DISCUSSION PAPER SERIES

IZA DP No. 10286

Inequalities in Educational Outcomes: How Important Is the Family?

Julia Bredtmann
Nina Smith

October 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Inequalities in Educational Outcomes: How Important Is the Family?

Julia Bredtmann
RWI and IZA

Nina Smith
Aarhus University,
CESifo, IZA and RWI Research Network

Discussion Paper No. 10286
October 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 10286
October 2016

ABSTRACT

Inequalities in Educational Outcomes: How Important Is the Family?*

In this paper, we investigate sibling correlations in educational outcomes, which serve as a broad measure of the importance of family and community background. Making use of rich longitudinal survey and register data for Denmark, our main aim is to identify the parental background characteristics that are able to explain the resemblance in educational outcomes among siblings. We find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in educational achievement can be explained by family and community background. Our results further reveal that parents' socio-economic background can account for a large part of the sibling correlation. Other family characteristics such as family structure, the incidence of social problems, and parents' educational preferences also play a role, though these factors only contribute to explaining sibling similarities at lower levels of the educational distribution.

JEL Classification: I21, I24, J13

Keywords: intergenerational mobility, sibling correlations, education

Corresponding author:

Julia Bredtmann
RWI – Leibniz Institute for Economic Research
Hohenzollernstr. 1-3
45128 Essen
Germany
E-mail: julia.bredtmann@rwi-essen.de

* The authors are grateful to Anders Björklund, Sebastian Otten, Daniel Schnitzlein and participants at the Economics Seminar at the Ruhr University Bochum, the 6th International Workshop on Applied Economics of Education, the 29th ESPE conference, the 30th Congress of the EEA and the VfS Annual Conference 2015 for helpful comments and suggestions. All remaining errors are our own. This work was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD).

1 Introduction

There is a general interest in society to understand the importance of family background for individual achievement. In particular, social scientists have long been interested in exploring the intergenerational relationship between parents' and offspring's outcomes, such as their educational attainment or income. These studies are motivated by the aim to assess the degree of equality of opportunity in a society. Family background, broadly defined, represents circumstances that members of the offspring generation cannot be held accountable for, hence a strong dependence of individual outcomes on family background implies low equality of opportunity (cf. Roemer, 1998).

Though the concept of intergenerational mobility is certainly a meaningful one, the major limitation of traditional parent-offspring associations is that they are based on one single characteristic of the family. However, family background has an impact on children in many ways that cannot be picked up by one single variable.[1] An alternative approach to measure the importance of family background is to investigate the sibling correlation in economic outcomes. A sibling correlation can be interpreted as the fraction of the total variation in an outcome that can be attributed to factors shared by siblings. As such, sibling correlations provide a broad measure of the overall importance of family and community background.

Estimates of sibling correlations in educational outcomes, which are the focus of this paper, have centered around 0.5 to 0.6 for the US and 0.4 for Norway and Sweden. This suggests that even in the Nordic countries, which are characterized by an extensive welfare state and a long history of offering free post-secondary and higher education, 40 percent of the variation in educational outcomes can be attributed to family background. In our study, we focus on Denmark, a country that has been shown to rank at the top of the educational mobility scale.[2,3] Basically, we are interested in whether in a high-mobility country such as Denmark, inequalities in educational outcomes still exist.

Although sibling correlations give us an estimate of *how much* of the variation in educational outcomes can be attributed to family and community background, they do not tell us anything about *which* background characteristics matter for children's educational achievement. Björklund and Jäntti (2012) compare the sibling correlation in years of schooling with the respective intergenerational correlation between children's and parents' education and find that siblings share much more than their parents' education. Hence, if parental education is not the main driver of educational inequalities, what exactly is it

[1] See Björklund and Jäntti (2012) for a more extensive discussion of the limitations of traditional analysis of intergenerational mobility.

[2] In a cross-country comparison of the intergenerational correlation in years of schooling across 42 nations, among them 13 Western countries, Hertz *et al.* (2007) find Denmark to possess the highest level of intergenerational educational mobility among the Western countries, and one of the highest levels across the world.

[3] Basic information on the Danish institutional setting and educational system is given in Appendix A.

that makes siblings similar in terms of their educational achievement?

One hypothesis is that because most siblings grow up in the same neighborhood, this could explain part of the sibling similarity. However, recent studies for the US, the UK, and Sweden suggest that neighborhood characteristics are of minor relevance in explaining the sibling resemblance in educational outcomes. Hence, there must be something within the family that accounts for the relatively high sibling correlation in educational achievement. Obviously, parents influence their children via several channels beyond parental education: investments in their children's education, transmission of cultural values, attitudes, or social skills, and genetic endowments are all possible candidates. Moreover, in addition to the investment decisions and endowments of the parents, family members beyond the parental generation, such as grandparents, may influence the economic position of the child generation. From an equality-of-opportunity perspective, it is crucial to understand what it is that is so important about family background. A major aim of our paper is therefore to shed light on which family background characteristics are able to explain the sibling resemblance in educational outcomes.

Our contributions to the literature are manifold: First, we provide first evidence on sibling similarities in educational outcomes for Denmark, adding upon previous literature for the US, the UK, and other Scandinavian countries. Second, we are the first to investigate whether family background is more important for obtaining an upper secondary educational degree or a tertiary educational degree, gaining insights into whether educational inequality increases or decreases at higher stages of the educational system. Lastly, we are the first to decompose the sibling correlation in educational outcomes in factors attributable to family and community characteristics, thereby considering a wide range of background characteristics, including parents' socio-economic status, cognitive skills and attitudes, as well as the role of grandparents and the neighborhood.

For a sample of children born between 1968 and 1984 in Denmark, we find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in these outcomes can be explained by family and community background. For both brothers and sisters, family background is found to be more important for obtaining a tertiary educational degree than for obtaining an upper secondary degree, which suggests that educational inequality is higher at the top of the educational distribution. A decomposition of the sibling correlation reveals that parents' socio-economic status can explain up to 44 percent of the sibling correlation in educational outcomes. On top of that, only a few other family characteristics, such as the structure of the family, the incidence of social problems as well as parents' educational preferences, play a role, though these factors only contribute to explaining sibling similarities in the completion of upper secondary education.

The outline of the paper is as follows. In Section 2, we summarize previous literature on the role of family background in economic outcomes. In Section 3, we explain our

econometric approach to estimate and decompose the sibling correlations and in Section 4, we describe the data and main variables used in our empirical analysis. The results of our analysis are presented in Section 5 and Section 6 concludes.

2 Literature

A long literature in sociology and economics has aimed to estimate the importance of family background for children's future economic success. Most of this research has focused on the intergenerational relationship between parents' and offspring's outcomes.[4] Beginning with the study of Corcoran *et al.* (1976), researchers started to examine the sibling correlation as an alternative approach to measuring the importance of family background.

While intergenerational correlations in economic outcomes measure the relationship between parents' and offspring's economic success based on one single characteristic (e.g., years of education or earnings), sibling correlations in such outcomes provide a much broader measure of the role of family background for these outcomes. Measures of sibling similarity take into account not only the influence of the observed parental resource used in the intergenerational mobility analysis, but also all other unobserved factors that are shared by siblings and uncorrelated with the parental resource. Traditional studies of intergenerational associations in economic outcomes are therefore likely to study only "the tip of the iceberg" (Björklund and Jäntti, 2012, p. 471).

The majority of studies investigating sibling correlations in economic outcomes focus on investigations of sibling (or brother) correlations in permanent earnings or income. For the US, Solon *et al.* (1991), Levine and Mazumder (2007), and Mazumder (2008) find brother correlations in permanent earnings of about 0.45 to 0.50. Results for other countries as well as cross-country comparisons of sibling correlations in earnings (see, e.g., Björklund *et al.*, 2002; Schnitzlein, 2014) reveal that these estimates are of about the same size in Germany, while they are much lower in the Scandinavian countries. For Denmark, Schnitzlein (2014) estimates the sibling correlation in permanent earnings to be around 20 percent for both brothers and sisters, which is comparable with previous estimates for Finland (Österbacka, 2001; Björklund *et al.*, 2002), Norway (Björklund *et al.*, 2002) and Sweden (Björklund *et al.*, 2002, 2010; Björklund and Jäntti, 2012).

With respect to years of schooling and other educational outcomes, sibling correlations are usually found to be higher than the respective correlations in income or earnings.[5] For the US, Solon *et al.* (2000) and Mazumder (2008) estimate the sibling correlation in years of education to lie in the range of 0.5 to 0.6, suggesting that more than half of the variation in educational attainment in the US can be explained by family and community

[4]See Solon (1999) for a review of the earlier and Black and Devereux (2011) for a review of the more recent literature on intergenerational mobility.

[5]For an overview of estimates of sibling correlations in years of education across countries, see Björklund and Salvanes (2011).

factors. Looking at test scores, Mazumder (2008, 2011) and Nicoletti and Rabe (2013) find similar results for the US and the UK, respectively. Slightly smaller sibling correlations in years of education are found for the Netherlands and West Germany (Sieben *et al.*, 2001). Again, economic inequality is lower in the Scandinavian countries: For Sweden, Björklund and Jäntti (2012) find an overall sibling correlation in years of schooling of 0.44, while the correlation is slightly higher for brothers (0.46) than for sisters (0.40). Raaum *et al.* (2006) and Lindahl (2011) obtain similar results using data for Norway and Sweden, respectively.

Only a few studies have tried to gain insights into which family and community factors drive the sibling correlation in economic outcomes. A part of the literature compares the sibling correlation in economic outcomes with the respective correlation in this outcome among neighboring children in order to impose a lower bound on the role of family background as opposed to neighbor and community effects for children's outcomes.[6] In general, these studies find a small role for neighborhoods in explaining sibling correlations in educational or economic outcomes. For instance, Lindahl (2011) finds brother and sister correlations in years of education of about 0.40, while the respective neighbor correlations are much smaller: 0.02 for males and 0.01 for females when basic family background characteristics (parental income and education) are accounted for.

Mazumder (2008) is the first to systematically decompose sibling correlations in economic outcomes into factors attributable to siblings' human capital, physical characteristics, socially deviant behaviors and psychological characteristics, such as self esteem. He finds that human capital can explain 50 percent or more of the brother correlation in wages and earnings, while non-cognitive measures such as deviant behavior and psychological characteristics can account for around 20 percent of these correlations. While Mazumder (2008) is mainly interested in identifying the underlying channels through which family and community affect children's future economic outcomes, Björklund *et al.* (2010) employ Mazumder's decomposition approach to investigate which specific characteristics of the parents are important for sibling similarities in long-run income. Using data on a sample of children born in 1953 who lived in the Stockholm metropolitan area in 1963, the authors find that parents' socio-economic status, as measured by parental education, income, as well as father's occupation, can only account for 13 percent (sisters) and 28 percent (brothers) of the raw sibling correlation in long-run income.[7] They further show that the explanatory power of the family characteristics rises to 58 percent for sisters and 71 percent for brothers when indicators of parents' involvement in schoolwork and parental attitudes are added, suggesting that parental characteristics beyond parents' socio-economic status play a role for sibling similarities in long-run income.

In this paper, we contribute to the above literature in several ways: First, we add upon

[6]See, amongst others, Solon *et al.* (2000) and Page and Solon (2003a,b) for the US, Nicoletti and Rabe (2013) for the UK, Raaum *et al.* (2006) for Norway, and Lindahl (2011) for Sweden.

[7]The raw sibling correlation in long-run income was estimated to be 0.23 for sisters and 0.25 for brothers.

previous literature for the US and some European countries and provide first evidence on sibling similarities in educational outcomes for Denmark. In doing so, we go beyond the traditional analysis of years of schooling as an outcome variable, but explicitly investigate whether the role of family background varies over different stages of the educational system. Lastly, we are the first to apply a decomposition analysis as proposed by Mazumder (2008) to decompose the sibling correlation in educational outcomes in factors attributable to family and community characteristics. By making use of a combination of rich Danish survey and register data, we are able to consider a wide range of background characteristics, including parents' socio-economic status, cognitive skills and attitudes, as well as the role of grandparents and the neighborhood.

3 Empirical Model and Estimation

The following statistical framework based on Solon *et al.* (1991) is used to measure the sibling correlation in educational outcomes. Each educational outcome (e.g., years of education) is denoted by y_{ij}, where j indexes siblings and i indexes families. The model for each outcome is then:

$$y_{ij} = \mu + \epsilon_{ij}, \text{ with } \epsilon_{ij} = a_i + b_{ij}, \tag{1}$$

where μ is the population mean and ϵ_{ij} is the residual. The latter can be decomposed into a permanent component common to all siblings in the family, a_i, and a permanent component that is individual-specific, b_{ij}, which captures individual deviations from the family component. Both a_i and b_{ij} are treated as random effects that are assumed to be independent of each other. The variance of y_{ij} is then simply:

$$\sigma_\epsilon^2 = \sigma_a^2 + \sigma_b^2. \tag{2}$$

The first term, σ_a^2, captures the variance in educational outcomes that is due to differences between families, whereas the second term, σ_b^2, captures the variance in educational outcomes within families. These two components are then used to calculate the correlation in permanent outcomes between siblings, ρ:

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{3}$$

This is also equivalent to the fraction of the overall variance in educational outcomes that is due to shared family and community background.

A sibling correlation can thus be thought of as an omnibus measure of the importance of family background and community effects. It includes the variance of anything shared by siblings, such as (observed and unobserved) parental resources and influences, e.g., parents'

education and income, their cognitive and non-cognitive skills as well as their preferences and aspirations. Moreover, it captures things not directly related to the parents, such as school and neighborhood effects as well as the influence of other family members, e.g., the siblings' grandparents. However, there are also factors related to family and community background that are not captured by the sibling correlation, such as genetic traits not shared by siblings, differential treatment of siblings, and changes across time in the family, neighborhoods and schools. Therefore, the sibling correlation is a lower-bound measure of the importance of such factors.

Following, amongst others, Mazumder (2008, 2011), Björklund *et al.* (2010), and Lindahl (2011), the variance components that are needed to calculate the sibling correlation are estimated using restricted maximum likelihood (REML).[8] The standard errors of the sibling correlations are calculated using the delta method.

To understand how different observable characteristics (e.g., parental education or income) influence the sibling correlation in educational outcomes, we follow the method developed by Mazumder (2008) and augment equation (1) with the vector X_{ij}, which contains different variables depending on the specification. These variables are treated as fixed effects in the REML framework and should reduce the residual variation in the outcome variable. Hence, adding the control variables X_{ij} to the model should produce lower estimates of the family component (σ_a^{2*}) and the sibling correlation (ρ^*) than what was found without their inclusion. The relative difference between the two sibling correlations, $\Delta^* = (\rho - \rho^*)/\rho$, can then be interpreted as an estimate of the fraction of the overall sibling correlation that can be attributed to the specific factors in question. This provides an upper-bound estimate of the causal effect because it includes all omitted factors that are also correlated with the included fixed effects. For example, the reduction in ρ due to the inclusion of parents' education would be comprised of both the direct effect of parents' education on children's education as well as any omitted factors that are correlated with parents' education and influence children's educational outcomes (e.g., parents' cognitive skills or preferences). Implementing this approach for a wide variety of possible explanatory variables, either by including them one at a time or by including them simultaneously, should tell us something about which family background characteristics are critical to explaining the sibling correlation in educational outcomes.

Following Björklund *et al.* (2010), we start with adding basic individual and parental characteristics (i.e., parental education, occupation, and income) to our model to see how

[8]REML has been shown to be superior to other estimation methods (as, e.g., ANOVA formulas) when the data are unbalanced, which is the case in our study because of varying family sizes. A drawback of using REML is that the error components a and b must be assumed to be normally distributed. For years of education this may be less problematic, but for our binary outcome variables the normality assumption may be more suspect. We therefore checked the robustness of our results by using ANOVA formulas instead of REML to calculate the error components, which did not change our results substantially. This is in accordance with the results of Mazumder (2008, 2011), Lindahl (2011), and Nicoletti and Rabe (2013), who find that the results based on REML are similar to those of other estimation methods.

much standard measures of parental socio-economic background can add to explaining the sibling correlation in educational outcomes. Our main question, though, is which family characteristics beyond parents' socio-economic status are able to explain the sibling resemblance in educational outcomes. To answer this question, we then take the sibling correlation obtained from the model controlling for parents' socio-economic background, ρ^*, as our new baseline correlation and subsequently add new variables to the vector X_{ij}. The relative difference between this new sibling correlation ρ^{**} and ρ^* then gives us an estimate of the additional contribution of these family characteristics to the sibling correlation once parents' socio-economic status is already controlled for.[9] Such an analysis is not only interesting in itself, but also reduces the problem of unobserved heterogeneity accruing from the fact that many family background characteristics (e.g., parents' cognitive skills) might be highly correlated with parents' socio-economic background. Therefore, they might mainly capture the indirect effect of parents' economic status on children's outcomes.[10] Nonetheless, we can not rule out that unobserved heterogeneity is still a problem in our analysis. The obtained estimates should therefore still be interpreted as upper-bound measures of the importance of the respective family characteristics for the sibling resemblance in educational outcomes.

4 Data and Variables

4.1 Data Sources and Sample

Our basic data source is the Danish Longitudinal Survey of Youth (DLSY)[11], which is augmented with data from the Danish registers. The DLSY is an ongoing longitudinal study of a nationally representative sample of 3,151 Danish respondents who were born in or around 1954. The main respondents were first interviewed in 1968 when they were around 14 years old and attended 7th grade of elementary school. The purpose of the 1968 DLSY survey was to analyze the determinants and consequences of educational achievement and attainment. For this purpose, the class teacher in the respondent's school class and one of the respondent's parents were also interviewed in 1968 and 1969, respectively. During the first interview, the DLSY respondents further took part in a 3-dimensional (verbal, spatial, inductive) intelligence test. The main DLSY respondents have since been followed and interviewed in 1970, 1971, 1973, 1976, 1992, 2001, and finally

[9] The additional contribution is calculated as $\left(\frac{\rho^* - \rho^{**}}{\rho^*}\right) \times (1 - \Delta^*)$.

[10] Our approach thus differs from Björklund *et al.* (2010) in that we explicitly look at the additional contribution of non-economic aspects of family background, while Björklund *et al.* (2010) are mostly interested in the overall contribution of economic and non-economic factors to the sibling correlation.

[11] We use the "Cumulative 1968–2004 File (version 2)" of the DLSY data. For a technical report of these data, see Jæger (2015).

in 2004 when they were around 50 years old.[12]

The DLSY respondents constitute our index generation, which is the second generation included in the DLSY data. We then extend the survey data in several ways. By making use of parental identifiers in the Danish register data, we first identify all biological children of the DLSY respondents (born between 1968 and 2012), which constitute our sibling sample.[13] In doing so, our sibling sample includes both multiples and singletons, the latter being useful for calculating the individual error component.[14] We then identify the second biological parent of these children, i.e., the (former) spouse of the DLSY respondent. Following most of the previous literature (e.g., Björklund *et al.*, 2009; Björklund and Jäntti, 2012; Mazumder, 2008), we then restrict our sample to siblings having the same biological mother and father. In looking at children's completed education, we further have to restrict the sample to those individuals who are old enough to have completed their educational track at the end of our observation window. Hence, we restrict the sample to individuals aged 28 and older in 2012, the age at which the vast majority of individuals in Denmark has entered the labor market. This leaves us with a final sample of 3,087 children, born between 1968 and 1984, of 1,934 parents.

It is important to note that while our index generation is nationally representative of all 7th grade pupils in 1968 (and therefore roughly representative of the 1954 birth cohort), our sample of children of these parents is not representative of the Danish population. In particular, the fact that our sibling sample consists of children born to members of the 1954 birth cohort creates a potential selectivity with respect to the parents' age at birth. By restricting our sample to children born before 1984, we have to exclude all children born to the initial DLSY respondents after the age of 30. This has two main consequences: First, our sample of children is likely to be negatively selected among all children of the respective birth cohort, as we observe children born to individuals who became parents relatively early in life. Second, first- and second-born children should be over-represented in our sample, while children of higher birth order are more likely to be excluded from our sample due to the described age restriction. While it is hard to tell how this potential selectivity may affect our estimation results, a comparison of our sibling sample to a representative sample of children that are not born to a specific birth cohort

[12]Over the 36 years of data collection, the response rates in the DLSY have consistently remained very high. In the latest 2004 survey around 76 percent of the original 3,151 sample members were successfully interviewed. The drop in response rates over the period is partially due to respondents moving out of the country or dying. However, since the DLSY data can be linked with register data from Statistics Denmark, a considerable amount of information exists for all 3,151 original sample members even though they have dropped out of the DLSY survey.

[13]We do not impose any restriction on the age difference between the siblings. However, restricting the sample to children born to the initial DLSY respondents before the age of 30 (see discussion below) actually imposes a natural bound to the spacing of the siblings. In fact, the maximum age difference between the siblings in our sample is 12 years. Restricting the maximum age difference to 8 years does not change our results significantly.

[14]By including singletons in the analysis, we follow Mazumder (2008, 2011). Mazumder (2008) also shows that including singletons has little effect on the estimated sibling correlations.

reveals that selectivity is not a major concern in our analysis (see Appendix B).

4.2 Measures of Educational Attainment and Family Background

To measure children's educational success, we consider three different outcome variables obtained from the register of the level of education maintained by Statistics Denmark. Our first outcome measure is the years of schooling of the highest completed education, which range from 9 years, the compulsory schooling grade, to 21 years for individuals holding a PhD degree. In order to be able to answer whether family background is more important at the bottom or at the top of the educational ladder, we define two further outcomes: (i) a variable that takes value 1 if the individual has completed upper secondary education and (ii) a variable that takes value 1 if the individual has completed tertiary education. All outcome measures are observed in 2012, the latest year of observation in the educational registers.

Our basic control variables at both the sibling and the parental level come from the Danish population and employment registers. In 1980, Denmark was the first country to conduct a totally register-based census, hence most variables are available from 1980 onwards (i.e., till 2012, the year of observation of our outcome variables). At the sibling level, we control for gender, age (and its square) as well as birth order, which has been shown to be highly relevant for children's educational outcomes. (e.g., Black *et al.*, 2005; Björklund and Jäntti, 2012)

At the parental level, our main control variables are measures of parents' socio-economic background, i.e., both parents' education, occupation, and income. Similarly to our outcome variable, parents' education is defined as mother's and father's completed years of education. Our measure of parents' occupation is a mixture of their labor force status and their occupational status over the period 1980 to 2012. For each year within this period, we observe parents' main economic status distinguishing between the self-employed, white-collar workers, blue-collar workers, other workers (not specified), the unemployed, and those out of the labor force. The predominant status over the 32 year period is then taken as our measure of parents' occupation.[15] Lastly, we control for both parents' logarithm of average income over the period 1980 to 2012.[16]

The variables described above serve as our main control variables. One aim of this paper, however, is to investigate which family background characteristics beyond parents' socio-economic status are crucial for children's educational success and are thus able to explain the sibling resemblance in educational outcomes. In order to answer this question,

[15] We also tried different definitions of our occupation variable. For instance, we controlled for mother's and father's share of years out of the labor market. This does not alter our results substantially.

[16] Income is defined as the sum of earned income, transfer income, property income (excl. imputed rent of owner-occupiers) and other non-classifiable income attributable directly to the individual, and before deduction of labor market contributions and special pension contributions. Income is measured in 2000 prices.

we make use of the rich information on family background available in both the register and the survey data by defining a large set of further control variables which are potentially relevant for children's educational outcomes.[17]

As argued by Björklund *et al.* (2010), both the structure of the family and the incidence of social or health problems should be strong predictors of children's later-life outcomes. We measure *family structure* by using information on mother's age at first birth, the number of siblings and whether the child lived with both parents until the age of 16, all of which is obtained from the registers. *Social problems* is measured by several different variables. These include indicators for whether the mother or the father died before age 16, for whether the mother or the father was ever convicted of a crime, and if so, whether this was followed by imprisonment, as well as the share of days the mother and father received any social assistance between 1984 and 2007, the period for which this information is available in the registers. Lastly, we use information on the respondent's self-rated health status from the 1992 wave of the DLSY.

Another important determinant of children's skill formation and thus their educational success is the parental stock of cognitive and non-cognitive skills, which influences children's educational outcomes both directly, through the genetic endowment transmitted from the parents to the children, and indirectly, through parental investments in children's human capital (Becker and Tomes, 1979, 1986). As a measure of parents' *cognitive skills*, we use the scores of the 3-dimensional (verbal, spatial, inductive) intelligence test the DLSY respondents took part in during the first interview, i.e., when they were around 14 years old. In addition to cognitive skills, economic research emphasizes that non-cognitive skills or personality traits are important determinants of individual labor market outcomes (e.g., Heckman *et al.*, 2006), and that these traits are intergenerationally transmitted from the parents to their children (e.g., Anger and Schnitzlein, 2016). While the DLSY data do not contain any direct measures of non-cognitive skills, such as measures of the "Big Five" or of locus of control, we use them to create a measure of parents' *ability to plan for the future*. This is obtained from the respondents' statements to different questions, such as "No reason to think too much about the future" or "Better to save money for the future", which they were asked at age 14, 22 and 38.

Children's human capital is further influenced by different types of family values, i.e., attitudes and preferences, which are transmitted from parents to their children both genetically (cf. Cesarini *et al.*, 2009a,b) and through a process of learning. The most important family values that are relevant in our context are parents' *educational preferences*, which might have explanatory power for children's educational outcomes even beyond the effect of parents' education and income. We proxy the educational preferences of the parents by different questions asked in the DLSY survey, among them questions on

[17]Descriptive statistics of our outcome and basic control variables are shown in Table B1. Descriptive statistics of further family background characteristics are shown in Tables B3 to B5.

whether the respondents like school at age 14, on their preferred years of compulsory schooling at age 19 as well as various questions regarding the respondents' evaluation of the importance of schooling and education asked at different ages.[18]

While there are various ways through which parental characteristics can influence children's educational outcomes, more distant family members than the parents might also have an influence on children's human capital. In particular, grandparents often take an active part in the upbringing of their grandchildren and children might receive extra benefits from having well-positioned grandparents, irrespective of where their own parents stand in social hierarchies (Mare, 2011). We aim to assess the degree to which grandparents account for the sibling resemblance in educational outcomes by including information on *grandparents' socio-economic status*. In our context, the most relevant characteristic of the grandparents is their educational attainment, information on which is obtained from both the register and the survey data.[19] In the 1969 survey, the parents of the DLSY respondents were further asked about their main current occupation and their taxable annual income in the two years preceding the survey. This information on the education, occupation, and income of the first generation represents our measure of grandparents' socio-economic status.

5 Results

5.1 Raw Sibling Correlations in Educational Outcomes

The basic estimates of sibling correlations in our three educational outcomes are shown in Figure 1.[20] Starting with the outcomes for mixed sexes, we find a sibling correlation of 0.33 for our years of education variable, suggesting that about a third of the variation in years of education can be explained by family and community background. This estimate is somewhat lower than previous estimates for Norway and Sweden, centering around 0.4 (cf. Raaum *et al.*, 2006; Lindahl, 2011; Björklund and Jäntti, 2012). This supports the finding of Hertz *et al.* (2007), who – based on intergenerational correlations in years of

[18] For these variables as well as for the various questions concerning parents' ability to plan for the future, we use principal component analyses to convert the information included in these possibly correlated variables into a set of values of linearly uncorrelated variables, so-called principal components. Instead of the actual variables, these principal components are then included in the regressions in order to avoid multicollinearity problems.

[19] As in the Danish registers parental identifiers are only complete for children born in or after 1960, we are only able to identify around 40% of the grandparents in the register data. For those grandparents whose registry information is missing, we use the survey information from 1968 and 1969. In the first DLSY survey, both the respondents and the parents themselves were asked about their highest level of schooling and their post-school education. Though this information is probably of lower quality than the register information, we can use it to obtain a measure of grandparents' years of education for the majority of siblings in our data. The conversion is done by imputing years of education by regressing the years of education variable in 2012 on the survey indicators for 1968 and 1969 using all individuals for whom educational information is available in both data sources.

[20] The estimated family and individual components of the respective regressions are shown in Table C1.

education – find Denmark to possess the highest level of intergenerational educational mobility among the Western countries, including Finland, Norway and Sweden.

Considering only same-sex siblings, we find a correlation in years of schooling of 31 percent for brothers and 39 percent for sisters, though the difference between the estimates is not statistically significant. This is in line with the results of Lindahl (2011) and Björklund and Jäntti (2012), who also find very small gender differences in sibling correlations in years of education for Sweden. Turning to our other outcome variables, we find a sibling correlation of 0.15 for the completion of upper secondary education and a correlation of 0.30 for the completion of tertiary education, suggesting that family background is twice as important for obtaining a higher educational degree than for obtaining a basic one. Both estimates are slightly higher for same-sex siblings, but of about a similar size for brothers and for sisters.

Summing up, we find sibling correlations in educational outcomes for Denmark that lie in the range of 0.15 to 0.39. Though these estimates are lower than comparable estimates for other Scandinavian countries and especially for the US (cf. Mazumder, 2008), they are still of a considerable magnitude. This is especially true if one considers that sibling correlations are lower-bound estimates of the overall importance of family and community background, because there are also factors attributable to the family that are not shared by siblings, such as differences in genes or a differential treatment of the siblings. While the above results are raw estimates that do not account for any individual or family background characteristics whatsoever, we now start to add control variables to our specification in order to explain which family characteristics make siblings similar in terms of their educational outcomes.

5.2 The Contribution of Parents' Socio-economic Status

Table 1 shows our estimates of the sibling correlations for mixed siblings when basic individual characteristics as well as indicators for mothers' and fathers' socio-economic background are (subsequently) added. For each of the three outcomes considered, we report both the estimated sibling correlation (column 1) and the percentage decrease in the sibling correlation due to adding the respective covariates (column 2). The latter statistic can be interpreted as an upper-bound measure of the contribution of the considered background characteristics to the sibling correlation in educational outcomes.

The second row displays the results when only some basic individual characteristics of the siblings (age, gender, and birth order) are controlled for. The estimated sibling correlations are somewhat smaller than the raw correlations (row 1), but overall adding individual controls does not alter the results substantially. This is not surprising, given that controlling for individual characteristics should mainly sob up some of the residual variation previously captured by the individual component rather than by the family

component.[21]

In the next step, we separately add different indicators for parents' socio-economic status, starting with controls for mother's education, occupation, and income (rows 3 to 5). Considering siblings' years of education, mother's education and occupation each can explain around 15 percent of the sibling correlation, while mother's income seems to be of less relevance (8 percent). The contribution of mother's background characteristics to the sibling correlation, however, varies over the stages of the educational system. It turns out that mother's occupation is able to explain a large part of the sibling correlation in completing upper secondary education (23 percent), while mother's education is relatively more important for obtaining a tertiary educational degree (15 percent). A look at the full estimation results displayed in Table C2 reveals that with respect to mother's occupation, having an unemployed or non-participating mother seems to be most harmful for completing upper secondary education. Overall, mother's socio-economic status is able to explain 20 to 25 percent of the sibling similarity in educational outcomes (row 6).

Turning to the estimation results for father's socio-economic status (rows 7 to 10), we find similar results as for mother's socio-economic status. Among the background characteristics considered, father's education and occupation seem to be most relevant for explaining the sibling resemblance in educational outcomes, while the latter is again most important for obtaining an upper secondary degree. Father's income, on the other hand, is not able to explain more than 11 percent of the sibling correlation. Overall, father's socio-economic status can account for up to 27 percent of the sibling correlation in educational outcomes (row 10). Hence, mother's and father's socio-economic backgrounds seem to be about equally relevant for children's educational outcomes.[22]

In the last step, we add mother's and father's background characteristics simultaneously. First, we only include mother's and father's years of education (row 11), i.e., we look at the intergenerational correlation between parents' and children's education. Our results reveal that these factors alone are only able to explain between 16 and 24 percent of the sibling correlation in educational outcomes. Hence, there seems to be much more than parental education that is responsible for inequalities in educational outcomes. This supports our notion (and that of other researchers) that intergenerational relationships between parents' and children's outcomes are likely to only capture the "tip of the iceberg" (Björklund and Jäntti, 2012).

In the last row of Table 1, we control for both mother's and father's socio-economic status. The results reveal that parental socio-economic background is able to explain about a third of the sibling correlation in years of education. The explanatory power of

[21] While this is true for gender and birth order, the siblings' age is certainly a factor related to family background, as – given that our siblings represent a sample of children born to members of a specific birth cohort – it partly captures the indirect effect of parents' age at birth.

[22] Of course, both parents' education, occupation, and income are likely to be highly correlated with each other, making it difficult to ascertain the importance of one single indicator.

parents' background characteristics, however, is much higher for the completion of an upper secondary degree (44 percent) than for the completion of a tertiary degree (29 percent). This finding is in line with Landersø and Heckman (2016), who compare educational mobility between Denmark and the US and find that parental background, as measured by parents' income and wealth, is more strongly correlated with children's high school completion in Denmark than in the US, while no such difference is found for children's college attendance.

In Table 2, we have conducted the same analysis separately for brothers and sisters.[23] The results reveal that there are indeed some differences between the genders. With respect to the completion of upper secondary education, we find that mother's occupation and also her income are most relevant for explaining brother correlations in this outcome, while they are less relevant for sisters. On the other hand, father's occupation and income are more important for girls than for boys. For both genders, parents' education does hardly show any explanatory power for the sibling similarity in completing upper secondary education. Especially mother's years of education, however, are an important predictor of inequalities in the completion of tertiary education of girls. Overall, we find that parents' socio-economic status is more relevant for brothers regarding the completion of upper secondary education and more relevant for sisters regarding the completion of tertiary education.

The above results show that parental socio-economic status is a major determinant of inequalities in educational outcomes, especially of inequalities in obtaining a high school or vocational education degree. However, our results also reveal that the dominating part of the sibling correlations in these outcomes cannot be explained by parents' education, occupation, or income. Hence, there must be something more than parents' socio-economic status that drives the sibling similarity in educational outcomes.

5.3 The Contribution of Other Family Characteristics

In the next step, we therefore add further, potentially relevant family background characteristics to our model, which might be able to explain the sibling resemblance in educational outcomes.[24] As we are interested in the additional contribution of these family characteristics to the sibling correlation once parents' socio-economic status is already controlled for, the sibling correlation obtained from the model controlling for parents' socio-economic background (last row of Table 1) serves as our new baseline correlation.[25]

As is evident from Table 3, indicators of the *structure of the family*, i.e., mother's age at first birth, the number of siblings and whether the child grew up with both parents, are able

[23] Full estimation results are shown in Table C3.

[24] Note that the reduced sample size for at least part of these analyses precludes us from conducting our analysis separately for boys and girls.

[25] Table C4 shows the respective results not conditioning on parents' socio-economic background. Full estimation results are shown in Tables C5 to C10.

to explain another 8 percent of the variation in siblings' years of education. Considering the different stages of the educational system, it further turns out that family structure is able to explain a large part of the sibling correlation in completing upper secondary education (15 percent), while it is hardly relevant for obtaining a tertiary educational degree (3 percent). A similar picture emerges when including proxies for *social problems*, i.e., for whether one of the parents died or was convicted of a crime, their social assistance receipt, and their health status. Conditional on parents' socio-economic status, social problems account for about another 19 percent of the sibling correlation in completing upper secondary education, but have no explanatory power for obtaining a tertiary degree. This reveals that inequalities in the completion of a lower educational degree cannot only be attributed to parents' socio-economic status, but that the structure of the family and the incidence of social problems also play an important role.[26]

Somewhat surprisingly, our next set of family characteristics, parents' *cognitive skills*, does not matter much for sibling similarities in educational outcomes. For all outcomes considered, the explanatory power of parents' cognitive skills, as measured by the result of the 3-dimensional intelligence test the parents took part in at age 14, does not exceed 2 percent. What is the reason for this? Examining the regression results directly (Table C7) reveals that parents' scores on both the verbal and the inductive intelligence test are strong predictors of their children's educational outcomes. However, this effect almost completely vanishes once parents' socio-economic status is controlled for. Hence, parents' cognitive skills do matter for children's educational outcomes, but as their effect mainly operates through parents' socio-economic status, their additional explanatory power for the sibling correlation is negligible. A similar picture emerges when considering parents' non-cognitive skills, in our case *parent's ability to plan for the future*. While these factors do matter for children's outcomes as long as parents' socio-economic status is not controlled for (Table C8), the additional explanatory power of these variables above and beyond parents' socio-economic status is small.

A factor that does seem to matter for the sibling correlation in educational outcomes is parents' *educational preferences*. When controlling for the educational preferences of the parents, the sibling correlation in years of education decreases by 3 percent and the sibling correlation in the completion of upper secondary education decreases by 11 percent. This suggests that parents' educational preferences exert an effect on children's educational outcomes that goes above and beyond the effect of parents' education, occupation, or income. As argued by Jæger and Holm (2007), parents' tastes and preferences may affect children's educational attainment because the home environment acts as a "learning lab" in the development of children's educational preferences, their cognitive skills as well as

[26] Adding both groups of variables simultaneously, however, does not substantially increase the explanatory power, suggesting that family structure and social problems to a large extent capture the same underlying mechanisms.

their knowledge of the education system. Parents with strong preferences for education are likely to pass on to their children a preference for an academically oriented education and are more likely to possess realistic information on the strategic importance of pursuing further education, e.g., by illustrating that it is a long-term investment in educational and occupational status. This argumentation, however, is not fully supported by our results. While we find parents' educational preferences to matter for obtaining an upper secondary degree, they have no explanatory power for completing tertiary education.

Lastly, we address the aspect that more distant family members than the parents might also have an influence on children's educational outcomes and control for *grandparents' socio-economic status*, i.e., their education, occupation, and income. While grandparents' socio-economic status does not contribute to explaining sibling similarities in obtaining a tertiary educational degree, it explains another 7 percent of the sibling similarity in the completion of upper secondary education. This supports the results of a more recent literature (e.g., Lindahl *et al.*, 2015) that extends the traditional two-generation model by considering three or four generations and finds a strong persistence in economic outcomes across multiple generations.

We conduct a number of sensitivity analyses to check the robustness of our results. First, there might be an issue with using principal component analyses (PCA) to convert the information included in some of the possibly correlated variables into a set of values of linearly uncorrelated variables, so-called principal components which are then included in the regressions. This has been done for the various questions asked with respect to parents' forward looking behavior (8 questions) as well as for those concerning parents' educational preferences (6 questions). Although the Kaiser-Meyer-Olkin measure of sampling adequacy exceeds 0.6 for both cases and therefore passes the critical value of 0.5, it might still be the case that the considered variables have not enough in common to warrant a PCA. In order to rule out that our results are driven by the use of PCA, we check the robustness of our results by including the actual variables instead of the principal components in our regressions. The respective results are both qualitatively and quantitatively similar.

Second, it might be a problem that only one of the parents, either the mother or the father, has participated in the DLSY survey from which most of the information on family characteristics is drawn. This is problematic if, for example, mother's skills or preferences exert a differential impact on children's outcomes than father's skills or preferences. In order to address this issue, we interact all explanatory variables obtained from the DLSY data with an indicator variable for the gender of the survey respondent. Although there are some variables for which a gender-specific impact cannot be ruled out, the basic picture, i.e., the explanatory power of these variables for the sibling correlation in educational outcomes, remains largely the same.

Lastly, we address the issue that some of the family characteristics considered, i.e., the measures of whether the child lived with both parents until the age of 16 and whether

one of the parents died before the child turned 16, are not constant across siblings, but do in fact vary over the siblings within the family. Controlling for these variables could therefore increase the individual component and thus lower the sibling correlation in educational outcomes. We therefore checked the robustness of our results by conditioning these variables on the parents' instead of the children's age to hold them fixed among siblings of the same family. Again, the results are qualitatively robust to this alternative specification.[27]

Our results raise some important questions: First, why are many family characteristics that theoretically influence children's educational success not able to explain the sibling similarity in educational outcomes? One explanation might be measurement error. If our measures of family characteristics are very imprecisely measured, then this might explain why they do not add much to explaining the sibling similarity in educational outcomes. Though measurement error is certainly an issue here, we argue that this cannot be the whole story. As can be seen from Tables C5 to C10, most variables do have explanatory power for children's educational outcomes. However, the explanatory power of these variables is often strongly reduced once parents' socio-economic status is controlled for. This finding is also important from a methodological point of view: By obtaining information on parents' socio-economic characteristics from the register data, we are able to measure father's and mother's education, occupation and income both very precisely and over a long period of time (1980 to 2012). With these precise measures of parents' socio-economic status at hand, other family background characteristics, such as parents' cognitive or non-cognitive skills, do not seem to exert a huge additional influence on children's educational outcomes.

Second, our results reveal that, if anything, family characteristics beyond parents' socio-economic status are able to explain inequalities in reaching a lower educational qualification, but not in obtaining a tertiary degree. This is somewhat puzzling since the initial sibling correlation and thus the inequality in obtaining a tertiary degree is actually larger than the respective sibling correlation in obtaining an upper secondary degree. However, when we consider the factors that do actually matter for children's educational outcomes, such as whether the child grew up with both parents, parents' crime incidence as well as their welfare receipt, it is not surprising that such conditions are particularly crucial in determining whether a child pursues some further education beyond compulsory schooling or not, especially given that the completion rate of upper secondary education in Denmark is particularly high (86 percent in our sample).

Also, although our data contain comprehensive information on parents' preferences, aspirations, and skills, there are obviously some background characteristics that cannot be captured by our data. Most importantly, we might fail to capture some relevant parental non-cognitive skills, such as parents' (academic) "identity" or their knowledge

[27]The results of all robustness checks are available from the authors upon request.

of specific information about the tertiary education system. Non-cognitive skills are strongly intergenerationally transmitted from the parents to their children (e.g., Anger and Schnitzlein, 2016) and have been shown to be especially important for the choice of higher education (Rustichini *et al.*, 2016), which might partly explain the sizable, so far unexplained, sibling correlation in tertiary educational attainment.

5.4 The Contribution of the Neighborhood

In our search for factors that are able to explain the sibling similarity in educational outcomes, we have so far focused on background characteristics of the family, broadly defined. Another hypothesis would be that it is rather neighborhood characteristics shared by siblings than family characteristics that are able to explain inequalities in educational outcomes. In order to test this hypothesis, we follow previous literature (e.g., Solon *et al.*, 2000; Raaum *et al.*, 2006; Lindahl, 2011; Nicoletti and Rabe, 2013) and estimate correlations in educational attainment among children growing up in the same neighborhood in order to impose a lower bound on the role of family background in determining children's outcomes.

In order to measure the neighbor correlations in educational outcomes, we estimate a model similar to Eq. (1):

$$y_{cij} = \lambda + \nu_{cij}, \text{ with } \nu_{cij} = u_c + v_{cij}, \tag{4}$$

where y_{cij} is the educational outcome of individual j in family i in neighborhood c. λ is the population mean and ν_{cij} is the residual, which can be decomposed into a neighborhood random component, u_c, and an individual-specific error term, v_{cij}. The variances of the individual and the neighborhood component can then be used to calculate the neighbor correlation in educational outcomes, $\varphi = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$, which captures the share of the between-neighborhood variation of the overall variance in educational outcomes.

As outlined by Solon *et al.* (2000), a neighbor correlation represents an upper-bound measure of the importance of the neighborhood for individuals' outcomes, as it captures both the "pure" neighborhood effect as well as an indirect effect, accruing from sorting of families into neighborhoods. As the sorting effect is assumed to be positive, the neighbor correlation as estimated from Eq. (4) represents an upper-bound measure capturing the indirect family effects as well.[28]

We define an individual's neighborhood in terms of where he or she lived at age 16.

[28]Specifically, there are two different types of sorting, which are both assumed to positively contribute to the neighbor correlation: (i) sorting of similar families into the same neighborhood and (ii) sorting of advantaged families into advantaged neighborhoods. In order to tighten the bound of the neighbor correlation, some authors use measures of neighbor correlations that are adjusted for observed family characteristics. As our aim is to impose a lower bound on the role of the family in generating inequalities in educational outcomes, we estimate unadjusted neighbor correlations. For a formal derivation of the estimation of the neighborhood covariance, see Solon *et al.* (2000).

Though this might be an imperfect measure of the neighborhood, as the family may have lived elsewhere before, data limitations preclude us from following individuals further back in time.[29] We define neighborhoods at the level of the postcode area. There are 1,029 postcode areas in Denmark, which have an average population of 5,317 individuals.[30] In order to be able to compare the neighbor correlations with our previously estimated sibling correlations, we construct a neighborhood sample that resembles our sibling sample. Specifically, our neighborhood sample is comprised of all children born to the 1954 birth cohort. Moreover, we restrict the sample to individuals aged 28 to 44 in 2012, which represent the minimum and maximum ages in our sibling sample.

The estimation results are shown in Figure 2. Though the neighbor correlations are all statistically significant, they are very small in magnitude. The neighbor correlation in years of education, for instance, amounts to 0.024 for the mixed sample, while it is slightly higher for women than for men (0.026 vs. 0.022). Considering the other outcomes, the estimated neighbor correlations are even smaller. This suggests that in Denmark, less than 3 percent of the variation in educational outcomes can be explained by neighborhood effects. This result is in line with previous literature (e.g., Raaum *et al.*, 2006; Lindahl, 2011), which also finds a small role for neighborhoods in explaining the sibling resemblance in educational outcomes.

Although the amount of variation explained by neighborhood correlations should exceed the explanatory power of standard regression-based neighborhood analyses (see, Page and Solon, 2003a), we further explore an alternative approach to measuring neighborhood effects, in which we add information on local neighborhood characteristics to our sibling estimates. First, we add information on population density, average years of education, the local unemployment rate, as well as the share of immigrants in the postcode area to our model. The results reveal that local neighborhood characteristics have hardly any explanatory power for sibling similarities in educational outcomes. As the choice of neighborhood variables is of course to some extent arbitrary, we next add postcode-area fixed effects to our sibling estimates. Again, the estimated sibling correlations in educational attainment remain largely unchanged. Lastly, we follow the argument of Raaum *et al.* (2006) that it might rather be regional characteristics than local conditions that affect children's outcomes and add municipality fixed effects to our sibling estimates. In this way, we should capture all of the variation in siblings' educational outcomes that is explained by the region in which they spent their childhood. Again, our estimated sibling correlations remain largely unchanged.[31] Hence, our conclusion that neighborhoods play a

[29]However, previous research has shown that even when families move, the neighborhoods to which they move are usually similar to the ones from where they move (Kunz *et al.*, 2003). We thus assume that the neighborhood at age 16 is a good proxy for the neighborhood environment the children grew up in.

[30]The numbers refer to the year 2012. In 2007, Denmark underwent a local government reform, which involved a complete reorganization of municipalities and also led to a change in postcodes. We therefore use the postcode areas as defined after 2007, even if observing individuals before 2007.

[31]All estimation results are available from the authors upon request.

minor role in explaining sibling similarities in educational outcomes remains.

6 Conclusion

In this paper, we analyze the correlation in educational outcomes among siblings in Denmark. A sibling correlation captures everything that is shared by siblings and can thus be thought of as an omnibus measure of the importance of family and community background for individuals' outcomes. Our main contribution to previous literature is that we do not only determine the extent of the sibling similarity in educational attainment, but further aim at identifying the determinants of educational inequalities across families.

For a sample of children born to participants in a Danish long-term study, which constitutes a representative sample of all 7th graders in 1968, we find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in these outcomes can be explained by family and community background. Hence, even in a highly egalitarian country such as Denmark, inequalities in educational attainment do exist.

For both brothers and sisters, we find family background to be more important for obtaining a tertiary educational degree than for obtaining an upper secondary degree, which suggests that educational inequality is higher at the top of the educational distribution. This result points to the possible existence of a "glass-ceiling effect" for children from disadvantaged families. While the Danish welfare state, with its universal system of publicly provided daycare for pre-school children, is able to mitigate inequalities in children's skills and schooling outcomes early in life, it seems less successful in removing existing barriers to pursue tertiary education for children from disadvantaged families. If the general political aim is to reduce educational inequality, promoting higher education for less advantaged children thus seems to be of particular importance.

A decomposition of the sibling correlation reveals that parents' socio-economic status, i.e., their education, occupation and income, is the main determinant of sibling similarities in educational outcomes. Non-economic factors such as family structure, the incidence of social problems, or parents' educational preferences are able to explain an additional part of the sibling correlation. Overall, these family characteristics have much more explanatory power for sibling similarities in obtaining an upper secondary degree than for sibling similarities in obtaining a tertiary degree. Parents' cognitive skills, the socio-economic background of the grandparents as well as neighborhood characteristics, in contrast, seem to be of minor relevance for explaining inequalities in educational outcomes. Hence, a large part of the educational inequality, especially of the inequality in completing higher education, remains unexplained by observable characteristics.

What is it then that explains the remaining parts of the sibling correlations in educational outcomes? Even though we cannot provide direct evidence for this, our interpretation

is that there exist some unobservable non-cognitive skills of the parents, as for example their (academic) "identity" or their knowledge of specific information about the tertiary education system, that cannot be captured by our data. Such skills have been shown to be especially important for the choice of tertiary education (Rustichini *et al.*, 2016) and may thus explain the large sibling resemblance in higher levels of education. This interpretation is in line with the findings of Landersø and Heckman (2016), who show that both in Denmark and in the US, the relationship between parental resources and children's education is to a large degree mediated by levels of cognitive and non-cognitive skills at earlier ages, and that this relationship is more pronounced for children's college attendance than for their high-school completion. If it is rather non-cognitive skills, transmitted from the parents to their children, than parental resources that contribute to the high sibling correlation in higher levels of education, then this might explain why in Denmark, despite its generous welfare state, including free college tuition, sizeable inequalities in tertiary educational attainment still exist.

Lastly, it is important to point to some limitations of our approach. First, our results are purely descriptive and cannot be interpreted causally. It is a major challenge for future research to obtain causal inference on which family background characteristics are important for children's outcomes and to which extent these factors violate equality norms. Second, our analysis has been primarily empirical. As already noted by Björklund and Jäntti (2012), the search for factors that explain sibling similarities in economic outcomes should ideally be guided by an all-encompassing theoretical model. While the standard Becker-Tomes model (Becker and Tomes, 1979, 1986) provides a theoretical model of the causal effect of parental income on offspring's outcomes, the literature on sibling correlations would clearly benefit from a broader theoretical framework that explains how family and community factors interact to influence children's outcomes.

References

Anger, S. and Schnitzlein, D. D. (2016). Cognitive Skills, Non-Cognitive Skills, and Family Background: Evidence from Sibling Correlations. IZA Discussion Paper No. 9918.

Becker, G. S. and Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, **87** (6), 1153–1189.

— and — (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, **4** (3), 1–39.

Björklund, A., Eriksson, T., Jäntti, M., Raaum, O. and Österbacka, E. (2002). Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States. *Journal of Population Economics*, **15** (4), 757–772.

— and Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, **19** (4), 465–474.

—, Jäntti, M. and Lindquist, M. J. (2009). Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932–1968. *Journal of Public Economics*, **93** (5–6), 671–680.

—, Lindahl, L. and Lindquist, M. J. (2010). What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents. *The B.E. Journal of Economic Analysis & Policy*, **10** (1), Article 102.

— and Salvanes, K. G. (2011). Education and Family Background: Mechanisms and Policies. In E. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education, Vol. 3*, Amsterdam: Elsevier, pp. 1559–1695.

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. In D. Card and O. Ashenfelter (eds.), *Handbook of Labor Economics, Vol. 4, Part B*, Amsterdam: Elsevier, pp. 1487–1541.

—, — and Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics*, **120** (2), 669–700.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P. and Wallace, B. (2009a). Genetic Variation in Preferences for Giving and Risk Taking. *The Quarterly Journal of Economics*, **124** (2), 809–842.

—, Johannesson, M., Lichtenstein, P. and Wallace, B. (2009b). Heritability of Overconfidence. *Journal of the European Economic Association*, **7** (2-3), 617–627.

Corcoran, M., Jencks, C. and Olneck, M. (1976). The Effects of Family Background on Earnings. *The American Economic Review Papers and Proceedings*, **66** (2), 430–435.

Datta Gupta, N., Smith, N. and Verner, M. (2008). The impact of Nordic countries' family friendly policies on employment, wages, and children. *Review of Economics of the Household*, **6** (1), 65–89.

Heckman, J. J., Stixrud, J. and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, **24** (3), 411–482.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A. (2007). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy*, **7** (2), Article 10.

Jæger, M. M. (2015). Danish Longitudinal Survey of Youth (DLSY) – Cumulative 1968–2004 File. The Danish National Centre for Social Research Working Paper 02:2015.

Jæger, M. M. and Holm, A. (2007). Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? *Social Science Research*, **36** (2), 719–744.

Kunz, J., Page, M. E. and Solon, G. (2003). Are point-in-time measures of neighborhood characteristics useful proxies for children's long-run neighborhood environment? *Economics Letters*, **79** (2), 231–237.

Landersø, R. and Heckman, J. J. (2016). The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S. *Scandinavian Journal of Economics. forthcoming*.

Levine, D. I. and Mazumder, B. (2007). The Growing Importance of Family: Evidence from Brothers' Earnings. *Industrial Relations*, **46** (1), 7–21.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income – evidence from Sweden. *The Journal of Economic Inequality*, **9** (2), 207–226.

Lindahl, M., Palme, M., Sandgren Massih, S. and Sjögren, A. (2015). Long-term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations. *The Journal of Human Resources*, **50** (1), 1–33.

Mare, R. D. (2011). A Multigenerational View of Inequality. *Demography*, **48** (1), 1–23.

Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of Population Economics*, **21** (3), 685–701.

— (2011). Family and community influences on health and socioeconomic status: sibling correlations over the life course. *The B.E. Journal of Economic Analysis & Policy*, **11** (3), Article 1.

Nicoletti, C. and Rabe, B. (2013). Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter? *Economica*, **80** (318), 197–218.

OECD (2013). *Education at a Glance 2013: OECD Indicators*. Paris: OECD Publishing.

— (2014). *Revenue Statistics 2014*. Paris: OECD Publishing.

Österbacka, E. (2001). Family Background and Economic Status in Finland. *Scandinavian Journal of Economics*, **103** (3), 467–484.

Page, M. E. and Solon, G. (2003a). Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban. *Journal of Labor Economics*, **21** (4), 831–855.

— and — (2003b). Correlations between sisters and neighbouring girls in their subsequent income as adults. *Journal of Applied Econometrics*, **18** (5), 545–562.

Raaum, O., Salvanes, K. G. and Sørensen, E. Ø. (2006). The Neighbourhood is Not What it Used to be. *The Economic Journal*, **116** (508), 200–222.

Roemer, J. E. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.

Rustichini, A., Iacono, W. G. and McGue, M. (2016). The Contribution of Skills and Family Background to Educational Mobility. *Scandinavian Journal of Economics. forthcoming*.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, **27** (1), 69–89.

Sieben, I., Huinink, J. and de Graaf, P. M. (2001). Family Background and Sibling Resemblance in Educational Attainment. Trends in the Former FRG, the Former GDR, and the Netherlands. *European Sociological Review*, **17** (4), 401–430.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Vol. 3*, Amsterdam: Elsevier, pp. 1761–1800.

—, Corcoran, M., Gordon, R. and Laren, D. (1991). A Longitudinal Analysis of Sibling Correlations in Economic Status. *The Journal of Human Resources*, **26** (3), 509–534.

—, Page, M. E. and Duncan, G. J. (2000). Correlations between Neighboring Children in Their Subsequent Educational Attainment. *The Review of Economics and Statistics*, **82** (3), 383–392.

Figures

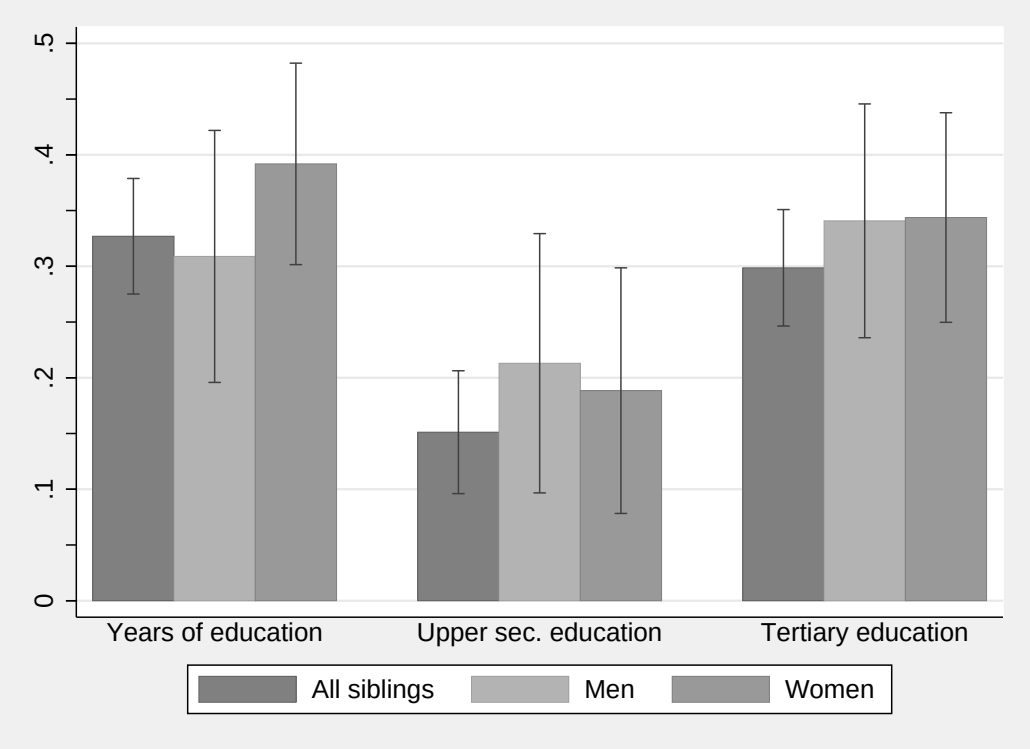

Figure 1: Raw Sibling Correlations

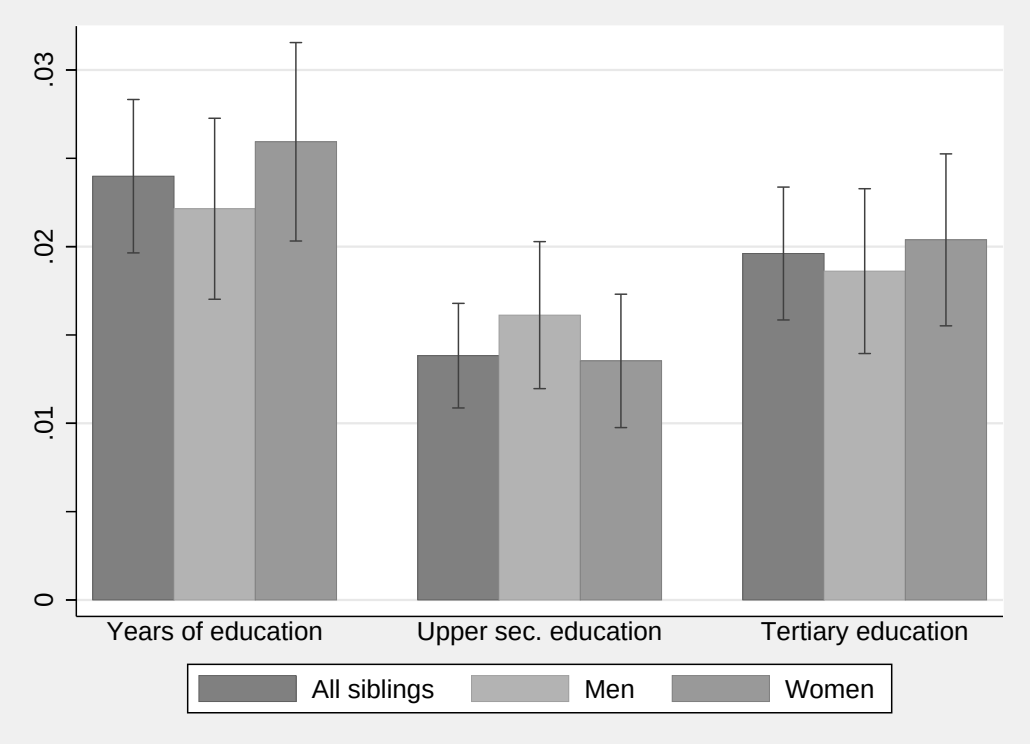

Figure 2: Raw Neighbor Correlations

Tables

Table 1: Sibling Correlations in Educational Outcomes and Parents' Socio-economic Background

	Years of education		Completed upp. sec. ed.		Completed tertiary ed.	
	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$
Raw correlations						
Sibling correlation	0.327	–	0.151	–	0.299	–
StdE	(0.026)		(0.028)		(0.027)	
Individual controls						
Sibling correlation	0.285	–	0.133	–	0.256	–
StdE	(0.028)		(0.028)		(0.028)	
Mother's education						
Sibling correlation	0.244	15%	0.121	9%	0.217	15%
StdE	(0.028)		(0.028)		(0.028)	
Mother's occupation						
Sibling correlation	0.244	15%	0.103	23%	0.220	14%
StdE	(0.028)		(0.028)		(0.028)	
Mother's income						
Sibling correlation	0.263	8%	0.122	8%	0.239	7%
StdE	(0.028)		(0.028)		(0.028)	
Mother's education, occupation, and income						
Sibling correlation	0.224	21%	0.099	25%	0.204	20%
StdE	(0.028)		(0.028)		(0.028)	
Father's education						
Sibling correlation	0.247	13%	0.119	10%	0.233	9%
StdE	(0.028)		(0.028)		(0.028)	
Father's occupation						
Sibling correlation	0.251	12%	0.106	21%	0.231	10%
StdE	(0.028)		(0.028)		(0.028)	
Father's income						
Sibling correlation	0.264	8%	0.118	11%	0.244	5%
StdE	(0.028)		(0.028)		(0.028)	
Father's education, occupation, and income						
Sibling correlation	0.223	22%	0.097	27%	0.215	16%
StdE	(0.028)		(0.028)		(0.028)	
Parents' education						
Sibling correlation	0.218	24%	0.112	16%	0.203	21%
StdE	(0.028)		(0.028)		(0.028)	
Parents' education, occupation, and income						
Sibling correlation	0.185	35%	0.075	44%	0.183	29%
StdE	(0.029)		(0.028)		(0.028)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – The individual controls (age, gender, birth order) are included in all regressions and the results displayed in the second row constitute our new baseline estimates. – Full estimation results are shown in Table C2.

Table 2: Brother and Sister Correlations in Educational Outcomes and Parents' Socio-economic Background

	Years of education				Completed upp. sec. education				Completed tertiary education			
	Brothers		Sisters		Brothers		Sisters		Brothers		Sisters	
	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$
Raw correlations												
Sibling correlation	0.309	–	0.392	–	0.213	–	0.188	–	0.341	–	0.344	–
StdE	(0.058)		(0.046)		(0.059)		(0.056)		(0.053)		(0.048)	
Individual controls												
Sibling correlation	0.269	–	0.350	–	0.214	–	0.160	–	0.290	–	0.299	–
StdE	(0.059)		(0.049)		(0.059)		(0.057)		(0.057)		(0.050)	
Mother's education												
Sibling correlation	0.259	4%	0.305	13%	0.214	0%	0.155	3%	0.273	6%	0.254	15%
StdE	(0.057)		(0.050)		(0.057)		(0.057)		(0.057)		(0.051)	
Mother's occupation												
Sibling correlation	0.244	10%	0.313	11%	0.177	18%	0.152	5%	0.269	7%	0.270	10
StdE	(0.059)		(0.050)		(0.060)		(0.056)		(0.057)		(0.051)	
Mother's income												
Sibling correlation	0.241	11%	0.329	6%	0.192	10%	0.156	2%	0.278	4%	0.280	6%
StdE	(0.061)		(0.050)		(0.060)		(0.057)		(0.058)		(0.051)	
Mother's education, occupation, and income												
Sibling correlation	0.238	12%	0.289	17%	0.179	17%	0.152	5%	0.263	9%	0.247	17%
StdE	(0.058)		(0.051)		(0.060)		(0.056)		(0.057)		(0.052)	
Father's education												
Sibling correlation	0.253	6%	0.320	8%	0.203	6%	0.160	0%	0.282	3%	0.279	7%
StdE	(0.058)		(0.050)		(0.059)		(0.056)		(0.056)		(0.051)	
Father's occupation												
Sibling correlation	0.253	6%	0.308	12%	0.195	9%	0.129	19%	0.280	3%	0.265	11%
StdE	(0.059)		(0.051)		(0.059)		(0.057)		(0.057)		(0.052)	
Father's income												
Sibling correlation	0.270	0%	0.316	10%	0.210	2%	0.145	9%	0.291	0%	0.273	9%
StdE	(0.058)		(0.050)		(0.058)		(0.057)		(0.057)		(0.051)	
Father's education, occupation, and income												
Sibling correlation	0.240	11%	0.283	19%	0.185	14%	0.128	20%	0.274	6%	0.247	17%
StdE	(0.058)		(0.052)		(0.059)		(0.057)		(0.056)		(0.052)	
Parents' education												
Sibling correlation	0.244	10%	0.290	17%	0.203	5%	0.156	2%	0.267	8%	0.245	18%
StdE	(0.056)		(0.050)		(0.058)		(0.056)		(0.056)		(0.051)	
Parents' education, occupation, and income												
Sibling correlation	0.212	21%	0.254	27%	0.153	29%	0.124	22%	0.250	14%	0.223	25%
StdE	(0.059)		(0.052)		(0.061)		(0.056)		(0.057)		(0.052)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – The individual controls (age, gender, birth order) are included in all regressions and the results displayed in the second row constitute our new baseline estimates. – Full estimation results are shown in Table C3.

Table 3: Contribution of Family Characteristics to Sibling Correlations

	Years of education		Completed upp. sec. ed.		Completed tertiary ed.	
	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$
Parents' socio-economic status						
Sibling correlation	0.185	–	0.075	–	0.183	–
StdE	(0.029)		(0.028)		(0.028)	
Family structure						
Sibling correlation	0.163	8%	0.053	15%	0.177	3%
StdE	(0.029)		(0.028)		(0.029)	
Social problems						
Sibling correlation	0.182	4%	0.075	19%	0.181	1%
StdE	(0.030)		(0.030)		(0.030)	
Parent's cognitive skills						
Sibling correlation	0.180	2%	0.071	1%	0.181	2%
StdE	(0.029)		(0.029)		(0.029)	
Parent's ability to plan for the future						
Sibling correlation	0.209	1%	0.125	4%	0.196	1%
StdE	(0.032)		(0.032)		(0.031)	
Parent's educational preferences						
Sibling correlation	0.217	3%	0.109	11%	0.206	1%
StdE	(0.033)		(0.033)		(0.032)	
Grandparents' socio-economic status						
Sibling correlation	0.176	1%	0.056	7%	0.191	0%
StdE	(0.032)		(0.032)		(0.032)	
All family characteristics						
Sibling correlation	0.191	10%	0.044	44%	0.196	4%
StdE	(0.033)		(0.035)		(0.033)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – In all models, the sibling and parental characteristics included in Table 1 are controlled for. – Note that the number of observations varies with the control variables considered. The percentage change in the sibling correlation is therefore calculated based on the baseline sibling correlation for the specific sample. – Note also that the number of observations reduces to 2,200 when all family characteristics are considered. Hence, the results have to be interpreted with some caution.

Appendix A

The Danish Institutional Setting and Educational System

As the other Scandinavian countries, Denmark started already in the 1960s to build up the "Scandinavian Welfare State" which is characterized by a large redistribution of income via a high tax pressure[32] as well as means-tested income transfers. Besides, Denmark has one of the highest coverage rates of publicly provided day care for pre-school children worldwide, and the quality of public childcare is relatively high (see Datta Gupta *et al.*, 2008). Childcare is highly subsidized by the public and for low-income families it is virtually for free. The public school system is also widely free and among the most expensive in the world (see OECD, 2013).

In 1968, when the parental generation considered in our analysis attended the 7th grade of elementary school, compulsory school in Denmark started at the age of 6 or 7 and included 7 years of compulsory schooling. In 1972, compulsory schooling was extended to 9 years by adding two extra compulsory grades, grades 8 and 9. In 2009, the former "kindergarten grade" or "grade 0" for children aged 5-6 years was made compulsory. Since the sample of children considered in this study is restricted to those born between 1968 and 1984, these children were subject to 9 years of compulsory schooling. There is no tracking in the compulsory school system in Denmark, i.e., up to the age of 15-16 years all children attend the same schools. After grades 0 to 9, the pupils have the options of 1 extra year at school (grade 10), to enroll at upper secondary education (high school or a vocational education), which typically takes 3 years, or to leave the formal educational system without any education. After upper secondary education, the students may enroll in a tertiary education at a university or a university college. A completed tertiary education takes between 2 and 6 years (for instance, 2 years for a laboratory technician, 4 years for a school teacher or nurse and 5-6 years for a master's degree at university level). Finally, on top of a university bachelor's or master's degree, the students may complete a PhD degree. This means that the total number of years of formal education for our sample of siblings may vary from 9 years to 9+3+6+3=21 years. Of course many students do not complete their formal education within the "standard" number of years, i.e., it may take more than 21 years to complete a PhD degree.

There are no fees in the Danish educational system, including all universities, except for MBA programs, etc. Students have to pay for their books and other costs related to studying. However, they are also entitled to fairly generous student grants and loans. According to the OECD, Denmark has one of the most generous student grants in the world (OECD, 2013). The student grant system was introduced in 1970 and has since then been extended and become more generous several times.

[32] In 2012, Denmark had the highest tax to GDP ratio (47.2%) among the OECD countries. The tax ratio in the OECD area as a whole is 33.7% (OECD, 2014).

Appendix B

Test of Sample Selectivity and Descriptive Statistics

As mentioned in Section 4.1, there is a concern because our sample of children is not representative of the Danish population. While we are not able to determine the direction of potential bias in our sibling estimates, we aim to assess the extent of selectivity in our estimation sample by comparing our sibling sample to a representative sample of children that is not born to a specific birth cohort. In particular, we define a sample that is in the same age range as our sibling sample, whereas the age range is defined by the median age in our sibling sample (i.e., age 32) +/- 4 years. For the resulting sample, which is a representative sample of the 1976 to 1984 birth cohort, we calculate the same descriptive statistics as in Table B1. As can be seen from Table B2, the two samples are largely similar with respect to their observable characteristics. As expected, there is a small difference in the mean birth order between the two samples, which amounts to 1.78 in the representative sample as compared to 1.43 in our sibling sample. This sort of selectivity, however, does not translate into differences between siblings' educational outcomes. An exception is the percentage of brothers holding a tertiary education degree, which is about 4 percentage points higher in the representative sample than in our sibling sample (0.35 vs. 0.31). Hence, our sample might be slightly negatively selected with respect to the tertiary educational attainment of men (though the difference in mean values between the samples is not statistically significant). Considering the characteristics of the siblings' parents, we find hardly any differences between the two samples. This makes us confident that, though our sample is not representative of the Danish population, selectivity is not a major concern in our analysis.

Table B1: Descriptive Statistics

	All siblings		Brothers		Sisters	
	Mean	StdD	Mean	StdD	Mean	StdD
Years of education	14.619	(2.311)	14.292	(2.284)	14.942	(2.293)
Completed upp. sec. education	0.860	(0.347)	0.832	(0.374)	0.887	(0.316)
Completed tertiary education	0.413	(0.492)	0.308	(0.462)	0.516	(0.500)
Female	0.503	(0.500)	–	–	–	–
Age in 2012	32.965	(3.261)	32.976	(3.256)	32.955	(3.267)
Birth order	1.430	(0.587)	1.426	(0.581)	1.433	(0.593)
Observations	3,087		1,534		1,553	

	Fathers		Mothers	
	Mean	StdD	Mean	StdD
Years of education	13.525	(2.822)	13.055	(2.646)
Main occupation, 1980-2012				
Self-employed	0.093	(0.290)	0.022	(0.146)
White-collar worker	0.402	(0.490)	0.507	(0.500)
Blue-collar worker	0.392	(0.488)	0.296	(0.457)
Other worker (not specified)	0.033	(0.178)	0.058	(0.235)
Unemployed	0.015	(0.122)	0.016	(0.124)
Out of the labor force	0.066	(0.248)	0.102	(0.303)
Log of average income, 1980-2012	12.554	(0.492)	12.236	(0.374)
Observations	1,934		1,934	

Table B2: Descriptive Statistics, Whole Danish Population Aged 28 to 36

	All siblings		Brothers		Sisters	
	Mean	StdD	Mean	StdD	Mean	StdD
Years of education	14.649	(2.412)	14.392	(2.416)	14.915	(2.379)
Completed upp. sec. education	0.851	(0.356)	0.822	(0.383)	0.880	(0.324)
Completed tertiary education	0.432	(0.495)	0.350	(0.477)	0.516	(0.500)
Female	0.492	(0.500)	–	–	–	–
Age in 2012	32.220	(2.583)	32.220	(2.585)	32.220	(2.582)
Birth order	1.783	(0.901)	1.784	(0.900)	1.783	(0.902)
Observations	460,899		234,324		226,575	

	Fathers		Mothers	
	Mean	StdD	Mean	StdD
Years of education	13.203	(3.085)	12.958	(2.858)
Main occupation, 1980-2012				
Self-employed	0.117	(0.321)	0.029	(0.169)
White-collar worker	0.382	(0.486)	0.455	(0.498)
Blue-collar worker	0.354	(0.478)	0.274	(0.446)
Other worker (not specified)	0.031	(0.173)	0.060	(0.238)
Unemployed	0.017	(0.129)	0.019	(0.136)
Out of the labor force	0.099	(0.299)	0.163	(0.369)
Log of average income, 1980-2012	12.536	(0.484)	12.158	(0.486)
Observations	346,214		346,214	

Table B3: Descriptive Statistics: Family Characteristics

	Mean	StdD
A. Family structure		
Mother's age at first birth	23.498	(3.291)
Number of siblings	1.150	(0.798)
Lived with both parents until age 16	0.637	(0.473)
Observations	3,087	
B. Social problems		
Mother died before child is aged 16	0.006	(0.075)
Father died before child is aged 16	0.016	(0.121)
Mother convicted of a crime between 1980 to 2012	0.059	(0.236)
Father's crime incidence 1980-2012		
No crime	0.819	(0.385)
Yes, but no imprisonment	0.158	(0.364)
Yes, imprisonment	0.024	(0.152)
Mother's welfare receipt, share of days 1984-2007	0.130	(0.170)
Father's welfare receipt, share of days 1984-2007	0.089	(0.162)
Parent's health status (age 38)		
Excellent	0.657	(0.475)
Good	0.242	(0.428)
Fair	0.072	(0.258)
Poor or very poor	0.029	(0.169)
Observations	2,783	
C. Parent's cognitive skills		
Parent's verbal test score (age 14)	36.210	(7.925)
Parent's spatial test score (age 14)	22.463	(7.289)
Parent's inductive test score (age 14)	22.027	(8.465)
Observations	2,964	
D. Parent's educational preferences (part I)		
Parent likes school (age 14)		
Hate it	0.043	(0.202)
Don't like it	0.091	(0.288)
Don't mind	0.376	(0.485)
Like it	0.414	(0.493)
Like it a lot	0.076	(0.264)
Parent's preferred years of compulsory schooling (age 19)		
7 years	0.239	(0.427)
8 years	0.071	(0.257)
9 years	0.424	(0.494)
10 years	0.239	(0.427)
10 or more years	0.027	(0.163)
Observations	2,314	

Table B4: Descriptive Statistics of Further Family Characteristics and Results of Principle Component Analyses

	Mean/StD	Mean/StD	Mean/StD	Mean/StD
B. Parent's educational preferences (part II)				
	Strongly agree	Agree	Disagree	Strongly disagree
It's stupid to quit school. You'll regret it later (age 14)	0.261	0.242	0.325	0.173
	(0.439)	(0.428)	(0.468)	(0.378)
Parents belive that 7 years of school is enough (age 14)	0.009	0.048	0.260	0.683
	(0.093)	(0.215)	(0.439)	(0.465)
In most cases higher education is a waste (age 14)	0.097	0.168	0.325	0.411
	(0.296)	(0.374)	(0.468)	(0.492)
People waste their youth by staying on in school (age 14)	0.092	0.131	0.256	0.521
	(0.289)	(0.338)	(0.437)	(0.500)
People waste their youth by staying on in school (age 22)	0.053	0.115	0.299	0.532
	(0.224)	(0.319)	(0.458)	(0.499)
	Very important	Somewhat important	Not very important	
Education is important to reach prominent position (age 38)	0.708	0.251	0.041	
	(0.455)	(0.433)	(0.199)	
Results of PCA analysis				
Components with Eigenvalues > 1	Eigenvalues	Proportion		
1. component	1.465	0.244		
2. component	1.061	0.177		
Kaiser-Meyer-Olkin criteria	0.611			
Observations	2,314			
B. Parent's ability to plan for the future				
	Strongly agree	Agree	Disagree	Strongly disagree
There is no reason to think too much about the future (age 14)	0.064	0.107	0.345	0.484
	(0.244)	(0.309)	(0.476)	(0.500)
There is no point in too much planning (age 14)	0.086	0.161	0.380	0.373
	(0.281)	(0.368)	(0.485)	(0.484)
It is better to save up money than to spend it (age 14)	0.301	0.314	0.264	0.120
	(0.459)	(0.464)	(0.441)	(0.326)
It is better to spend money than to save it up (age 14)	0.197	0.203	0.337	0.263
	(0.398)	(0.402)	(0.473)	(0.440)
There is no point in too much planning (age 22)	0.046	0.168	0.416	0.370
	(0.209)	(0.374)	(0.493)	(0.483)
It is better to save up money than to spend it (age 22)	0.117	0.261	0.461	0.161
	(0.322)	(0.439)	(0.499)	(0.367)
It is better to spend money than to save it up (age 22)	0.050	0.164	0.487	0.299
	(0.219)	(0.371)	(0.500)	(0.458)
	Very important	Somewhat important	Not very important	Not important at all
Attitude of life: Being able to plan life many years ahead (age 38)	0.109	0.379	0.440	0.072
	(0.311)	(0.485)	(0.497)	(0.258)
Results of PCA analysis				
Components with Eigenvalues > 1	Eigenvalues	Proportion		
1. component	1.646	0.206		
2. component	1.300	0.163		
3. component	1.016	0.127		
Kaiser-Meyer-Olkin criteria	0.604			
Observations	2,482			

Table B5: Descriptive Statistics: Grandparents' Socio-economic Status

	Mean	StdD
Grandfather's years of education	10.346	(2.636)
Grandmother's years of education	9.376	(2.281)
Grandmother's main occupation		
Not known	0.149	(0.356)
Unemployed/out of the labor force	0.397	(0.489)
Unskilled worker	0.165	(0.371)
Skilled worker	0.030	(0.171)
Routine/non-manual occupation	0.099	(0.298)
Self-employed	0.160	(0.367)
Grandfather's main occupation		
Not known	0.151	(0.358)
Unemployed/out of the labor force	0.005	(0.069)
Unskilled worker	0.193	(0.395)
Skilled worker	0.147	(0.354)
Routine/non-manual occupation	0.217	(0.412)
Self-employed without employees	0.146	(0.354)
Self-employed with employees	0.141	(0.348)
Log of grandparents' income, 1967 and 1968	9.964	(1.837)
Observations	2,487	

Appendix C

Additional Estimation Results

Table C1: Family and Individual Component of REML Estimates for Different Educational Outcomes

	All siblings	**Brothers**	**Sisters**
Years of education			
Family component	1.750	1.618	2.061
StdE	(0.161)	(0.320)	(0.271)
Individual component	3.602	3.621	3.199
StdE	(0.243)	(0.305)	(0.243)
Observations	3,087	1,534	1,553
Completed upp. sec. education			
Family component	0.018	0.030	0.019
StdE	(0.003)	(0.009)	(0.006)
Individual component	0.102	0.110	0.081
StdE	(0.004)	(0.009)	(0.006)
Observations	3,087	1,534	1,553
Completed tertiary education			
Family component	0.073	0.073	0.086
StdE	(0.007)	(0.012)	(0.013)
Individual component	0.170	0.141	0.164
StdE	(0.007)	(0.012)	(0.012)
Observations	3,087	1,534	1,553

Table C2: Siblings' Educational Outcomes and Parents' Socio-Economic Background

	Years of education			Completed upp. sec. education			Completed tertiary education		
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Siblings									
Female	$0.619^{\dagger}$	$0.613^{\dagger}$	$0.594^{\dagger}$	$0.055^{\dagger}$	$0.054^{\dagger}$	$0.050^{\dagger}$	$0.203^{\dagger}$	$0.202^{\dagger}$	$0.199^{\dagger}$
	(0.078)	(0.075)	(0.074)	(0.012)	(0.012)	(0.012)	(0.016)	(0.016)	(0.016)
Age in 2012	0.618^{***}	0.608^{***}	0.551^{***}	0.069^{**}	0.068^{**}	0.053	0.086^{*}	0.084^{*}	0.077^{*}
	(0.220)	(0.212)	(0.211)	(0.035)	(0.034)	(0.034)	(0.047)	(0.045)	(0.045)
Age in 2012 (squared)	$-0.011^{\dagger}$	-0.010^{***}	-0.009^{***}	-0.001^{**}	-0.001^{**}	-0.001^{*}	-0.002^{**}	-0.001^{**}	-0.001^{*}
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Birth order	$-0.449^{\dagger}$	-0.181^{***}	−0.104	$-0.041^{\dagger}$	-0.019^{*}	−0.011	$-0.084^{\dagger}$	-0.032^{**}	−0.016
	(0.070)	(0.069)	(0.069)	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)
Mother									
Years of education	–	$0.193^{\dagger}$	$0.129^{\dagger}$	–	$0.015^{\dagger}$	0.009^{***}	–	$0.041^{\dagger}$	$0.026^{\dagger}$
		(0.017)	(0.019)		(0.003)	(0.003)		(0.004)	(0.004)
Main occupation, 1980-2012 (Ref.: White-collar worker)									
Self-employed	–	–	−0.159	–	–	−0.014	–	–	−0.096
			(0.291)			(0.045)			(0.062)
Blue-collar worker	–	–	-0.319^{***}	–	–	−0.012	–	–	$-0.108^{\dagger}$
			(0.106)			(0.016)			(0.023)
Other worker (not specified)	–	–	-0.372^{**}	–	–	−0.033	–	–	$-0.134^{\dagger}$
			(0.186)			(0.028)			(0.040)
Unemployed	–	–	$-1.303^{\dagger}$	–	–	$-0.215^{\dagger}$	–	–	-0.149^{**}
			(0.337)			(0.052)			(0.072)
Out of the labor force	–	–	$-0.625^{\dagger}$	–	–	$-0.120^{\dagger}$	–	–	-0.101^{***}
			(0.172)			(0.026)			(0.037)
Log of average income, 1980-2012	–	–	0.214	–	–	0.012	–	–	0.032
			(0.135)			(0.021)			(0.029)
Father									
Years of education	–	$0.156^{\dagger}$	$0.110^{\dagger}$	–	$0.014^{\dagger}$	$0.009^{\dagger}$	–	$0.026^{\dagger}$	$0.017^{\dagger}$
		(0.016)	(0.016)		(0.002)	(0.003)		(0.003)	(0.003)
Main occupation, 1980-2012 (Ref.: White-collar worker)									
Self-employed	–	–	0.160	–	–	0.060^{***}	–	–	−0.021
			(0.151)			(0.023)			(0.032)
Blue-collar worker	–	–	-0.266^{***}	–	–	0.000	–	–	$-0.090^{\dagger}$
			(0.101)			(0.015)			(0.021)
Other worker (not specified)	–	–	-0.467^{**}	–	–	−0.044	–	–	-0.086^{*}
			(0.237)			(0.036)			(0.050)
Unemployed	–	–	-0.980^{**}	–	–	$-0.235^{\dagger}$	–	–	−0.095
			(0.383)			(0.060)			(0.082)
Out of the labor force	–	–	-0.488^{**}	–	–	-0.092^{***}	–	–	−0.065
			(0.193)			(0.030)			(0.041)
Log of average income, 1980-2012	–	–	$0.360^{\dagger}$	–	–	0.033^{**}	–	–	$0.065^{\dagger}$
			(0.092)			(0.014)			(0.020)
Constant	6.872^{*}	0.684	−4.172	−0.043	−0.544	−0.762	−0.516	-1.704^{**}	-2.474^{***}
	(3.682)	(3.567)	(4.101)	(0.581)	(0.577)	(0.654)	(0.780)	(0.758)	(0.873)
Sibling correlation	0.285	0.218	0.185	0.133	0.112	0.075	0.256	0.203	0.183
StdE	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Observations	3,087	3,087	3,087	3,087	3,087	3,087	3,087	3,087	3,087

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*

Table C3: Brothers' and Sisters' Educational Outcomes and Parents' Socio-economic Background

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Brothers Coef/StdE	Sisters Coef/StdE	Brothers Coef/StdE	Sisters Coef/StdE	Brothers Coef/StdE	Sisters Coef/StdE
Siblings						
Age in 2012	0.885***	0.187	0.109**	−0.004	0.118*	0.039
	(0.305)	(0.295)	(0.052)	(0.044)	(0.062)	(0.066)
Age in 2012 (squared)	−0.013***	−0.003	−0.002**	−0.000	−0.002**	−0.001
	(0.005)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Birth order	−0.088	−0.120	−0.019	−0.008	−0.007	−0.020
	(0.101)	(0.095)	(0.017)	(0.014)	(0.020)	(0.021)
Mother						
Years of education	0.112†	0.142†	0.009**	0.010**	0.019†	0.032†
	(0.026)	(0.026)	(0.004)	(0.004)	(0.005)	(0.006)
Main occupation, 1980-2012 (Ref.: White-collar worker)						
Self-employed	−0.642	0.275	−0.035	0.027	−0.156*	−0.049
	(0.406)	(0.402)	(0.068)	(0.058)	(0.083)	(0.089)
Blue-collar worker	−0.248*	−0.412***	0.001	−0.027	−0.127†	−0.091***
	(0.148)	(0.145)	(0.025)	(0.021)	(0.030)	(0.032)
Other worker (not specified)	−0.420	−0.368	−0.000	−0.065*	−0.206†	−0.067
	(0.260)	(0.256)	(0.044)	(0.037)	(0.053)	(0.057)
Unemployed	−1.525†	−1.173**	−0.263†	−0.163**	−0.182**	−0.130
	(0.445)	(0.482)	(0.075)	(0.070)	(0.091)	(0.107)
Out of the labor force	−0.703***	−0.594**	−0.124***	−0.114†	−0.134***	−0.075
	(0.235)	(0.238)	(0.039)	(0.034)	(0.048)	(0.052)
Log of average income, 1980-2012	0.134	0.243	0.026	−0.003	0.007	0.056
	(0.195)	(0.178)	(0.033)	(0.026)	(0.040)	(0.039)
Father						
Years of education	0.119†	0.097†	0.012***	0.006*	0.020†	0.014***
	(0.023)	(0.023)	(0.004)	(0.003)	(0.005)	(0.005)
Main occupation, 1980-2012 (Ref.: White-collar worker)						
Self-employed	0.177	0.084	0.070*	0.049*	−0.023	−0.032
	(0.220)	(0.199)	(0.037)	(0.029)	(0.045)	(0.044)
Blue-collar worker	−0.215	−0.285**	−0.010	0.012	−0.089***	−0.083***
	(0.138)	(0.140)	(0.023)	(0.020)	(0.028)	(0.031)
Other worker (not specified)	−0.844**	−0.156	−0.093*	−0.006	−0.159**	−0.013
	(0.330)	(0.327)	(0.055)	(0.047)	(0.068)	(0.072)
Unemployed	−0.455	−1.416***	−0.148	−0.322†	−0.011	−0.153
	(0.543)	(0.513)	(0.092)	(0.075)	(0.111)	(0.114)
Out of the labor force	−0.593**	−0.274	−0.132***	−0.047	−0.047	−0.069
	(0.262)	(0.280)	(0.044)	(0.040)	(0.054)	(0.062)
Log of average income, 1980-2012	0.294***	0.625†	0.036*	0.037	0.049**	0.121***
	(0.112)	(0.168)	(0.019)	(0.024)	(0.023)	(0.037)
Constant	−7.970	−1.037	−1.964**	0.450	−2.625**	−2.614**
	(5.855)	(5.799)	(0.996)	(0.860)	(1.191)	(1.291)
Sibling correlation	0.212	0.254	0.153	0.124	0.250	0.223
StdE	(0.059)	(0.052)	(0.061)	(0.056)	(0.057)	(0.052)
Observations	1,534	1,553	1,534	1,553	1,534	1,553

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*

Table C4: Contribution of Family Characteristics to Sibling Correlations

	Years of education		Completed upp. sec. ed.		Completed tertiary ed.	
	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$	ρ	$\%\Delta\rho$
Raw correlation (only individual controls)						
Sibling correlation	0.285	–	0.133	–	0.256	–
StdE	(0.028)		(0.028)		(0.028)	
Family structure						
Sibling correlation	0.255	10%	0.100	22%	0.244	5%
StdE	(0.028)		(0.028)		(0.028)	
Social problems						
Sibling correlation	0.240	15%	0.095	38%	0.237	6%
StdE	(0.029)		(0.029)		(0.030)	
Parent's cognitive skills						
Sibling correlation	0.249	12%	0.118	8%	0.226	12%
StdE	(0.028)		(0.029)		(0.029)	
Parent's ability to plan for the future						
Sibling correlation	0.284	5%	0.162	7%	0.251	5%
StdE	(0.030)		(0.031)		(0.031)	
Parent's educational preferences						
Sibling correlation	0.284	9%	0.133	16%	0.256	7%
StdE	(0.032)		(0.033)		(0.032)	
Grandparents' soecio-economic status						
Sibling correlation	0.243	7%	0.084	15%	0.253	3%
StdE	(0.032)		(0.032)		(0.031)	
All family characteristics						
Sibling correlation	0.224	25%	0.049	56%	0.226	15%
StdE	(0.033)		(0.035)		(0.033)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – Note that the number of observations varies with the control variables considered. The percentage change in the sibling correlation is therefore calculated based on the raw sibling correlation for the specific sample. – Note also that the number of observations reduces to 2,200 when all family characteristics are considered. Hence, the results have to be interpreted with some caution.

Table C5: Siblings' Educational Outcomes and Family Structure

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Mother's age at first birth	$0.623^{\dagger}$	0.310^{*}	0.069^{***}	0.040	$0.128^{\dagger}$	0.069^{**}
	(0.174)	(0.163)	(0.026)	(0.025)	(0.037)	(0.035)
Mother's age at first birth (squared)	-0.010^{***}	−0.005	-0.001^{**}	−0.001	-0.002^{***}	−0.001
	(0.004)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
No. of siblings	−0.000	−0.024	−0.008	−0.009	0.010	0.007
	(0.056)	(0.052)	(0.008)	(0.008)	(0.012)	(0.011)
Lived with both parents until age 16	$0.599^{\dagger}$	$0.581^{\dagger}$	$0.098^{\dagger}$	$0.080^{\dagger}$	$0.066^{\dagger}$	$0.076^{\dagger}$
	(0.092)	(0.089)	(0.014)	(0.014)	(0.019)	(0.019)
Constant	−0.333	-7.888^{*}	−0.610	-1.222^{*}	-2.088^{***}	$-3.214^{\dagger}$
	(3.777)	(4.142)	(0.591)	(0.659)	(0.802)	(0.888)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.255	0.163	0.100	0.053	0.244	0.177
StdE	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.029)
Observations	3,078	3,078	3,078	3,078	3,078	3,078

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method.

Table C6: Siblings' Educational Outcomes and Social Problems

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Mother died before child is aged 16	-1.426^{**}	-1.182^{**}	-0.074	-0.032	-0.267^{**}	-0.255^{**}
	(0.579)	(0.561)	(0.086)	(0.086)	(0.125)	(0.121)
Father died before child is aged 16	-0.238	0.101	-0.035	0.021	-0.015	0.030
	(0.355)	(0.351)	(0.052)	(0.054)	(0.077)	(0.076)
Mother convicted of a crime between 1980 to 2012	-0.305	-0.389^{**}	-0.083^{***}	-0.090^{***}	-0.003	-0.019
	(0.198)	(0.188)	(0.028)	(0.028)	(0.043)	(0.041)
Father's crime incidence 1980-2012 (Ref.: No crime)						
Yes, but no imprisonment	-0.381^{***}	-0.344^{***}	-0.053^{***}	$-0.059^{\dagger}$	-0.040	-0.026
	(0.125)	(0.120)	(0.018)	(0.018)	(0.027)	(0.026)
Yes, imprisonment	$-1.310^{\dagger}$	-0.986^{***}	$-0.224^{\dagger}$	$-0.207^{\dagger}$	-0.161^{**}	-0.094
	(0.322)	(0.309)	(0.047)	(0.047)	(0.070)	(0.067)
Mother's welfare receipt, share of days 1984-2007	$-2.228^{\dagger}$	$-1.784^{\dagger}$	$-0.333^{\dagger}$	$-0.310^{\dagger}$	$-0.325^{\dagger}$	-0.219^{***}
	(0.298)	(0.385)	(0.043)	(0.057)	(0.064)	(0.083)
Father's welfare receipt, share of days 1984-2007	-0.906^{***}	-0.127	-0.139^{***}	-0.050	-0.127^{*}	-0.021
	(0.310)	(0.428)	(0.045)	(0.064)	(0.067)	(0.092)
Parent's health status at age 38 (Ref.: Excellent)						
Good	-0.220^{**}	-0.159	-0.035^{**}	-0.032^{**}	-0.032	-0.018
	(0.107)	(0.102)	(0.015)	(0.015)	(0.023)	(0.022)
Fair	-0.150	-0.099	0.019	0.016	-0.054	-0.039
	(0.182)	(0.173)	(0.026)	(0.026)	(0.039)	(0.037)
Poor or very poor	0.165	0.061	0.072^{*}	0.070^{*}	-0.039	-0.066
	(0.280)	(0.267)	(0.040)	(0.040)	(0.060)	(0.058)
Constant	9.251^{**}	-2.229	0.595	-0.357	-0.335	-2.413^{**}
	(3.774)	(4.340)	(0.580)	(0.672)	(0.818)	(0.937)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.240	0.182	0.095	0.075	0.237	0.181
StdE	(0.029)	(0.030)	(0.029)	(0.030)	(0.030)	(0.030)
Observations	2,783	2,783	2,783	2,783	2,783	2,783

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*

Table C7: Siblings' Educational Outcomes and Parent's Cognitive Skills

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Parent's verbal test score	$0.036^{\dagger}$	0.011	0.002	-0.000	$0.008^{\dagger}$	0.003^{*}
	(0.007)	(0.007)	(0.001)	(0.001)	(0.002)	(0.002)
Parent's spatial test score	0.010	0.007	0.001	0.001	0.001	0.000
	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Parent's inductive test score	$0.027^{\dagger}$	0.014^{**}	0.003^{**}	0.002	$0.006^{\dagger}$	0.003^{**}
	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	3.821	-2.987	-0.441	-0.743	-1.044	-2.138^{**}
	(3.754)	(4.200)	(0.593)	(0.665)	(0.796)	(0.898)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.249	0.180	0.118	0.071	0.226	0.181
StdE	(0.028)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Observations	2,964	2,964	2,964	2,964	2,964	2,964

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.* – *The intelligence test was conducted when the DLSY respondents were 14 years old.*

Table C8: Siblings' Educational Outcomes and Parent's Ability to Plan for the Future

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Parent's ability to plan for the future (1. comp. PCA)	0.198†	0.056	0.012**	0.003	0.049†	0.019**
	(0.038)	(0.037)	(0.006)	(0.006)	(0.008)	(0.008)
Parent's ability to plan for the future (2. comp. PCA)	0.046	0.054	0.011*	0.010*	0.003	0.006
	(0.043)	(0.040)	(0.006)	(0.006)	(0.009)	(0.009)
Parent's ability to plan for the future (3. comp. PCA)	0.208†	0.097**	0.029†	0.021***	0.026**	0.003
	(0.049)	(0.046)	(0.007)	(0.007)	(0.010)	(0.010)
Constant	7.188*	−6.902	0.126	−0.934	−0.391	−2.944***
	(4.117)	(4.810)	(0.628)	(0.740)	(0.890)	(1.041)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.284	0.209	0.162	0.125	0.251	0.196
StdE	(0.030)	(0.032)	(0.031)	(0.032)	(0.031)	(0.031)
Observations	2,482	2,482	2,482	2,482	2,482	2,482

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.* – *1., 2. and 3. comp. of PCA refers to the respective principal component (eigenvector) of a principal component analysis conducted on different questions covering the respective issue.* – *Questions regarding the DLSY respondent's ability to plan ahead were asked at age 14, 22 and 38.*

Table C9: Siblings' Educational Outcomes and Parent's Educational Preferences

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Parent likes school (Ref.: Don't mind)						
Hate it	−0.615**	−0.399	−0.079**	−0.061*	−0.128**	−0.082
	(0.262)	(0.245)	(0.037)	(0.036)	(0.056)	(0.053)
Don't like it	−0.804†	−0.595†	−0.119†	−0.103†	−0.126***	−0.081**
	(0.186)	(0.175)	(0.026)	(0.026)	(0.040)	(0.038)
Like it	0.163	0.036	−0.000	−0.009	0.046*	0.021
	(0.113)	(0.106)	(0.016)	(0.016)	(0.024)	(0.023)
Like it a lot	0.403**	0.340*	0.061**	0.057**	0.027	0.011
	(0.202)	(0.189)	(0.028)	(0.028)	(0.043)	(0.041)
Parent's preferred years of compulsory schooling (Ref.: 7 years)						
8 years	0.161	0.295	0.014	0.036	0.043	0.053
	(0.218)	(0.204)	(0.030)	(0.030)	(0.046)	(0.044)
9 years	0.166	0.103	0.031*	0.025	0.038	0.027
	(0.129)	(0.121)	(0.018)	(0.018)	(0.028)	(0.026)
10 years	0.132	0.026	0.027	0.025	0.025	−0.003
	(0.147)	(0.138)	(0.021)	(0.020)	(0.031)	(0.030)
10 or more years	0.038	−0.136	0.038	0.040	−0.012	−0.065
	(0.314)	(0.297)	(0.045)	(0.044)	(0.067)	(0.064)
Parent's educational preferences (1. comp. PCA)	0.176†	0.035	0.004	−0.004	0.044†	0.013
	(0.043)	(0.041)	(0.006)	(0.006)	(0.009)	(0.009)
Parent's educational preferences (2. comp. PCA)	0.063	0.061	0.008	0.008	0.013	0.011
	(0.051)	(0.048)	(0.007)	(0.007)	(0.011)	(0.010)
Constant	2.273	−11.931**	−0.616	−1.476*	−1.064	−3.492***
	(4.326)	(5.026)	(0.655)	(0.767)	(0.939)	(1.092)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.284	0.217	0.133	0.109	0.256	0.206
StdE	(0.032)	(0.033)	(0.033)	(0.033)	(0.032)	(0.032)
Observations	2,314	2,314	2,314	2,314	2,314	2,314

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.* – *1. and 2. comp. of PCA refers to the respective principal component (eigenvector) of a principal component analysis conducted on different questions covering the respective issue.* – *The question on whether the DLSY respondent likes school was asked at age 14.* – *The question on the DLSY respondent's preferred years of schooling was asked at age 19.* – *Questions regarding the DLSY respondent's educational preferences were asked at age 14, 22 and 38.*

Table C10: Siblings' Educational Outcomes and Grandparents' Socio-economic Status

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Grandfather's years of education	0.011	−0.023	−0.000	−0.002	0.002	−0.006
	(0.022)	(0.020)	(0.003)	(0.003)	(0.005)	(0.004)
Grandmother's years of education	0.041	−0.006	0.001	−0.003	0.011**	0.000
	(0.025)	(0.024)	(0.004)	(0.004)	(0.005)	(0.005)
Grandmother's main occupation (Ref.: Routine/ non-manual occupation)						
Not known	−0.405	−0.148	−0.078	−0.064	−0.049	0.010
	(0.329)	(0.308)	(0.048)	(0.046)	(0.071)	(0.067)
Unemployed/out of the labor force	−0.045	0.047	0.011	0.015	−0.035	−0.012
	(0.179)	(0.168)	(0.026)	(0.025)	(0.038)	(0.036)
Unskilled worker	−0.510**	−0.319	−0.065**	−0.051*	−0.083*	−0.045
	(0.208)	(0.196)	(0.030)	(0.030)	(0.045)	(0.042)
Skilled worker	−0.343	−0.166	0.022	0.036	−0.093	−0.051
	(0.315)	(0.296)	(0.046)	(0.045)	(0.068)	(0.064)
Self-employed	0.096	0.112	0.019	0.019	−0.015	−0.010
	(0.223)	(0.210)	(0.032)	(0.032)	(0.048)	(0.045)
Grandfather's main occupation (Ref.: Routine/ non-manual occupation)						
Not known	−0.550*	−0.294	0.013	0.046	−0.156**	−0.113*
	(0.314)	(0.295)	(0.045)	(0.044)	(0.068)	(0.064)
Unemployed/out of the labor force	−1.289*	−0.524	−0.268***	−0.174*	−0.179	−0.030
	(0.709)	(0.673)	(0.102)	(0.101)	(0.152)	(0.145)
Unskilled worker	−0.637†	−0.391**	−0.032	−0.014	−0.141†	−0.086**
	(0.167)	(0.157)	(0.024)	(0.024)	(0.036)	(0.034)
Skilled worker	−0.501***	−0.221	−0.020	−0.000	−0.134†	−0.070**
	(0.166)	(0.157)	(0.024)	(0.024)	(0.036)	(0.034)
Self-employed without employees	−0.279	−0.106	0.001	0.009	−0.070*	−0.026
	(0.186)	(0.176)	(0.027)	(0.027)	(0.040)	(0.038)
Self-employed with employees	−0.101	−0.068	0.024	0.020	−0.059	−0.044
	(0.182)	(0.171)	(0.026)	(0.026)	(0.039)	(0.037)
Log of grandparents' taxable annual income (1967, 1968)	−0.034	−0.031	−0.003	−0.003	−0.009	−0.008
	(0.026)	(0.024)	(0.004)	(0.004)	(0.006)	(0.005)
Constant	3.952	−9.141*	−0.583	−1.455*	−0.361	−2.535**
	(4.246)	(4.917)	(0.665)	(0.773)	(0.907)	(1.056)
Parents' SES	no	yes	no	yes	no	yes
Sibling correlation	0.243	0.176	0.084	0.056	0.253	0.191
StdE	(0.032)	(0.032)	(0.032)	(0.032)	(0.031)	(0.032)
Observations	2,485	2,485	2,485	2,485	2,485	2,485

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*