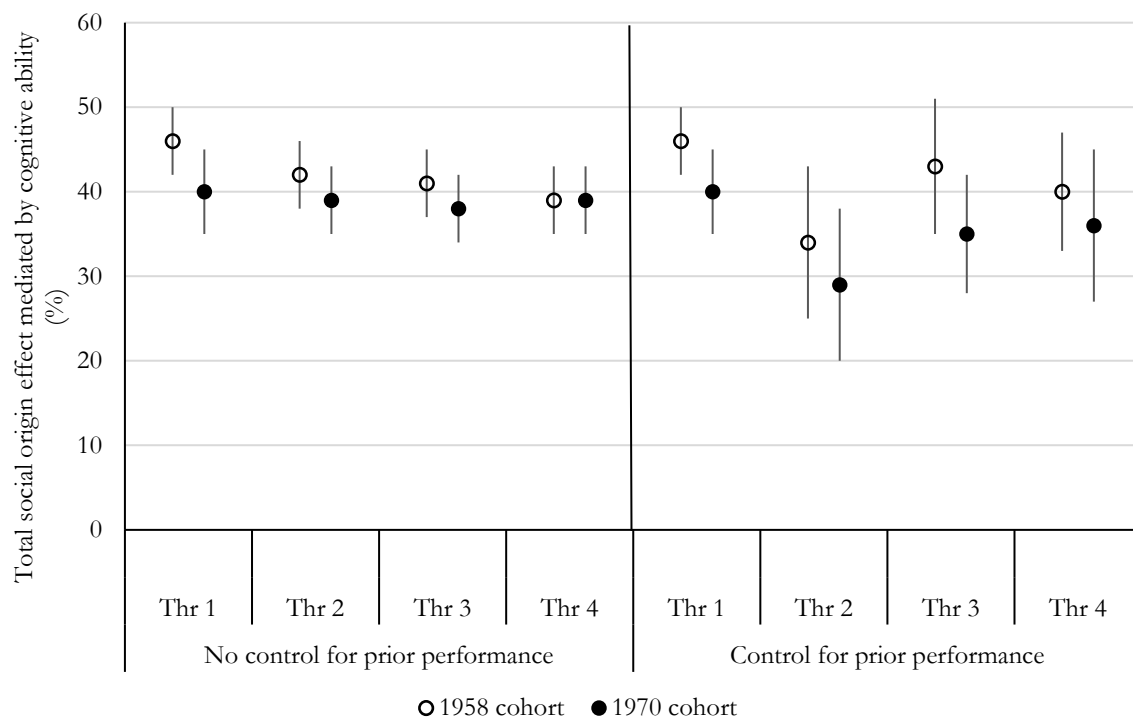
Weakening association

No change

Strengthening association

Notes: Changing strength of association assessed by comparing regression coefficients across cohort-specific models shown in Table 4.

Figure 4. The proportion of the total effect of combined social origins on four educational thresholds at age 38 mediated by cognitive ability, 1958 and 1970 cohorts (95% CIs)



Notes: All models control for gender. Estimates refer to the proportion of the total effect of belonging to the consistently advantaged group compared with belonging to the consistently disadvantaged group that is mediated by cognitive ability. None of the differences across the thresholds are significant at the 0.05 level. However, when controlling for prior performance, differences between thresholds 1 and 2 are significant at the 0.10 level for both cohorts.

Online Supplementary Appendix

Appendix A: Comparability of variables across the datasets

We use this appendix to describe the process of variable construction in each of the datasets we use. We took great efforts to harmonise our key variables in order to facilitate reliable cross-cohort analysis; as we demonstrate below, we have been able to achieve a high degree of comparability. We first present the extent of missing information in each of the datasets, which we address with multiple imputation (described in *Appendix C*). Then we give a detailed description of how each key variable was constructed, along with comments on how the distributions of the measures change across cohorts.

Missing data. Each of the cohort studies suffers from some sample attrition as well as item missing data. Table A1 summarises the amount of missing data in our key variables within each of the datasets.

Table A1. Missingness on key variables

	NCDS (1958 cohort)				BCS70 (1970 cohort)				ALSPAC (1990s cohort)			
	Per variable missing		Cumulative remaining		Per variable missing		Cumulative remaining		Per variable missing		Cumulative remaining	
	N	%	N	%	N	%	N	%	N	%	N	%
Initial sample	--	--	18558	100.00	--	--	18737	100.00	--	--	15445	100.00
Educational attainment	3947	21.27	14611	78.73	6600	35.22	12137	64.78	3361	21.76	12084	78.24
Cognitive ability	4427	23.85	12450	67.09	7342	39.18	8808	47.01	8825	57.14	5651	36.59
Parental social class	7400	39.87	9323	50.24	5604	29.91	7957	42.47	1862	12.06	5455	35.32
Parental social status	7400	39.87	9323	50.24	5592	29.84	7865	41.98	1862	12.06	5455	35.32
Parental education	7034	37.90	7078	38.14	5276	28.16	7752	41.37	1717	11.12	5429	35.15

Parental social class. In all datasets, the measure of parental class, which we classify according to the National Statistics Socio-Economic Classification (NS-SEC), is based on SOC2000 occupational codes and a five-category measure of employment status; i.e. we follow the ‘reduced method’ of derivation, as described in the ONS NS-SEC User Manual (2005). The measure for NCDS is based on fathers’ occupation only, as no information on mothers’ occupation was available. For the other datasets both parents’ occupations were considered (using the dominance approach (Erikson, 1984)). It should be noted that the datasets vary slightly with regards to the quality of the information provided; particularly the information on employment status given in the NCDS lacks detail. Table A2 displays the distributions of parental social class across the datasets.

The distribution of social class changes across the cohorts in ways we would expect: a steady increase in the proportions of those in higher and lower managerial and professional occupations; a decrease in the proportions of those in lower supervisory and technical occupations, as well as those in semi-routine and routine classes; a largely stable proportion making up the intermediate

class; and a fluctuating proportion of the class comprising small employers and own account workers.

Table A2. Distribution of parental social class (NS-SEC) by cohort (%)

	NCDS (1958)	BCS70 (1970)	ALSPAC (1990)
Class 1: Higher managerial and professional	8	11	21
Class 2: Lower managerial and professional	12	20	26
Class 3: Intermediate	9	9	13
Class 4: Small employers and own account workers	13	13	5
Class 5: Lower supervisory and technical	20	18	12
Class 6: Semi routine	16	14	14
Class 7: Routine	22	15	9
<i>Total</i>	100	100	100

Parental social status. Parents' social status is measured according to the Chan–Goldthorpe (2004) scale. SOC90 occupational codes are required for conversion to the scale. Parents' occupational information is coded according to different schemas in the different datasets. A dataset supplied by Paul Gregg *et al.* (2012) provides occupations coded for the NCDS cohort using the three-digit SOC90 codes. The BCS70 data include occupational codes using the three-digit OPCS Classification of Occupations 1980 (CO80), which we translated into three-digit SOC90 codes. The ALSPAC dataset includes SOC90 codes. The status measure in NCDS is based on fathers' occupation only, as no information on mothers' occupation was available. In the other datasets both parents' occupations were considered.

In order to allow for comparison across cohorts, the variables were normalised to a 0 to 1 scale using the theoretical minimum and maximum of -0.6 and +0.6 for all datasets. The measures are summarised in Table A3. As we would expect, we observe an over-time increase in parental status scores.

Table A3. Mean and standard deviation of parental Chan–Goldthorpe scores
(normalised, theoretical min/max)

	NCDS (1958)	BCS70 (1970)	ALSPAC (1990)
Mean	0.24	0.45	0.61
S.D	0.28	0.29	0.27

Parental education. The measure of parental education was generated by combining measures of educational attainment for mothers and fathers. Where the information from one parent was missing, the modal value for individuals with partners in a given educational attainment category was used. The available information on mothers' and fathers' educational attainment differs across surveys. The NCDS only provides information on the age at which mothers and fathers left full time education, so our measures of qualifications are estimated from this. BCS70 and ALSPAC measure qualifications attained, but differ somewhat with respect to the specific qualification categories given. The qualifications scale we derived using the available information is given in Table A4 with the distributions for each cohort.

Table A4. Distribution of combined measure of parental education (%)

	NCDS (1958)	BCS70 (1970)	ALSPAC (1990s)
7. Both parents have degree-level qualifications	2	2	10
6. One parent has degree-level qualification; other parent has lower qualification	4	11	14
5. Both parents have higher secondary or lower tertiary qualification	2	2	7
4. One parent has higher secondary or lower tertiary qualification; other parent has lower qualification	8	12	20
3. Both parents have secondary or lower qualification	22	20	37
2. One parent has secondary or lower qualification; other parent has no qualification	21	20	8
1. Neither parent has any qualification	41	33	4
<i>Total</i>	100	100	100

As would be expected over decades of educational expansion, the proportions of children with parents in the highest educational category – those who both have degrees – has increased over time, while the proportions of children whose parents have no qualifications between them has decreased.

To allow for meaningful cross-cohort comparisons, we treat parental education in relative terms. A relative measure of parental education was generated by assigning scores to each category of the composite measure of parental education according to the percentage of parents falling below that level in the cumulative distribution. Table A5 presents the mean and standard deviation of this relative measure across the surveys.

Table A5. Mean and standard deviation of relative measure of parental education (normalised, 0–1)

	NCDS (1958)	BCS70 (1970)	ALSPAC (1990s)
Mean	0.38	0.40	0.44
S.D	0.35	0.33	0.35

Respondent cognitive ability. In each of the surveys, cognitive ability was measured via scores attained on a range of tests administered by survey interviewers or psychologists. In NCDS, scores from a general ability test taken at age 11, comprising the assessment of both verbal and non-verbal skills, were used (see Schoon (2008) for more information on these tests). Table A6, which lists the names of the variables taken from the original data files for each cohort, shows the two measures from NCDS used, which are summed scores across the range of items. In BCS70, scores from four subtests of the British Ability Scales (BAS), taken at age 10, were used: Word Definitions and Word Similarities to measure verbal ability, and Recall of Digits and Matrices to measure non-verbal ability. In ALSPAC, scores from the Wechsler Intelligence Scale for Children – Third Edition (WISC-III), taken at age 8, were used. These tests returned scores on 11 items measuring both verbal skills (Information, Similarities, Arithmetic, Vocabulary, and Comprehension) and performance skills (Digit span, Picture completion, Coding, Picture Arrangement, Block Design and Object Assembly).

Table A6. Cognitive ability – Variable names and descriptions as provided in original data files

Dataset	Variable name	Description
NCDS	n914	Verbal score on general ability test
	n917	Non-verbal score on general ability test
BCS70	i3504 – i3540	BAS word definitions test (37 items)
	i3575 – i3616	BAS word similarities test (42 items)
	i3541 – i3574	BAS recall of digits test (34 items)
	i3617 – i3644	BAS matrices test (28 items)
ALSPAC	f8ws050 – f8ws060	WISC-III: All subtest scaled scores (11 items – excluding two digit span)
	f8ws100	WISC-III: Sum of all verbal subtests
	f8ws101	WISC-III: Sum of all performance subtests

We seek to capture the notion of ‘general intelligence’ – often referred to as ‘*g*’ (Spearman, 1904). This rests on correlations among test scores and, according to Colom *et al.* (2002), should be distinguished from ‘intelligence in general’ – or IQ – which rests on the summation of standardised test scores. Capturing the common variance across ability tests is the approach that has been commonly employed in previous studies using NCDS and BCS70 (e.g. Schoon, 2008, 2010), as well as in studies using ALSPAC (e.g. Kaufman 1994; Kamphaus 1993; Roid *et al.*, 1993), and is seen as the closest approximation to determining the latent measure of intelligence ‘*g*’ (Deary, 2001).

We are aware of the distinctions that have been drawn between different types of intelligences. Cattell (1971), some time ago, drew the later-influential distinction between ‘fluid’ intelligence, which represents individuals’ innate capacity to reason abstractly and apply logic, and ‘crystallized’

intelligence, which represents individuals' ability to draw on the skills, knowledge and experience they have gained throughout their life-course when solving problems. The former is captured through tests of number abilities and matrix problems, and so on, while the latter is captured via tests of vocabulary, reading comprehension and general knowledge. Most tests of cognitive ability, via their non-verbal and verbal components respectively, aim to capture both types of intelligences (see, e.g., Prifitera *et al.*, 2008). The view shared by most psychometric researchers is that, even with this understanding of intelligence as hierarchically structured, while sub-dimensions can be analysed in isolation they do not replace 'g' and moreover, there is often no reason to do so (Marks, 2014: 53; Jensen, 1998). As Marks (2014: 53) points out, 'one point that is often overlooked in the debates about dimensionality of intelligence is that highly correlated multiple dimensions are not useful for the analysis of the influence of ability, vis-à-vis other factors, on educational and socioeconomic outcomes. In regression-type analyses, high multicollinearity – a direct consequence of the highly correlated dimensions – will produce erroneous and useless estimates'.

For each cohort, a principal component analysis (PCA) was carried out on the subtests, using the measures specified in Table A6. In each case, scores were saved from the first unrotated component extracted and this was taken to represent cognitive ability. Table A7 gives the results of the PCAs, presenting Eigenvalues, factor loadings, the percentage of variance across measures that is accounted for by the first principal component score, and summaries of the final variables generated. Using ALSPAC, for the purposes of a robustness check, we also constructed a second cognitive ability measure via a PCA with the composite variables 'f8ws100' and 'f8ws101', as shown in Table A6 (this is labelled WISC-2 in Table A7). For the main analysis, however, we use the measure labelled WISC-11.

Table A7. Cognitive ability measures and PCA results

Measure (ALSPAC only)	Eigen.	PCA				New variable		
		Loading		Variance explained		Min	Max	Mean (SD)
NCDS	1.81	Verbal	Non-verbal	0.90		-3.61	3.13	0.00 (1.34)
		0.71	0.71					
BCS70	2.31	Word definitions	Word similarities	0.58		-5.64	5.13	0.00(1.52)
		0.55	0.55					
ALSPAC	WISC-2 WISC-11	Verbal	Performance	0.75 0.33		-5.07 -7.58	3.84 6.13	0.00 (1.23) 0.00 (1.92)
		0.71 --	0.71 --					

Notes: Factor loadings for WISC-11 have been omitted since there are too many to list.

Figures A1–A3 give the percentile distributions of each cognitive ability measure across the datasets. Expectedly, scores tend to be normally distributed within each of the datasets. In some analyses, we instead use cognitive ability quintiles.

Figure A1. Distribution of cognitive ability: NCDS

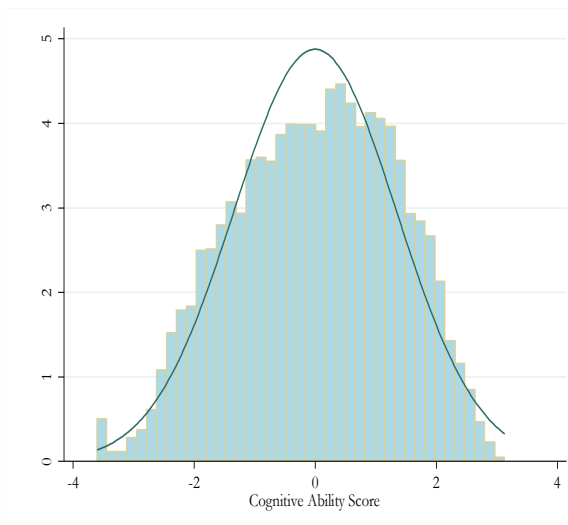


Figure A2. Distribution of cognitive ability: BCS70

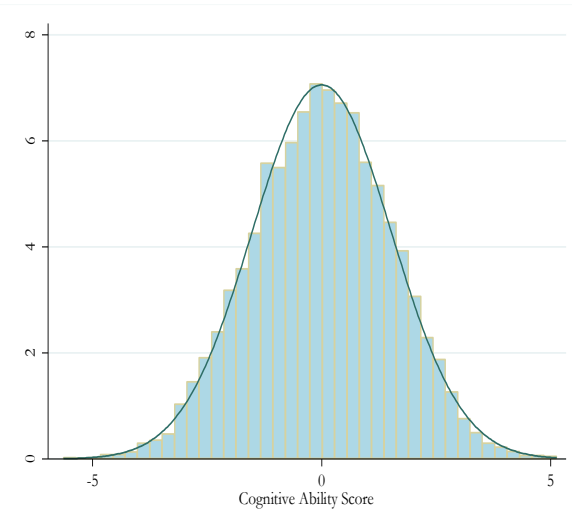
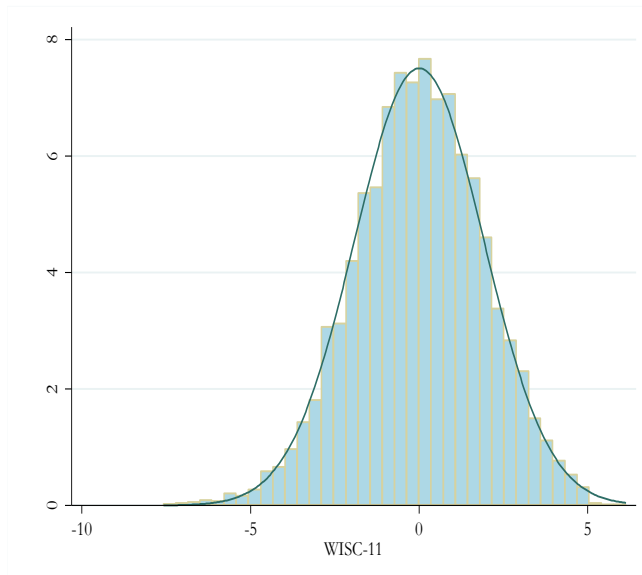


Figure A3. Distribution of cognitive ability: ALSPAC



Respondent educational attainment. The measures for respondents' educational attainment were generated by using the information on educational qualifications attained by cohort members that was given in each survey of the NCDS and BCS70. A key consideration in generating these measures was how to treat individuals for whom no educational qualifications had been recorded. This lack of information on qualifications may signal that the cohort member had not attained any education at all, or it may constitute a case of missing information. This was addressed by coding the information on the educational attainment of individuals as 'missing' where individuals did not participate in the relevant survey wave, and coding all other individuals as not having gained any educational qualification. For ALSPAC, information on educational attainment was determined using linked administrative data from the National Pupil Database (NPD).

In the British educational system, all individuals undertake 12 years of schooling which culminates in compulsory national examinations at the end of lower secondary education, at age 16. These examinations – currently called the General Certificate of Secondary Education (GCSE), but

previously GCE Ordinary Level (O-level)¹⁶ – involve students sitting tests of the core subjects, English, Maths and Science, as well as a number of optional subjects in addition to these. GCSEs (and O-levels) are graded between U (‘unclassified’; i.e. a no-show or nothing written on the exam paper) and A* at the highest, with achieved A*–C grades being considered passes. Five A*–C grade GCSEs has long been taken as a benchmark of performance at this level – it is commonly cited in league tables and is often used as the minimum entry criterion for admission to further education institutions.¹⁷

After GCSEs, individuals who stay in the education system will pursue Advanced Level (A-level) qualifications (the academic route) at the upper-secondary level, or other further education qualifications (an occupational or vocational route). Most individuals that pursue the academic route will undertake a two-year course, sitting AS-level exams in year 1 and, given successful completion of year 1, A-level exams in year 2. The non-academic post-16 educational route is highly complex in Britain. There is no unified system of vocational education, as is found in countries such as Germany; there are a vast number of qualifications for which their value in the labour market is often unclear (Hansen & Vignoles, 2005). National Vocational Qualifications (NVQs), require individuals to demonstrate their competence with relevance to a particular occupation. These have come to form the basis of a scale of ordered qualifications against which the level of other vocational qualifications can be compared. We represent vocational qualifications by their NVQ equivalencies.

Following further education, individuals who stay on to study will pursue tertiary-level qualifications. The most prestigious of these, and that which typically carries the highest value in the labour market, is the academic degree. But there are other tertiary-level qualifications that are commonly pursued – again usually with an occupational or vocational orientation – such as diplomas, and teaching or nursing degrees (referred to as tertiary-level sub-degrees in Table 1).

Table A8 shows the distribution of educational attainment at age 20 and Table A9 shows the distribution of educational attainment at age 38 for the older two cohorts. We observe a general upgrading in educational attainment over time.

Table A8. Distribution of educational attainment at age 20 (%)

	NCDS (1958)	BCS70 (1970)	ALSPAC (1990s)
1. 2+ A-level passes	19	19	40
2. 5+ O-level passes or 1 A-level pass, NVQ 3	18	25	22
3. 1–4 O-level passes, NVQ 2	21	22	20
4. Below O-level, NVQ 1	11	8	14
5. No qualifications	30	25	4
<i>Total</i>	100	100	100

¹⁶ Before 1988, less academic pupils sat the Certificate of Secondary Education (CSE) at age 16, while others sat O-levels. These were combined to form GCSEs in 1988 which meant that from that time, all pupils were to sit the same examinations, without having to choose in advance between the lower- or higher-level option. Since the introduction of GCSEs, far more pupils now leave school with at least some qualifications (Hansen & Vignoles, 2005).

¹⁷ The average number of GCSEs taken nowadays is approximately 8 (ONS, 2012).

Table A9. Distribution of educational attainment at age 38 (%)

	NCDS (1958)	BCS70 (1970)
1. Degree or higher degree, NVQ 5 or 6	12	19
2. Tertiary-level sub-degree, NVQ 4	13	14
3. 2+ A-level passes	4	3
4. 5+ O-level passes or 1 A-level pass, NVQ 3	17	21
5. 1–4 O-level passes, NVQ 2	22	21
6. Below O-level, NVQ 1	12	7
7. No qualifications	20	15
<i>Total</i>	100	100

Appendix B: How far can ALSPAC be taken as a good approximation to a nationally representative sample, using LSYPE as a comparator?

We use this appendix to consider the extent to which ALSPAC can adequately approximate a representative sample of Britain for this cohort, by comparing the distributions of our key variables with identically-constructed measures from the LSYPE. Further, we address the question of the extent to which the underlying associations between variables are also similar across the datasets. First, Table B1 presents the distributions of each of our key variables in both ALSPAC and LSYPE (cognitive ability is excluded since it is not available in LSYPE).

Overall, we find distributions that are very similar across the two samples, but we note the following discrepancies. In ALSPAC, there are larger proportions in the higher managerial and professional class, coupled with smaller proportions in the routine class, compared with LSYPE. The slightly higher socioeconomic position of ALSPAC parents compared with those in the national population has previously been documented (Fraser *et al.*, 2012). Further, there are slightly fewer self-employed parents than we might expect (small employers and own account workers). Similarly, we find that ALSPAC parents have slightly higher social status scores and educational qualifications on average.

	ALSPAC	LSYPE
<i>Parental social class (%)</i>		
Class 1: Higher managerial and professional	21	16
Class 2: Lower managerial and professional	26	29
Class 3: Intermediate	13	7
Class 4: Small employers and own account workers	5	9
Class 5: Lower supervisory and technical	12	14
Class 6: Semi routine	14	13
Class 7: Routine	9	12
<i>Parental social status (Mean (SD))</i>		
Using theoretical min/max, rescaled 0–1	0.61 (0.27)	0.57 (0.27)
<i>Parental education (Mean (SD))</i>		
Using relative scale, rescaled 0–1	0.44 (0.35)	0.40 (0.33)
<i>Respondent's educational attainment – age 20 (%)</i>		
2 + A-level passes, NVQ 4	38	39
5+ O-level passes or 1 A-level pass, NVQ 3	22	28
1 – 4 O-level passes, NVQ 2	21	16
Below O-level, NVQ 1 / No qualifications	19	17

Now we present the results of a set of binary logistic regression models predicting individuals' likelihood of exceeding a series of educational thresholds, comparing the results obtained between ALSPAC and LSYPE. Only cohorts and social origins are included as independent variables, with a control for gender. Table B2 presents average marginal effects from the first set of models which takes as the outcome variable a binary measure indicating whether individuals achieved at least 1–4 O-level passes/NVQ 2 (=1) or any lower level of attainment (=0).

Table B2. Comparing ALSPAC and LSYPE – The probability of attaining 1–4 O-level passes/NVQ 2 vs. any lower attainment (AMEs from binary logistic regression models)

	NCDS, BCS70, ALSPAC		NCDS, BCS70, LSYPE	
	M1	M2	M1	M2
<i>Cohort</i>				
NCDS	0.02**	0.01	0.02**	0.01
BCS70	ref	ref	ref	ref
ALSPAC/LSYPE	0.12**	0.10**	0.13**	0.12**
<i>Parental class</i>				
Class 1	ref	ref	ref	ref
Class 2	-0.05**	-0.04**	-0.03**	-0.03**
Class 3	-0.04**	-0.03**	-0.04**	-0.03**
Class 4	-0.09**	-0.08**	-0.07**	-0.06**
Class 5	-0.07**	-0.06**	-0.08**	-0.07**
Class 6	-0.11**	-0.10**	-0.11**	-0.10**
Class 7	-0.14**	-0.13**	-0.15**	-0.14**
<i>Parental status</i>				
0–1	0.16**	0.16**	0.14**	0.14**
<i>Parental education</i>				
0–1	0.17**	0.18**	0.17**	0.19**
<i>Parental class*Cohort</i>				
Class 1*NCDS		ref		ref
Class 2*NCDS		-0.03		-0.03
Class 3*NCDS		-0.03		-0.03
Class 4*NCDS		-0.09		-0.09
Class 5*NCDS		-0.07		-0.07
Class 6*NCDS		-0.13		-0.13
Class 7*NCDS		-0.20		-0.20
Class 1*ALSPAC/LSYPE		ref		ref
Class 2*ALSPAC/LSYPE		-0.05		-0.01
Class 3*ALSPAC/LSYPE		-0.04		-0.02
Class 4*ALSPAC/LSYPE		-0.09		-0.01
Class 5*ALSPAC/LSYPE		-0.04		-0.05
Class 6*ALSPAC/LSYPE		-0.06		-0.07
Class 7*ALSPAC/LSYPE		-0.09		-0.11
<i>Parental status*Cohort</i>				
*NCDS		0.13		0.14
*ALSPAC/LSYPE		0.25**		0.17
<i>Parental education*Cohort</i>				
*NCDS		0.13		0.13
*ALSPAC/LSYPE		0.24		0.26
N	28164	28164	24997	24997

Notes: Gender is also included.

Generally, we find results that are more-or-less consistent across ALSPAC and LSYPE. All social origin variables have independent and significant effects in M1, and the coefficients for the main effect of both cohorts are similar in magnitude. When interactions are included in models M2, these similarities remain. The cohorts-by-class and cohorts-by-parental education interaction terms are non-significant for both samples; however, the cohorts-by-status interaction is significant and

positive for ALSPAC, while it is not for LSYPE. Finally, we also find some differences between the samples for the self-employed, in the size of the coefficients for the interaction effects.

Table B3 presents the second set of models, which takes as the outcome variable a binary measure indicating whether individuals achieved at least 5+ O-level/GCSE passes or 1 A-level pass (=1), versus any lower level of attainment (=0).

Table B3. Comparing ALSPAC and LSYPE – The probability of attaining 5+ O-level passes or 1 A-level pass/NVQ 3 vs. any lower attainment (AMEs from binary logistic regression models)

	NCDS, BCS70, ALSPAC		NCDS, BCS70, LSYPE	
	M1	M2	M1	M2
<i>Cohort</i>				
NCDS	0.03**	0.02**	0.02**	0.02**
BCS70	ref	ref	ref	ref
ALSPAC/LSYPE	0.12**	0.10**	0.18**	0.18**
<i>Parental class</i>				
Class 1	ref	ref	ref	ref
Class 2	-0.07**	-0.06**	-0.06**	-0.06**
Class 3	-0.08**	-0.07**	-0.08**	-0.08**
Class 4	-0.12**	-0.12**	-0.10**	-0.10**
Class 5	-0.12**	-0.11**	-0.14**	-0.14**
Class 6	-0.16**	-0.15**	-0.17**	-0.16**
Class 7	-0.20**	-0.18**	-0.22**	-0.21**
<i>Parental status</i>				
0–1	0.21**	0.21**	0.16**	0.16**
<i>Parental education</i>				
0–1	0.27**	0.28**	0.24**	0.25**
<i>Parental class*Cohort</i>				
Class 1*NCDS		ref		ref
Class 2*NCDS		-0.10		-0.09
Class 3*NCDS		-0.11		-0.11
Class 4*NCDS		-0.18		-0.18
Class 5*NCDS		-0.18		-0.18
Class 6*NCDS		-0.25**		-0.25**
Class 7*NCDS		-0.30**		-0.30**
Class 1*ALSPAC/LSYPE		ref		ref
Class 2*ALSPAC/LSYPE		-0.06		-0.04
Class 3*ALSPAC/LSYPE		-0.05		-0.05
Class 4*ALSPAC/LSYPE		-0.12		-0.04
Class 5*ALSPAC/LSYPE		-0.06		-0.12
Class 6*ALSPAC/LSYPE		-0.10		-0.12
Class 7*ALSPAC/LSYPE		-0.10**		-0.16
<i>Parental status*Cohort</i>				
*NCDS		0.13		0.13
*ALSPAC/LSYPE		0.33**		0.21
<i>Parental education*Cohort</i>				
*NCDS		0.19**		0.20**
*ALSPAC/LSYPE		0.35		0.31
N	28164	28164	24977	24977

Notes: Gender is also included.

Here we find that the main effect cohort coefficient for LSYPE is larger than for ALSPAC. Again, the class-by-cohorts interaction effects are mostly insignificant. Again, we see small discrepancies between the samples; most notably in the size of the coefficients for Classes 4, 5 and 7. The parental education-by-cohorts interaction is non-significant in both models and, again, the parental status-by-cohort interaction is significant for ALSPAC but not for LSYPE.

Table B4 presents the final set of models, which take as the outcome variable a binary measure indicating whether individuals attained at least 2+ A-level passes/NVQ 4 (=1), versus any lower level of education (=0).

Here there are fewer differences between the samples. Consistent with the previous two thresholds considered, we find that the negative influence of being in Class 4, relative to Class 1, is more pronounced in ALSPAC, and the negative influence of being in Class 5 is more pronounced in LSYPE; but nonetheless, the class-by-cohorts interaction effects are non-significant in both samples. The positive effect of higher parental status scores and higher parental education are, for this threshold, significant in both samples.

Overall, the social origin indicators in ALSPAC and LSYPE samples are largely similar with regards to their influence on the likelihood of exceeding each educational threshold we consider. The most notable differences exist in relation to the changing role of parental status; we find that the effect of status is stronger in the ALSPAC sample.

Table B4. Comparing ALSPAC and LSYPE – The probability of attaining 2 + A-level passes/NVQ 4 vs. any lower attainment (AMEs from binary logistic regression models)

	NCDS, BCS70, ALSPAC		NCDS, BCS70, LSYPE	
	M1	M2	M1	M2
<i>Cohort</i>				
NCDS	0.06**	0.06**	0.06**	0.05**
BCS70	ref	ref	ref	ref
ALSPAC/LSYPE	0.15**	0.14**	0.15**	0.15**
<i>Parental class</i>				
Class 1	ref	ref	ref	ref
Class 2	-0.06**	-0.06**	-0.06**	-0.06**
Class 3	-0.06**	-0.06**	-0.06**	-0.06
Class 4	-0.10**	-0.11**	-0.08**	-0.08**
Class 5	-0.11**	-0.11**	-0.13**	-0.13**
Class 6	-0.13**	-0.13**	-0.13**	-0.13**
Class 7	-0.15**	-0.14**	-0.16**	-0.16**
<i>Parental status</i>				
0–1	0.17**	0.18**	0.14**	0.14**
<i>Parental education</i>				
0–1	0.27**	0.26**	0.23**	0.23**
<i>Parental class*Cohort</i>				
Class 1*NCDS		ref		ref
Class 2*NCDS		-0.07		-0.07
Class 3*NCDS		-0.08		-0.08
Class 4*NCDS		-0.14		-0.13
Class 5*NCDS		-0.15		-0.15
Class 6*NCDS		-0.17		-0.17
Class 7*NCDS		-0.22		-0.22
Class 1*ALSPAC/LSYPE		ref		ref
Class 2*ALSPAC/LSYPE		-0.05		-0.06
Class 3*ALSPAC/LSYPE		-0.05		-0.05
Class 4*ALSPAC/LSYPE		-0.12		-0.03
Class 5*ALSPAC/LSYPE		-0.07		-0.14
Class 6*ALSPAC/LSYPE		-0.10		-0.11
Class 7*ALSPAC/LSYPE		-0.10		-0.13
<i>Parental status*Cohort</i>				
*NCDS		0.11		0.11
*ALSPAC/LSYPE		0.28**		0.21**
<i>Parental education*Cohort</i>				
*NCDS		0.21		0.20
*ALSPAC/LSYPE		0.34**		0.30**
N	28164	28164	24977	24977

Notes: Gender is also included.

Appendix C: Multiple Imputation

We consider multiple imputation (MI) to be a necessary and useful exercise for the purposes of this study. We wish to make full use of the information available in the data. As a consequence of sample attrition (see Table A1), a complete-case analyses would have involved discarding information from around half of cases in NCDS and BCS70 and around two-thirds of cases in ALSPAC, which we see as unacceptable. In using MI, we also aim to improve the reliability of comparisons across datasets by limiting variations in sample selection and bias as far as possible. The case of missing data in ALSPAC is most severe, in part due to the high frequency of interviews. Cognitive ability, one of our key measures, has a particularly large amount of missing information (57% missing) which has been addressed via MI by other users of ALSPAC in the past (e.g. Washbrook *et al.*, 2014).

We took great care in the specification of our imputation models. MI relies on the missing at random (MAR) assumption, which cannot be directly tested but we believe that, for reasons given in Washbrook *et al.* (2014: 768–769) for ALSPAC, and considered by Kuha (2013) for NCDS and BCS70, it is highly plausible in this instance. Our imputation models included all the variables that entered the analyses model, as well as sets of relevant auxiliary variables. Examples of the auxiliary variables we used include scores gained by respondents in other tests of IQ and academic competence, scores on tests of non-cognitive skills, alternative operationalisations of parental socioeconomic status, and measures of the family's financial circumstances. Including auxiliary variables makes the missing at random assumption more plausible, improves the accuracy of imputed values (von Hippel & Lynch, 2013), and is of particular importance when imputing the dependent variable as we do here. Using imputed values of dependent variables in analysis models is acceptable practice in the presence of good auxiliary information (Johnson & Young, 2011) and we consider it preferable to excluding those cases.

Tables C1 and C2 show that the distributions of most of our variables are very similar before and after MI. An exception to this was the cognitive ability measure in ALSPAC. After imputing, we found fewer cases in the top quintile and more cases in the bottom quintile, which was probably indicative of the likelihood of missingness being higher for those with low cognitive ability (providing further justification for using MI). We therefore constructed the cognitive ability measures post-imputation, using Stata's *mi* passive command, which corrected the issue. The distributions given are each based on sets of 20 multiply imputed datasets. The MI procedure, in short, involves estimating the statistic of interest using each of the datasets, and then combining these using Rubin's rules (Rubin, 1987). Using MI increased the size of the sample in the NCDS dataset from 7,708 (complete cases) to 18,558 (with MI), the BCS70 dataset from 7,752 to 18,737 and the ALSPAC dataset from 5,429 to 15,242.

Table C1 presents the distribution of our educational attainment measures, before and after MI, for each cohort individually.

Table C1. Distribution of dependent variables, before and after multiple imputation

	NCDS		BCS70		ALSPAC	
	Pre-MI	Post-MI	Pre-MI	Post-MI	Pre-MI	Post-MI
Educational attainment at age 20 (Mean)	2.88	2.86	3.09	3.05	3.76	3.79
Educational attainment at age 38 (%)						
Exceeded threshold 1	48.14	46.06	57.44	56.98		
Exceeded threshold 2	31.49	29.54	36.40	35.63		
Exceeded threshold 3	27.44	25.73	33.76	32.97		
Exceeded threshold 4	13.22	12.31	19.27	18.54		

Table C2 presents the distribution of our focal independent variables (in each of the different ways they are used in analyses), before and after multiple imputation, for each cohort individually.

Table C2. Descriptive statistics for independent variables, before and after multiple imputation

	NCDS		BCS70		ALSPAC	
	Pre-MI	Post-MI	Pre-MI	Post-MI	Pre-MI	Post-MI
<i>Cognitive ability quintiles (%)</i>						
Top	19.98	20.09	19.99	20.00	20.18	20.00
4 th	19.92	19.94	20.01	20.00	19.98	20.00
3 rd	20.03	20.01	20.00	20.00	20.00	20.00
2 nd	20.06	19.99	20.00	20.00	19.98	20.00
Bottom	20.01	19.97	20.00	20.00	19.85	20.00
<i>Parental class</i>						
Hierarchical component, rescaled 0–1 (mean)	0.42	0.42	0.50	0.49	0.59	0.58
<i>Parental status</i>						
Approximation to quintiles, rescaled 0–1 (mean)	0.48	0.47	0.48	0.47	0.49	0.48
<i>Parental education</i>						
Approximation to quintiles, rescaled 0–1 (mean)	0.30	0.30	0.39	0.39	0.45	0.44
<i>Combined social origins (%)</i>						
Consistently advantaged	20.32	19.96	25.00	24.20	28.43	27.23
Intermediate	52.08	51.66	46.24	45.71	43.11	42.75
Consistently disadvantaged	27.60	28.38	28.75	30.09	28.46	30.02

Appendix D: Additional information and analyses

Table D1. Construction of the measure of combined origin

	Components of parental background			Distributions by cohort (%)		
	Parental class	Parental status	Parental education	NCDS	BCS70	ALSPAC
Consistently advantaged	Class 1/2	Tertile 1	Tertile 1	20	24	27
	Class 1/2	Tertile 2	Tertile 1			
	Class 1/2	Tertile 1	Tertile 2			
	Class 3/4/5	Tertile 1	Tertile 1			
Intermediate	All other combinations of the three components of social origins			52	46	43
Consistently disadvantaged	Class 6/7	Tertile 3	Tertile 3	28	30	30
	Class 6/7	Tertile 2	Tertile 3			
	Class 6/7	Tertile 3	Tertile 2			
	Class 3/4/5	Tertile 3	Tertile 3			

Research Question 1

Table D2. Results of sequential linked regression analyses using combined measure of social origins (intermediate as reference category) - unstandardised regression coefficients

	NCDS		BCS70		ALSPAC	
	Coef.	SE	Coef.	SE	Coef.	SE
<i>DV: Educational attainment</i>						
Cognitive ability	0.49	0.01	0.33	0.01	0.44	0.01
Consistently advantaged	0.44	0.03	0.47	0.03	0.37	0.02
Intermediate (ref.)	--	--	--	--	--	--
Consistently disadvantaged	-0.25	0.03	-0.31	0.03	-0.42	0.02
R ²	0.28		0.19		0.44	
N	18558		18737		15242	
<i>DV: Cognitive ability</i>						
Consistently advantaged	0.74	0.03	0.8	0.03	0.82	0.04
Intermediate (ref.)	--	--	--	--	--	--
Consistently disadvantaged	-0.5	0.03	-0.65	0.03	-0.68	0.03
R ²	0.10		0.14		0.15	
N	18558		18737		15242	

Notes: Models control for gender and are run separately by cohorts.

Table D3. Results of sequential linked regression analyses, taking each component of social origins separately - standardised regression coefficients

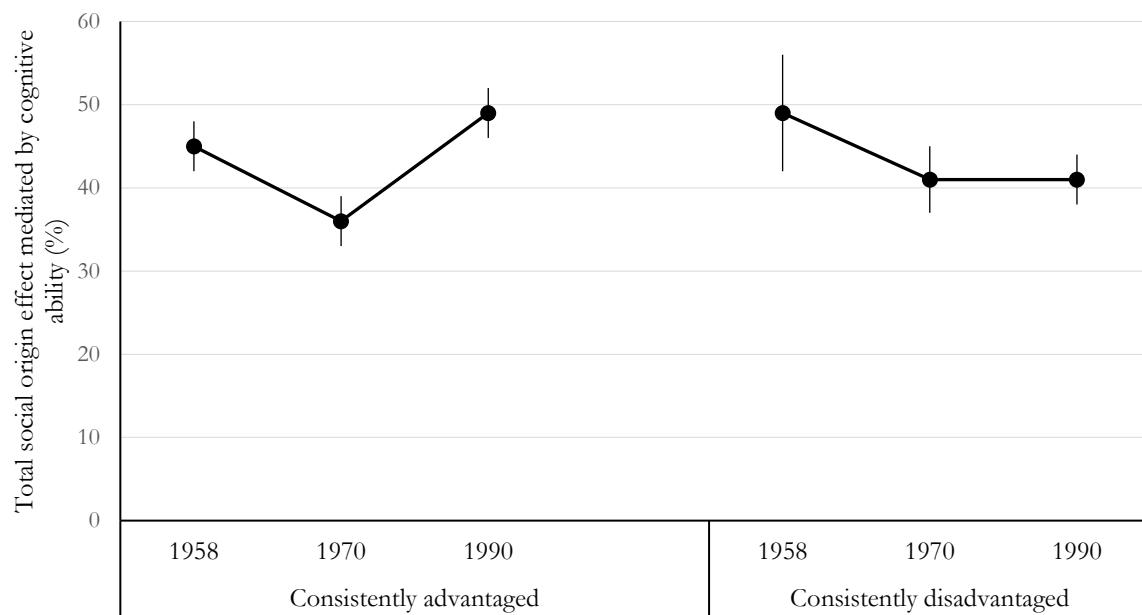
	NCDS		BCS70		ALSPAC	
	Coef.	SE	Coef.	SE	Coef.	SE
<i>DV: Educational attainment</i>						
Cognitive ability	0.45	0.01	0.31	0.01	0.48	0.01
Parental class	0.09	0.05	0.07	0.06	0.04	0.04
Parental status	0.05	0.05	0.06	0.05	0.13	0.04
Parental education	0.11	0.04	0.12	0.05	0.15	0.04
R ²	0.29		0.19		0.45	
N	18558		18737		15242	
<i>DV: Cognitive ability</i>						
Parental class	0.17	0.06	0.10	0.06	0.06	0.07
Parental status	0.10	0.04	0.13	0.05	0.16	0.05
Parental education	0.17	0.05	0.22	0.04	0.30	0.06
R ²	0.12		0.15		0.21	
N	18558		18737		15242	

Notes: Models control for gender and are run separately by cohorts.

Table D4. Proportion of the total effects of social origins on educational attainment at age 20 mediated by cognitive ability by cohort, based on standardised regression coefficients

Mediation %	NCDS	BCS70	ALSPAC
Parental class	45	30	40
Parental status	48	43	37
Parental education	40	36	49

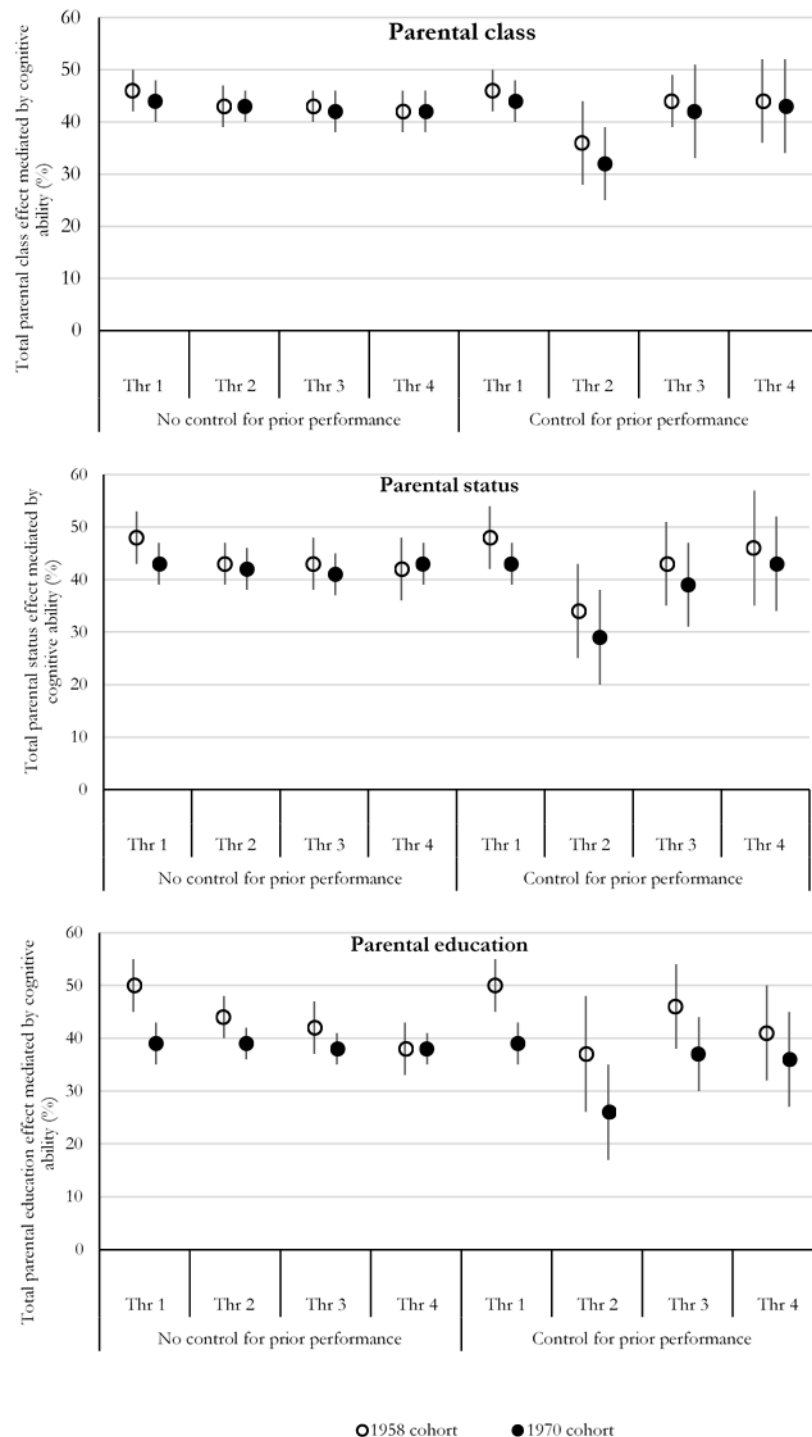
Figure D1. Mediation percentages for combined social origins (intermediate group as reference, as Table D4 (95% CIs)



Notes: All models control for gender and are run separately by cohort. For combined social origins, estimates refer to the proportion of the total effect of belonging to the consistently advantaged group compared with belonging to the intermediate group, and the proportion of the total effect of belonging to the consistently disadvantaged group compared with belonging to the intermediate group that is mediated by cognitive ability.

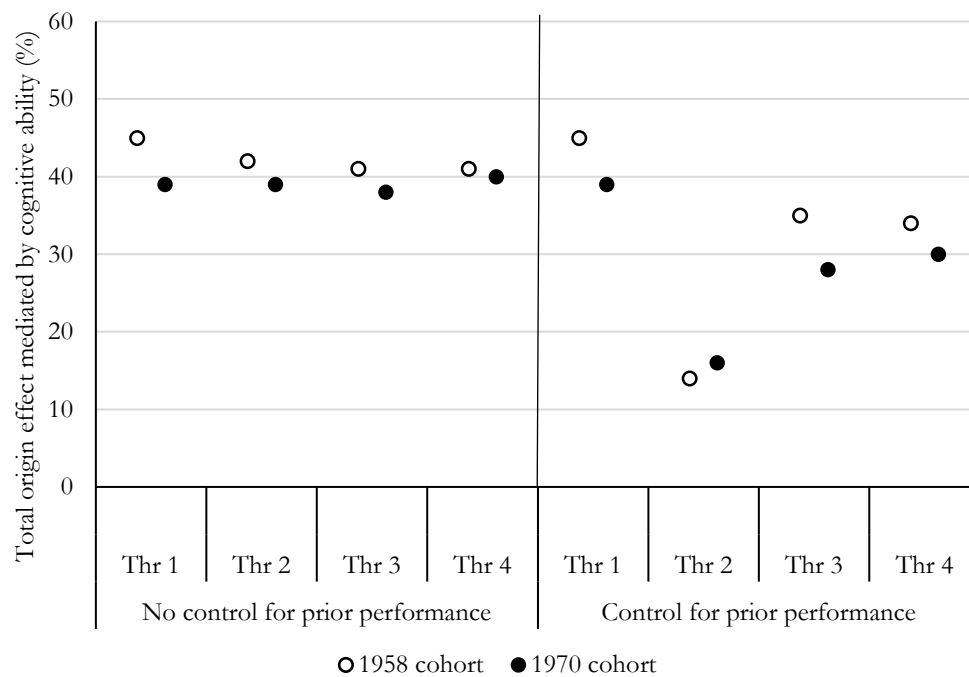
Research Question 2

Figure D2. The total effect of each component of social origins on four educational thresholds at age 38 mediated by cognitive ability



Notes: Estimates refer to the proportion of the total effect of belonging to highest social origin group (treated as a threefold collapse in the case of parental class and approximations to tertiles in the case of parental status and education) compared with belonging to the lowest social origin group that is mediated by cognitive ability. Controls included for gender and model run separately by cohorts. None of the differences across the thresholds are significant at the 0.05 level. However, when controlling for prior performance, differences in the mediation percentages between thresholds 1 and 2 are significant at the 0.10 level for both cohorts, and for each social origin component.

Figure D3. The proportion of the total effect of combined social origins on four educational thresholds at age 38 mediated by cognitive ability (using the KHB method)



Notes: Estimates refer to the proportion of the total effect of belonging to the consistently advantaged group compared with belonging to the consistently disadvantaged group that is mediated by cognitive ability. Software for the KHB method does not provide a means for calculating standard errors, so significance tests have not been carried out for this analysis.

References

- Cattell, R. B. (1971). *Abilities: Their Structure, Growth and Action*. Boston, MA: Houghton Mifflin.
- Colom, R., Abad, F. J., Garcia, L. F. & Juan-Espinosa, M. (2002). Education, Wechsler's Full Scale and g, *Intelligence*, 30, 449-462.
- Deary, I. J. (2001). *Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.
- Erikson, R. (1984). Social Class of Men, Women and Families, *Sociology*, 18, 500-514.
- Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M. & Lawlor, D. A. (2012). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers Cohort, *International Journal of Epidemiology* 42, 97-110.
- Hansen, K. & Vignoles, A. (2005). The United Kingdom Education System in a Comparative Context. In S. Machin and A. Vignoles (eds.) *What's the Good of Education? The Economics of Education in the UK*. Princeton: Princeton University Press.
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.
- Johnson, D. R. & Young, R. (2011). Toward Best Practices in Analysing Datasets with Missing Data: Comparisons and Recommendations, *Journal of Marriage and Family*, 73, 926-945.
- Kamphaus, R. W. (1993). *Clinical Assessment of Children's Intelligence*. Boston, MA: Allyn & Bacon.
- Kaufman, A. S. (1994). *Intelligent Testing with the WISC-III*. New York: John Wiley & Sons.
- Kuha, J. (2013). *Trends in Intergenerational Class Mobility in Britain: New Findings From the Analysis of Birth Cohort Data. A Note on the Multiple Imputation of Missing Data*. Oxford: University of Oxford.
- Marks, G. N. (2014). *Education, Social Background and Cognitive Ability: The Decline of the Social*. London: Routledge.
- Office for National Statistics (ONS) (2005). *The National Statistics Socio-economic Classification: User Manual*. London: Palgrave Macmillan.
- Office for National Statistics (ONS) (2012). 'Quality Adjustment for Public Service Education: Triangulation', 29 March

Prifitera, A., Saklofske, D. H. & Weiss, L. G. (2008). *WISC-IV: Clinical Assessment and Intervention*. San Diego, CA: Elsevier.

Roid, G. H., Prifitera, A. & Weiss, L. G. (1993). 'Replication of the WISC-III Factor Structure in an Independent Sample', in B. A. Bracken and R. S. McCallum (eds) *Journal of Psychoeducational Assessment Monograph Series, Advances in Psychoeducational Assessment: Wechsler Intelligence Scale for Children – Third Edition*. Garmantown, TN: Psychoeducational Corporation (pp 6-21).

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Schoon, I. (2008). A Transgenerational Model of Status Attainment: The Potential Mediating Role of School Motivation and Education, *National Institute Economic Review*, 205, 72-82.

Schoon, I. (2010). Childhood Cognitive Ability and Adult Academic Attainment: Evidence from Three British Cohort Studies, *Longitudinal and Life Course Studies*, 1, 241-258.

Spearman, C. (1904). "General Intelligence", Objectively Determined and Measured, *American Journal of Psychology*, 15, 201-293.

von Hippel, P. & Lynch, J. (2013). *Efficiency Gains from Using Auxiliary Variables in Imputation*. Cornell University Library.

Washbrook, E., Gregg, P. & Propper, C. (2014). A Decomposition Analysis of the Relationship between Parental Income and Multiple Child Outcomes, *Journal of the Royal Statistical Society: Series A – Statistics in Society*, 177, 757-782.