I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13441

Family Disadvantage, Gender and the Returns to Genetic Human Capital

Victor Ronda
Esben Agerbo
Dorthe Bleses
Preben Bo Mortensen
Anders Børglum
David M. Hougaard
Ole Mors
Merete Nordentoft
Thomas Werge
Michael Rosholm

JULY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13441

Family Disadvantage, Gender and the Returns to Genetic Human Capital

Victor Ronda
Aarhus University

Esben Agerbo
Aarhus University

Dorthe Bleses
Aarhus University

Preben Bo Mortensen
Aarhus University and iPSYCH

Anders Børglum
iPSYCH and Aarhus University

David M. Hougaard
iPSYCH and Statens Serum Institut

Ole Mors
iPSYCH and Aarhus University

Merete Nordentoft
iPSYCH and Mental Health Centre Copenhagen

Thomas Werge
iPSYCH and Mental Health Centre Sct. Hans

Michael Rosholm
Aarhus University and IZA

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Family Disadvantage, Gender and the Returns to Genetic Human Capital

This paper relies on a large-scale sample of genotyped individuals linked with detailed register data in Denmark to investigate the context-dependence of genetic influences on human capital formation. We show that the returns to genetic endowments, measured by a polygenic score for educational attainment, are significantly attenuated by childhood disadvantage. We replicate the findings in a within-family analysis, where we exploit exogenous genetic variation across siblings to control for unobserved family influences. We also explore gender differences in the context-dependence of genetic influences and find the attenuation effect of childhood disadvantage on educational attainment to be significantly stronger for males than for females. We show our findings extend to a representative sample of the Danish population. Our results highlight an important mechanism driving the persistence of disadvantage across generations. We show that children who experience childhood disadvantage are not able to fully realize their educational potential, even in the context of the generous Danish welfare-state.

JEL Classification: I1,I21,I24

Keywords: family disadvantage, education, genomics, gene-environment interactions

Corresponding author:
Victor Ronda
Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
8210 Aarhus V
Denmark

E-mail: ronda@econ.au.dk

1 Introduction

A central question in the study of inequality is how social and economic disadvantage persists across generations. It is well understood that inequality in family resources translates into inequality in children's outcomes and that adult life outcomes are to a large extent explained by skills acquired in childhood (Cunha, Heckman, and Navarro, 2005; Heckman and Mosso, 2014). It is also well established that skills are partly determined by genetic endowments realized at conception (see, e.g. Plomin and von Stumm, 2018; Lee et al., 2018). It is less understood, however, whether and how genetic endowments interact with family resources in the process of human capital formation. These interactions have important implications for intergenerational mobility and economic inequality. Heterogeneity in genetic influences across childhood environments is informative about different children's opportunities for realizing their full genetic potential, and hence, for disadvantaged children's ability to escape their parents' social and economic circumstances.

In this paper, we demonstrate that genetic influences on human capital formation are attenuated by childhood disadvantage in Denmark. We use individual-level genotypic data obtained from a large sample of genotyped Danish individuals and rely on recent advances in molecular biology to construct a polygenic score (PGS) as an individual measure of the genetic potential for education. The PGS is constructed using summary statistics from the most recent genome-wide association study (GWAS) of educational attainment (Lee et al., 2018). Polygenic scores for educational attainment have been shown to predict early life skills (Belsky et al., 2016), achievement in school (Ward et al., 2014), educational attainment (Rietveld et al., 2013; Domingue et al., 2015; Okbay et al., 2016; Lee et al., 2018), as well as earnings, socioeconomic mobility and wealth, over and above the effect of education (Papageorge and Thom, 2020; Belsky et al., 2018; Barth, Papageorge, and Thom, 2020). We match the genotyped individuals with the Danish population registries to obtain detailed measures of educational attainment and academic achievement as well as detailed measures of early childhood environment.

While individual genes are pre-determined at conception, polygenic scores are not guaranteed to be exogenous in most models. Children's genes are inherited from their parents, and, as a result, a child's genetic potential for education is also a measure of her parents' genetic potential for education and the parents' investment potential in the child (see the concept of genetic nurture in Kong et al., 2018; Young et al., 2018). For this and other reasons (e.g. population stratification, Novembre et al., 2008) polygenic scores often overestimate genetic influences. One way to address these problems is to rely on genetic variation between siblings, since conditional on parental genotype, child genotypes are randomly as-

signed at conception. As a result, genetic variation across a pair of siblings is independent of any environmental effect (see Conley and Fletcher, 2017).

We demonstrate this point by comparing the relation between the polygenic score and human capital outcomes in our full sample and in a sibling sample, where we control for family fixed effects. Using the full sample, we show that the PGS explains a large share (7 to 9%) of the variation in educational attainment and school achievement in Denmark. We also show that a one standard deviation increase in the score is associated with 0.56-0.60 more years of schooling, 11-12%-points larger probability of attending a post-secondary education, and 6-7 percentile better ranking in Danish and mathematics school leaving test scores. However, using the sibling sample, we show that these effects are over-estimated and decrease by 40-50% once we control for family fixed effects. These results demonstrate that the coefficient estimated on the polygenic score in the full sample captures information about children's environment that is important for education, in addition to children's genetic potential. This is an additional cause of concern when estimating genetic effects across groups, as differences in estimated genetic effects across groups can reflect differences in the confounding bias across groups. This issue can also be addressed by relying on the genetic variation between siblings within families.

In our main set of results, we show that the genetic influences on human capital formation are significantly smaller for children growing up in disadvantage than for children growing up in normal circumstances. Our measure of disadvantage captures different dimensions of childhood environment, including parental human capital, family resources, family stability, and parental mental health. We show that the genetic influences on educational attainment and academic achievement are 20-40% smaller in the group that experienced disadvantage. We find an even larger effect of childhood disadvantage when estimating the effects in the sibling sample and controlling for family fixed effects. These findings demonstrate that the payoffs to genetic endowments depend on the family environment experienced by the child, a similar idea to a growing literature in economics that demonstrates that payoffs to skills are context-dependent (Lundberg, 2013; Papageorge, Ronda, and Zheng, 2019). Our results are also related to a growing literature documenting the context-dependence of genetic influences using polygenic scores (e.g. Barcellos, Carvalho, and Turley, 2018; Papageorge and Thom, 2020; Trejo et al., 2018). In particular, our findings are mostly related to Papageorge and Thom (2020), which documents that the relationship between the polygenic score for education and college graduation in the US is stronger for individuals who grew up in households with higher socioeconomic status relative to those who grew up in poorer households. We extend their findings as we show that these group differences are causal and stronger within families. We also complement their paper, as we show that children growing up in

disadvantage are not able to fully realize their educational potential even in the context of the generous Danish/Nordic welfare-state.

We also document significant gender differences in the relationship between genetic influences on human capital and family disadvantage. An emerging literature in economics motivates this analysis, as it documents substantial gender differences in response to family disadvantage (Autor et al., 2019; Bertrand and Pan, 2013). If there are gender differences in response to the childhood environment, then it is also possible that there are gender differences in genetic influences across the family environment. Consistent with the literature, we find that family disadvantage has a stronger attenuation effect on the genetic influences on educational attainment for males than for females. These findings suggest that gender gaps in education are not only heterogeneous across the family environment but also with respect to the genetic predisposition for education.

Lastly, we show that the context-dependence of genetic influences on human capital formation is present in a representative sample of the Danish population. One possible concern with our data is that it includes many individuals diagnosed with a psychiatric problem, such as schizophrenia and ADHD, and, as a result, is not representative of the Danish population. The importance of representative samples for analyzing genetic influences is highlighted in a recent paper by Pirastu et al. (2020). They also demonstrate that, in their case, this is not a major concern in the iPSYCH sample. Nevertheless, to check for the external validity of our results, we re-estimate our main findings separately for a group of individuals randomly selected from the Danish population. We find the same attenuation of genetic influences by childhood disadvantage in the representative group, which suggests that our main results are externally valid to the rest of the Danish population.

The rest of the paper is organized as follows: Section 2 introduces our data set and describes our measure of genetic potential in more detail. Section 3 describes our empirical framework and how sibling genetic variation can be used to estimate the causal genetic effects. We describe our empirical findings in Section 4. Section 5 discusses the implications of our results and concludes.

2 Data

The Integrative Psychiatric Research (iPSYCH) study is a large Danish population-based Case-Control sample aimed at unraveling the genetic and environmental architecture of severe mental disorders. It follows 77,639 individuals selected among the Danish population born between 1981 and 2005 (1,472,762 individuals). About one-third of the sample, 27,605

individuals, is comprised of individuals randomly selected from the 1981-2005 Danish cohorts. This is an important part of the research design as it allows us to check the validity of our findings in a representative sample of the danish population. The remaining sample consists of individuals diagnosed with one or more psychiatric disorders. Of the total sample, 2,738 were diagnosed with schizophrenia, 1,452 with bipolar disorder, 14,812 with autism, 17,249 with attention-deficit/hyperactivity disorder, and 22,809 with an affective disorder (Pedersen et al., 2018). The initial sample included 86,189 individuals, but of these 8,550 were excluded due to unsuccessful genotyping or quality control. DNA was extracted from the neonatal dried blood spot samples obtained from the Danish Neonatal Screening Biobank and genotyped using the Illumina PsychChip.

Individuals were matched to the Danish population registries. As a result, the iPSYCH sample also includes longitudinal information on health outcomes, educational attainment, schooling outcomes, and other social and socioeconomic data. Parents and other relatives are also matched, which allows us to combine individual outcomes to detailed information on their socioeconomic background. For more information, we refer to Pedersen et al. (2018), which describes the iPSYCH study in more detail.

We restricted our analysis to study members with European descent (90.5%), matched with both biological parents at the birth registers (99.4%), without missing information on the family environment (82.8%) and with at least one relevant outcome (81.96%). Our restricted sample includes 49,756 individuals in the study, of which 16,581 (33.3%) are controls, and 33,175 (66.7%) are psychiatric cases.

2.1 Childhood Disadvantage

We consider four different dimensions of poor childhood environment. The first dimension captures parental human capital. We measure human capital disadvantage as having both parents without any post-secondary education (18% of the sample). Our second dimension captures family resources. We measure resources disadvantage as growing up in a family in the lowest quintile of disposable family income. We measure disposable family income as the average maternal and paternal disposable income at ages 1 to 10. Our third dimension captures family stress. We proxy family stress with an indicator for whether either parent was ever diagnosed with a mental health condition (22% of the sample). Our fourth dimension captures family instability. We measure family instability as growing up in a broken family, with non-cohabiting parents, between the ages 0 to 10 (28% of the sample). Table 1 provides summary statistics on the measures of childhood environment by disadvantage group

About half of the sample, 22,927 (46%), experienced no dimension of childhood disad-

Table 1: Summary Statistics

	[All]	[No Dis.]	[Disad.]
Childhood Disadvantage:			
Low Parental Education	0.183	0.067	0.553
	(0.387)	(0.249)	(0.497)
Low Family Income	0.200	0.086	0.561
	(0.400)	(0.280)	(0.496)
Parental M.H. Diag.	0.221	0.097	0.612
	(0.415)	(0.296)	(0.487)
Broken Family	0.284	0.144	0.727
	(0.451)	(0.351)	(0.445)
Disadvantage Group	0.240	0.000	1.000
	(0.427)	(.)	(.)
Genetic Endowments:			
EA Polygenic Score	0.000	0.084	-0.267
	(1.000)	(0.994)	(0.971)
Human Capital Outcomes:			
Years of Education at 25	12.669	13.006	11.692
	(2.235)	(2.228)	(1.948)
Any Post Secondary Education	0.313	0.370	0.148
	(0.464)	(0.483)	(0.355)
Danish Percentile	38.461	40.375	31.146
	(22.170)	(22.255)	(20.238)
Mathematics Percentile	39.245	41.777	29.327
	(25.268)	(25.314)	(22.496)
Child Characteristics:			
Child is a Female	0.512	0.515	0.503
	(0.500)	(0.500)	(0.500)
Birth Year	1990	1990	1990
	(5.085)	(5.138)	(4.878)
Birth-order	1.713	1.720	1.692
	(0.735)	(0.726)	(0.763)
Psychiatric Case	0.667	0.630	0.782
	(0.471)	(0.483)	(0.412)
Obs.	49,756	37,809	11,974

Notes: Summary statistics for the analytic sample of 49,756 individuals. Statistics are reported separately for all individuals (Column [1]), for individuals in the non-disadvantage group (Column [2]) and for individuals part of the disadvantage group (Column [3]). The childhood disadvantage measures and the post secondary education, psychiatric case and gender categories entries are in the form of percentages divided by 100. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.

vantage, 14,882 (30%) experienced one dimension of disadvantage, and 11,974 (24%) experienced two or more dimensions of disadvantage. In our benchmark specification, we classify the latter group as the disadvantaged group. We believe that this classification is a fair

representation of childhood disadvantage as it captures more than one dimension and corresponds to about one-quarter of our sample. In a robustness exercise, in the supplemental materials, we consider alternative specifications of childhood disadvantage, including each dimension individually and the number of dimensions experienced during childhood. We show our main results are robust to different definitions of childhood disadvantage. Due to the over-representation of mental health cases in our sample, our sample has a higher percentage of disadvantaged individuals than the whole population. As a comparison, individuals in the psychiatric case sample (i.e., individuals with schizophrenia or other iPSYCH disorders) were almost twice as likely to be part of the disadvantaged group than individuals in the control sample. With that concern in mind, in Section 4.4, we show our main results replicate when we use only the representative sample.

2.2 Genetic Endowments

We measure genetic variation using a polygenic score that predicts educational attainment (EA PGS). The polygenic score is a weighted sum of single-nucleotide polymorphisms (SNPs). The weights used to create the score were based on the summary statistics from the most recent genome-wide association study (GWAS) for educational attainment that identified 1,271 SNP's significantly associated with educational attainment (Lee et al., 2018).[1] Importantly, the Lee et al. study did not include the individuals part of the iPSYCH sample.

The polygenic score was constructed using all imputed SNPs after quality control and pruning [2]. All iPSYCH samples were genotyped using Illumina PsychChip (Illumina, San Diego, CA, USA). Genotypes were processed using standard procedures for stringent quality control, which included filters for call rate, Hardy-Weinberg equilibrium, and heterozygosity rates. Each cohort was then phased and imputed using the 1000 Genomes Project phase 3 imputation reference panel (see Grove et al. (2017) for more information on quality control and imputation in iPSYCH).

It is important to highlight that, while we describe the polygenic score as a measure of genetic endowment, it is far from clear what is being captured by the score. The polygenic score is an index composed of many genetic variants from across the genome, each with a small effect on educational attainment. The same variants that predict education also predict a variety of additional traits, such as bipolar disorder, schizophrenia, and height (Bulik-Sullivan et al., 2015). In addition, the score can carry information about the individual's

[1] We use the public available summary statistics to construct the polygenic score. The summary statistics includes all discovery cohorts except 23andMe.

[2] We include all SNPs during pruning, imposing strict significance thresholds do not significantly alter our key findings.

environment.

This can be seen in Table 1, where we provide summary statistics on the polygenic scores by disadvantage group. The score was normalized to have mean zero and standard deviation one in the whole sample. The distribution of the score is different across the two groups. The average score for individuals growing up in disadvantage is 0.27 standard deviation smaller than the average for the whole sample. The difference across groups can also be seen in Figure 1, where we plot the distribution of the polygenic score for educational attainment separately for individuals in the two groups. The difference in the distribution of the score across groups is not surprising given that individual's genotypes were inherited from their parents, and parental genetic endowments influenced parental outcomes, such as education and income. The distribution of the polygenic score does not depend on gender. The main reason is that the polygenic score does not include genetic information from the sex chromosomes.[3]

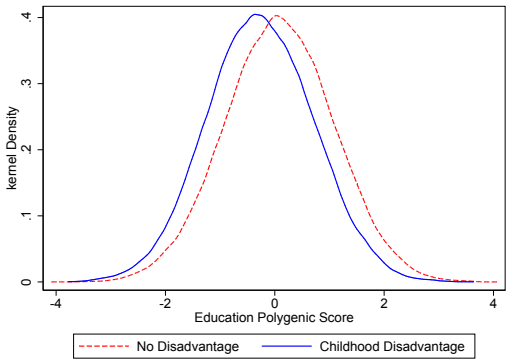


Figure 1: EA PGS by Childhood Disadvantage: Figure 1 shows the estimated distribution of the educational attainment polygenic score across childhood disadvantage groups. The figure shows that individuals growing up in better childhood environments have, on average, a higher genetic propensity for educational attainment.

2.3 Human Capital Outcomes

We consider four measures of human capital. The measures were constructed to capture academic achievement and educational attainment. We measure educational attainment as years of education at age 25 and any post-secondary education by age 25. Information on educational attainment is available for 28,331 individuals aged 25 or older in 2016. We

[3]Even so, Pirastu et al. (2020) demonstrate that in some selective samples, such as 23andMe and UK Biobank, it is still possible to predict sex from the autosomes. Their interpretation is that genes to some extent determine whether or not one decides to participate in these samples, again underscoring the importance of representative samples.

measure school achievement from Danish school-leaving exams at the end of primary school. We separate school achievement into Danish and mathematics test scores and construct the scores as cohort-specific percentile ranks.[4] The exam grades are available since the year 2000 for individuals graduating from primary school by the year 2016. As a result, Danish and mathematics scores are available for a total of 33,286 and 32,528 individuals born between the years 1985 and 2000.[5]

Table 1 provides summary statistics on the four human capital outcomes by disadvantaged groups. On average, individuals in our sample completed 12.7 years of education by age 25. Individuals that experienced two or more dimensions of disadvantage complete 1.3 fewer years of education on average than individuals in the non-disadvantage group. Similarly, on average 31.3% of the individuals in our sample attended a post-secondary school program, while only 14.8% of the individuals who experienced two or more dimensions of disadvantage did so. We also find large gaps by childhood disadvantage on school achievement. On average, individuals in the childhood disadvantage group scored 9 and 12 percentile points lower in the Danish and mathematics school leaving exams than individuals in the non-disadvantage group.

3 Estimation Model

To investigate the context-dependence of genetic influences on human capital formation, we consider two linear models: a between-family model, where we include all individuals in our sample, and a within-family model, where we include families with two or more siblings in the sample and control for family fixed effects.

The between-family model is the standard model used to study gene-by-environment interactions (e.g. Papageorge and Thom, 2020).[6] It has the advantage that it can be estimated on the whole sample, increasing our power to detect differences in association across family

[4]Danish and mathematics test scores were calculated as averages of cohort-specific percentile ranks on four Danish and four mathematics exams. We rely on cohort-specific ranks as different cohorts were exposed to a different combination of exams. Each exam was graded on a 7 point scale, and as a result, the average percentile score is not precisely 50%.

[5]Given the data limitations, our sample size decreases substantially to about 13,000 individuals when we consider individuals with measures of both educational attainment and school achievement. The between-family results replicate when considering the sample of individuals with all four outcomes. However, considering only individuals with all four outcomes decreases the sibling sample substantially, and we lose the power to identify effects using the within-family model.

[6]We use the term between-family loosely as the specification includes families with more than one individual. As a result, our between-family estimates includes both between-family and within-family effects. In results available upon request, we re-estimated the model with only one child per family and found no significant differences in the findings.

and gender groups. Formally, the between-family model can be described as

$$Y_{igj} = \alpha_j^{bf}(g)PGS_{ig} + X_{ig}\beta_j^{bf}(g) + \epsilon_{igj}^{bf} \tag{1}$$

where Y_{igj} is the jth human capital outcome of individual i in group g, PGS_{ig} is the polygenic score of individual i in group g, X_{ig} includes a set of control variables, and ϵ_{igj}^{bf} is an individual specific error term for outcome j. X_{ig} includes year of birth fixed effects, controls for gender and birth order and the first ten principal components of the genetic matrix to control for population stratification (Price et al., 2006). We allow the parameters to differ across groups, as we estimate Equation 1 separately for each family disadvantage and gender group and compare the estimated genetic effects (α_j^{bf}).

While the between-family model is standard in the current literature, it is not ideal for studying gene-by-environment interaction. The main concern is that the estimated genetic effects could be biased due to omitted environmental influences, such as genetic nurture effects (Kong et al., 2018; Young et al., 2018). This is a concern for gene-by-environment interactions, since any estimated differences in returns to the polygenic score could be driven by differences in unobserved environmental influences correlated with the polygenic score. Another concern is that between-family estimates could be influenced by population stratification, as recent evidence suggests that principal components might not be sufficient to control for stratification (Berg et al., 2019).

Within-family models solve these problems, exploiting exogenous genetic variation across siblings by controlling for family fixed effects. The fixed effect controls for any environmental influences shared by siblings and, as a result, isolate the impact of genetic differences from the environmental differences (Domingue et al., 2015). Within-family models rely on the fact that conditional on parental genotype, child genotype is randomly assigned at conception (see Conley and Fletcher, 2017). As a result, estimates from within-family models are generally interpreted as causal (Rietveld et al., 2014; Lee et al., 2018).

Formally, the within-family model can be described as

$$Y_{ifgj} = \alpha_j^{wf}(g)PGS_{ifg} + X_{ifg}\beta_j^{wf}(g) + u_{fgj} + \epsilon_{ifgj}^{wf} \tag{2}$$

where Y_{ifgj} now refers to jth human capital outcome of individual i in family f and in group g. We include u_{fgj}, which are the family fixed effects for outcome j. For the within-family models, the set of controls in X_{ifg} include year of birth fixed effects, controls for gender, and birth order fixed effects. We allow the parameters to vary across groups (g), as we estimate Equation 2 separately for each disadvantage group. Additionally, to test for gender

differences, in the supplemental materials, we estimate Equation 2 including only siblings of the same gender.

The drawback of the within-family design is that it excludes individuals without siblings in the sample. Only about 5% of the individuals in our sample can be part of the within-family design. This is a concern as the reduced sample size decreases the power to estimate differences across groups. Another possible concern is that our model assumes away any interactions between siblings. Sibling-peer effects or competition for resources would cause within-family and between-family to differ even in the absence of omitted environmental effects.

4 Findings

In this section, we report and discuss our main findings. We start by describing the estimated effects of the EA PGS on human capital outcomes in Denmark. We show that the EA PGS explains 7-9% of the variation in education and school achievement in Denmark. We also show that the estimated effects decrease substantially once we control for family characteristics and for the family fixed effect in the within-family specification. We then test for heterogeneity in the impact of the EA PGS by childhood environment. We show that childhood disadvantage significantly attenuates the relationship between the EA PGS and schooling outcomes and that the attenuation is relatively larger in the within-family analysis. We also test for gender differences and find that the attenuation effect of disadvantage is significantly stronger for males than females. Lastly, we test for differences in effects across psychiatric cases and controls. We find no significant difference between the two groups, suggesting that our results extend to the remaining Danish population.

4.1 EA PGS and Human Capital Formation

We start by exploring the predictive power of the polygenic score for educational attainment on schooling achievement and educational attainment in the Danish context. We report the between-family results in Panel A of Table 2. The EA PGS explains 7-9% of the variation in educational attainment and schooling outcomes in Denmark, as measured by the change in R^2 when including the polygenic score in the regression. We find that one standard deviation increase in the EA PGS is associated with 0.606 (s.e. = 0.012) additional years of education, a 12.1 (s.e. = 0.2) percentage point larger probability of attending post-secondary education, and a 6.323 (s.e. = 0.112) and 7.432 (s.e. = 0.131) higher percentile point ranking in Danish and mathematics school leaving exams at the end of lower secondary school.

When interpreting these effects, it is important to acknowledge that the estimated effects are likely attenuated due to measurement error in the polygenic score. The predictive power and accuracy of the polygenic score are dependent on the estimation error and sample size in the original GWA study. In fact, the predictive power of the polygenic score for educational attainment has increased over time as the educational attainment GWA studies included more individuals and samples (see Plomin and von Stumm (2018) for a recent discussion).

Next, we explore how the association between the EA PGS and human capital changes after we control for family characteristics associated by family disadvantage. As shown in Figure 1, the distribution of the EA PGS is different across family environments. As a result, the early childhood environment experienced by the individual, which also influences human capital formation, could mediate the associations we found. To test for this possibility, we re-estimate the regressions in Panel A in Table 2 controlling for measures of childhood environment. We control for average log disposable family income between ages 1 to 10, maternal and paternal years of education, maternal and paternal mental health and family structure between the ages 0 to 10. We report these between-family results in Panel B of Table 2.

We find that childhood environment partially mediates the association between the educational attainment polygenic score and human capital outcomes. After controlling for childhood environment, we find that one standard deviation increase in the EA PGS is associated with 0.380 (s.e. = 0.012) additional years of education, a 7.6 (s.e. = 0.3) percentage point increase in the probability of attending a post-secondary education, and a 4.533 (s.e. = 0.112) and 5.421 (s.e. = 0.132) percentile point increase in Danish and mathematics test scores.

These results suggest that unobserved dimensions of childhood environment could to some extent explain the association between the educational polygenic score and human capital formation. As discussed in the estimation model section, within-family models can control for these unobserved environmental effects. As a result, we re-do our initial analysis in a subset of our sample composed of siblings with the same biological parents. That is, we re-estimate the regressions in Panel A in Table 2 controlling for family fixed effects, gender, and birth order. Controlling for family fixed effects is akin to relating differences in sibling polygenic scores to differences in sibling outcomes. We report between-family results for the sibling-sample in Panel A, the model controlling for observable childhood environment in Panel B, and within-family results based on a fixed effect specification in Panel C of Table 3.

First, we find that between-family estimates are comparable in the siblings-sample (Panel

Table 2: Full Sample: EA PGS and Human Capital Formation

Dep. Var.	(1) Y. Edu.	(2) Any P.S.E.	(3) Danish	(4) Math.
Panel A:				
EA PGS	0.606	0.121	6.323	7.432
	(0.012)	(0.002)	(0.112)	(0.131)
Family Controls	(N)	(N)	(N)	(N)
Family F.E.	(N)	(N)	(N)	(N)
R^2	0.103	0.101	0.151	0.097
Incr. R^2 EA PGS	0.074	0.069	0.079	0.084
Panel B:				
EA PGS	0.380	0.076	4.533	5.421
	(0.012)	(0.003)	(0.112)	(0.132)
Family Controls	(Y)	(Y)	(Y)	(Y)
Family F.E.	(N)	(N)	(N)	(N)
N	28,331	28,331	33,286	32,528

Notes: This table reports parameter estimates from between-family regressions used to link the polygenic score for educational attainment to human capital outcomes. To test the effect of the EA PGS, we regress each human capital measure on the polygenic score together with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. Family controls in Panel B include the average log disposable family income between ages 1 to 10, maternal and paternal years of education, maternal and paternal mental health history and family structure between the ages 0 to 10. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.

A of Table 3) and for the whole sample (Panel A of Table 2). Second, we find that the EA PGS predicts higher educational outcomes even after we control for family fixed effects. Using the sibling sample, we find that one standard deviation increase in the EA PGS increases educational attainment by 0.296 (s.e. = 0.094) additional years, the probability of attending post-secondary education by an additional 6.9 (s.e. = 2.0) percentage points, and Danish and mathematics tests by 2.774 (s.e. = 0.842) and 3.616 (s.e. = 0.982) percentile points. However, these effects are about half the size of the effects in the between-family estimates without family controls. These results support the contention that the polygenic score for educational attainment confounds information about children's environment that is important for education. These results highlight the importance of replicating the results in a within-family analysis.

Table 3: Sibling Sample: EA PGS and Human Capital Formation

Dep. Var.	(1) Y. Edu.	(2) Any P.S.E.	(3) Danish	(4) Math.
Panel A:				
EA PGS	0.561	0.114	6.248	6.722
	(0.053)	(0.010)	(0.469)	(0.558)
Family Controls	(N)	(N)	(N)	(N)
Family F.E.	(N)	(N)	(N)	(N)
R^2	0.123	0.123	0.179	0.103
Incr. R^2 EA PGS	0.070	0.070	0.077	0.072
Panel B:				
EA PGS	0.352	0.073	4.542	4.780
	(0.055)	(0.011)	(0.491)	(0.570)
Family Controls	(Y)	(Y)	(Y)	(Y)
Family F.E.	(N)	(N)	(N)	(N)
Panel C:				
EA PGS	0.296	0.069	2.774	3.616
	(0.094)	(0.020)	(0.842)	(0.982)
Family Controls	(N)	(N)	(N)	(N)
Family F.E.	(Y)	(Y)	(Y)	(Y)
N	1,487	1,487	1,838	1,793

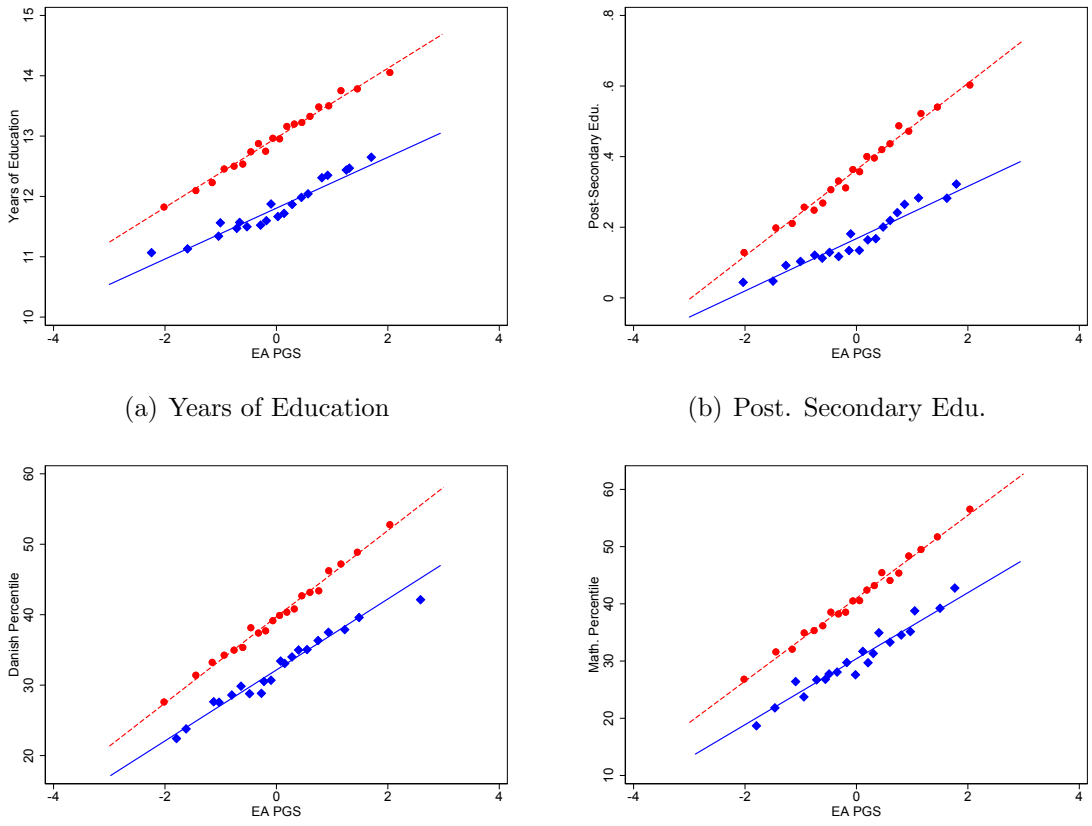
Notes: This table reports parameter estimates from regressions used to link the polygenic score for educational attainment to human capital outcomes for individuals in the sibling sample. Panel A reports between-family results with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. Family controls in Panel B include the average log disposable family income between ages 1 to 10, maternal and paternal years of education, maternal and paternal mental health history and family structure between the ages 0 to 10. Panel C reports within-family results, where we control for family fixed effects. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.

4.2 Attenuation Effect of Family Disadvantage

In the previous section, we show that the EA PGS influences different educational outcomes. In this section, we ask whether childhood disadvantage attenuates the effect of the EA PGS on these outcomes. We do so by regressing the educational attainment polygenic score on each outcome separately for individuals who experienced two or more dimensions of childhood

disadvantage and for individuals who did not and comparing the estimated effects across the two groups. We report our findings for the between family model in Panel A of Table 4. These results are also represented graphically in Figure 2.

We find a significant attenuation effect of childhood disadvantage on the relationship between the EA PGS and all four educational outcomes. We find that one standard deviation increase in the EA PGS is associated with 0.575 (s.e. = 0.014) additional years of education for the group that did not experience disadvantage but with 0.408 (s.e. = 0.023) additional years of education for the disadvantaged group. The 29% decrease in returns is statistically significant with a p-value below 0.001. We plot the relationship between the EA PGS and years of education for the two groups in Panel A of Figure 2.



(a) Years of Education

(b) Post. Secondary Edu.

(c) Danish Score

(d) Mathematics Score

Figure 2: Full Sample - EA PGS and Human Capital: Figure 2 plots the relationship between the educational attainment polygenic score and the four human capital outcomes separately for the disadvantage and non-disadvantage groups. Each dot corresponds to the mean value of the outcome and explanatory variable for different percentile. Blue dots and lines correspond to the disadvantage group and red dots and lines to the non-disadvantage group. Each dot corresponds to the average outcome value across EA PGS 5-percentile bins for each group.

Table 4: Attenuation Effect of Childhood Disadvantage

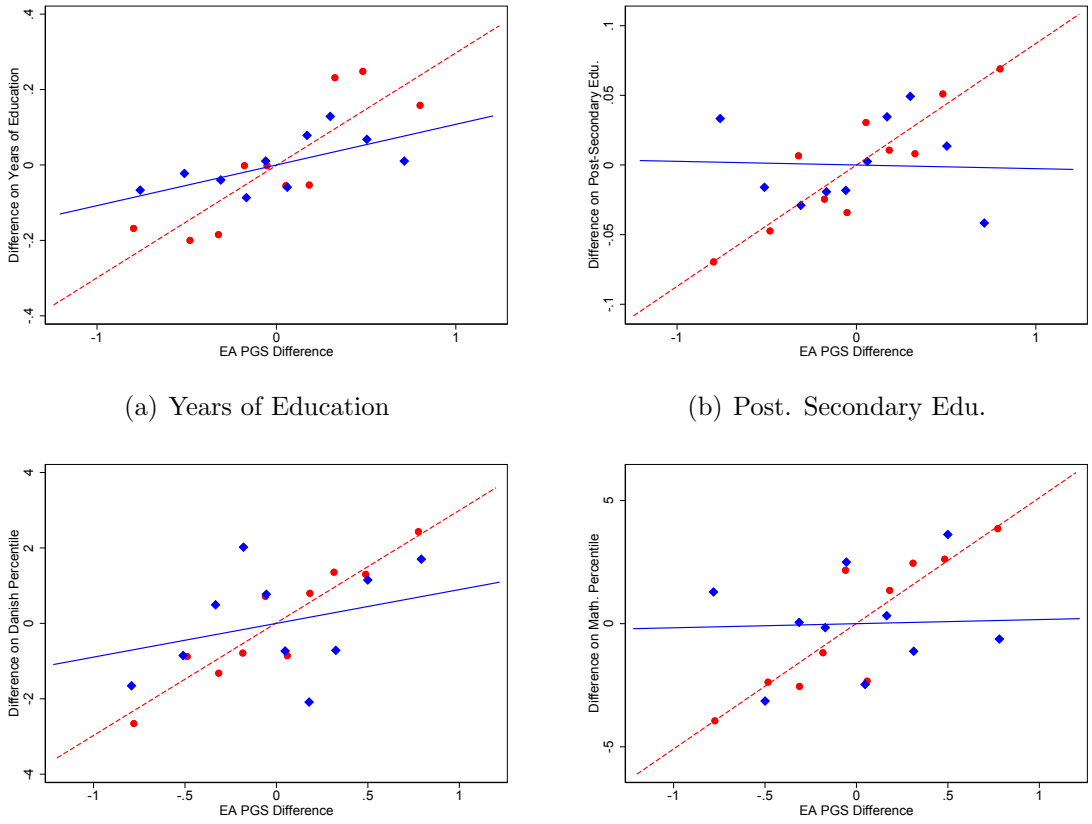
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Years Edu.		Post Sec. Edu.		Danish		Mathematics	
Group:	[No Dis]	[Disad.]	[No Dis]	[Disad.]	[No Dis]	[Disad.]	[No Dis]	[Disad.]
Panel A:								
EA PGS	0.575	0.408	0.122	0.072	6.166	5.028	7.197	5.799
	(0.014)	(0.023)	(0.003)	(0.004)	(0.126)	(0.240)	(0.148)	(0.273)
Group Differences	-0.167		-0.050		-1.138		-1.398	
	(0.027)		(0.005)		(0.271)		(0.311)	
	[0.000]		[0.000]		[0.000]		[0.000]	
Family F.E.	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)
N	21,064	7,267	21,064	7,267	26,381	6,905	25,913	6,615
Panel B:								
EA PGS	0.358	0.181	0.098	0.009	3.589	0.918	5.159	-0.091
	(0.119)	(0.154)	(0.026)	(0.030)	(0.906)	(1.925)	(1.105)	(2.034)
Group Differences	-0.177		-0.089		-2.671		-5.158	
	(0.194)		(0.040)		(2.120)		(2.308)	
	[0.362]		[0.027]		[0.208]		[0.023]	
Family F.E.	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)
N	963	524	963	524	1,400	438	1,357	436

Notes: This table reports parameter estimates from regressions used to test for context-dependence of genetic influences on human capital formation. Panel A reports between-family results for the full sample, with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. Panel B reports within-family results for the sibling sample, where we control for family fixed effects. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses and p-values in brackets.

Similarly, we find that one standard deviation increase in the EA PGS is associated with a 12.2 (s.e. = 0.3) percentage point increase in the probability of attending post-secondary education for the non-disadvantaged group but with a 7.2 (s.e. = 0.4) percentage point increase for the disadvantaged group. The 41% decrease in returns is statistically significant with a p-value below 0.001. We plot the relationship between the EA PGS and any post-secondary education for the two groups in Panel B of Figure 2.

We find similar attenuation effects on school achievement. For the non-disadvantaged group, we find that one standard deviation increase in the EA PGS is associated with a 6.166 (s.e. = 0.126) and 7.197 (s.e. = 0.148) percentile point increase in Danish and mathematics

test scores. For the disadvantaged group, the effects are smaller; namely 5.028 (s.e. = 0.240) and 5.799 (s.e. = 0.273), respectively. That is, childhood disadvantage attenuates the effect of the EA PGS by 17 and 19 percent, respectively. The group differences in the effects are statistically different with a p-value below 0.001. We plot the relationship between the EA PGS and school achievement for the two disadvantaged groups in Panel C and D of Figure 2.



(a) Years of Education

(b) Post. Secondary Edu.

(c) Danish Score

(d) Mathematics Score

Figure 3: Sibling Differences - EA PGS and Human Capital: Figure 3 plots the relationship between sibling differences in the educational attainment polygenic score and sibling differences in the four human capital outcomes separately for the disadvantage and non-disadvantage groups. Each dot corresponds to the mean value of the outcome and explanatory variable for different percentile. Blue dots and lines correspond to the disadvantage group and red dots and lines to the non-disadvantage group. Each dot corresponds to the average outcome value across EA PGS 10-percentile bins for each group.

For these analyses, we define the disadvantaged group as individuals who experienced two or more dimensions of childhood disadvantage. Since this is an arbitrary assumption, we tested for the robustness of our findings to other specifications in the supplemental materials.

In Table S1 and Figures S1 and S2, we distinguish individuals according to the number of dimensions of disadvantage experienced during childhood. We show that the effect of the EA PGS decreases with each additional dimension of disadvantage experienced during childhood. Similarly, in Table S2, we report attenuation effects when considering each dimension of childhood disadvantage individually. Parental education seems to have the most substantial impact, but all other measures of childhood disadvantage are related to lower returns of the EA PGS on human capital formation.

One worry is that the observed differences in the returns to the polygenic score could be explained by unobserved environmental effects correlated with the genetic score. To test this hypothesis, we re-do our initial analysis using the sibling sub-sample and controlling for family fixed effects. We estimate the within-family impact of the EA PGS for families in the two disadvantaged groups and report results from this exercise in Panel B of Table 4. The results can also be seen graphically in Figure 3, where we plot the relationship between sibling differences in the EA PGS and sibling differences in educational outcomes for siblings in the two disadvantaged groups.

We find similar attenuation effects of childhood disadvantage after controlling for family fixed effects. As a result of the smaller sample size, the within-family estimates are less precise than between-family estimates. For the non-disadvantaged group, we find that one standard deviation increase in the EA PGS increases educational attainment by 0.358 (s.e. = 0.119) additional years. It also increases the probability of attending post-secondary education by an additional 9.8 (s.e. = 2.6) percentage points, and Danish and mathematics tests by 3.589 (s.e. = 0.906) and 5.159 (s.e. = 1.105) percentile points. Whereas, for the disadvantaged group, the effects are 0.181 (s.e. = 0.154), 0.9 (s.e. = 3.0), 0.918 (s.e. = 1.925), -0.091 (s.e. = 2.034), respectively. The difference in effects on post-secondary educational attendance and mathematics test scores are statistically significant at the 5% confidence level.

4.3 Gender Differences in the Attenuation Effect of Family Disadvantage

Following results from an emerging literature in economics that study gender differences in response to family disadvantage (Bertrand and Pan, 2013; Autor et al., 2019; Lundberg, 2017), we test for gender differences in the context-dependence of genetic influences on educational outcomes by childhood disadvantage. We do so by re-estimating the between-family model for each disadvantage group and gender and testing for differences in the effect of the EA PGS across specifications. We report our findings in Table 5. The gender differences can also be seen graphically for years of education and any post secondary education in Figure 4,

where we plot the relationship between the EA PGS and educational attainment separately across gender and the two disadvantaged groups.

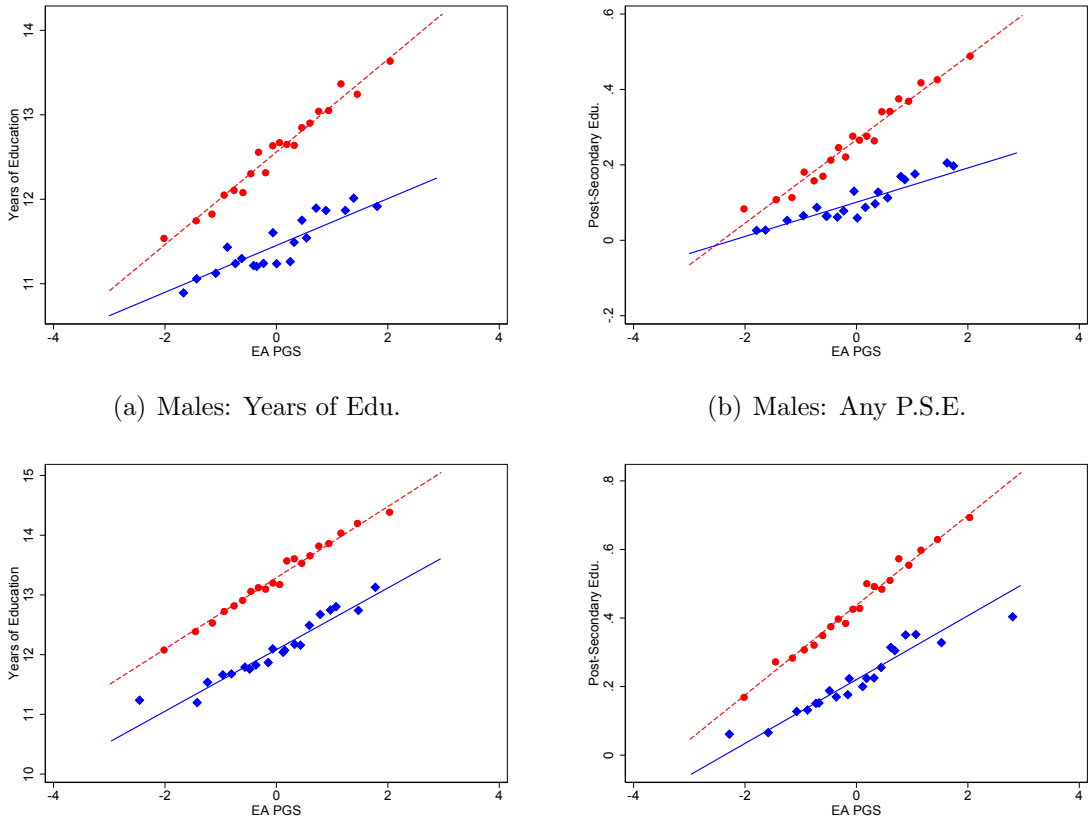
Table 5: Gender Differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Years Edu.		Post Sec. Edu.		Danish		Mathematics	
Gender Group:	[M]	[F]	[M]	[F]	[M]	[F]	[M]	[F]
EA PGS	0.549	0.596	0.111	0.131	5.597	6.681	6.933	7.419
	(0.021)	(0.019)	(0.004)	(0.004)	(0.179)	(0.177)	(0.219)	(0.202)
EA PGS × Disad.	-0.262	-0.091	-0.064	-0.039	-1.400	-0.988	-1.604	-1.221
	(0.038)	(0.038)	(0.007)	(0.008)	(0.384)	(0.380)	(0.467)	(0.415)
Gender Differences	-0.171		-0.025		-0.412		-0.383	
	(0.054)		(0.010)		(0.541)		(0.625)	
	[0.001]		[0.014]		[0.447]		[0.540]	
Family F.E.	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)
N	12,470	15,861	12,470	15,861	16,075	17,211	15,818	16,710

Notes: This table reports parameter estimates from regressions used to test for gender differences in the context-dependence of genetic influences. The reported coefficients are between-family results for the full sample with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses and p-values in brackets.

We find significant gender differences in the attenuation effect of childhood disadvantage on the returns of the EA PGS for educational attainment. For males, childhood disadvantage reduces the association between the EA PGS and years of education by 0.262 (s.e. = 0.039) years from 0.549 (s.e. = 0.021). That is, childhood disadvantage reduces the association by a half. The attenuation effects are much milder for females, where the association decreases by 0.091 (s.e. = 0.038) years from 0.596 (s.e. = 0.019). The gender differences in the gene-by-environment interaction are are statistically significant at the 1% confidence level. We observe similar gender differences when considering post-secondary education attendance. For males, the association between the EA PGS and post-secondary education attendance in the childhood-disadvantaged group is 4.7 percentage points, 6.4 percentage points smaller than in the non-disadvantaged group. For females, the association in the disadvantaged group is 9.2 percentage points, 3.9 percentage points lower than in the non-disadvantaged group. Again, the gender differences in the gene-by-environment interaction are statistically significant at the 5% confidence level.

We observe similar patterns for school leaving exam rankings, but gender differences are



(a) Males: Years of Edu.

(b) Males: Any P.S.E.

(c) Females: Years of Edu.

(d) Females: Any P.S.E.

Figure 4: Gender Differences - EA PGS and Human Capital: Figure 4 plots the relationship between the educational attainment polygenic score, years of education and the probability of attending a post-secondary program separately for the disadvantage and non-disadvantage groups, and separately for males and females. Each dot corresponds to the mean value of the outcome and explanatory variable for different percentile. Blue dots and lines correspond to the disadvantage group and red dots and lines to the non-disadvantage group. Each dot corresponds to the average outcome value across EA PGS 5-percentile bins for each group.

not statistically significant. That is, we find that the associations between the EA PGS and Danish and mathematics test scores are significantly lower in the disadvantaged group for both gender groups. Again, the decrease in the association is more substantial for males than for females, but the difference is not statistically significant (p=0.432 for Danish test scores and p=0.629) for mathematics test scores.

In the supplemental materials, we test for gender differences in a within-family specification. To do so, we re-estimate the within-family model in families where all observed siblings are male and families where all observed siblings are female. The findings from these analyses are reported in Panels B and C of Table S3. Consistent with the between-family results

we find that the attenuation effects of childhood disadvantage on educational attainment are larger for males than for females. However, when considering schooling test scores, if anything, the attenuation effect is larger for females than for males. Unfortunately, given the small sample sizes, we do not have the power to statistically distinguish most of these effects.

4.4 Attenuation Effect of Family Disadvantage in a Representative Sample

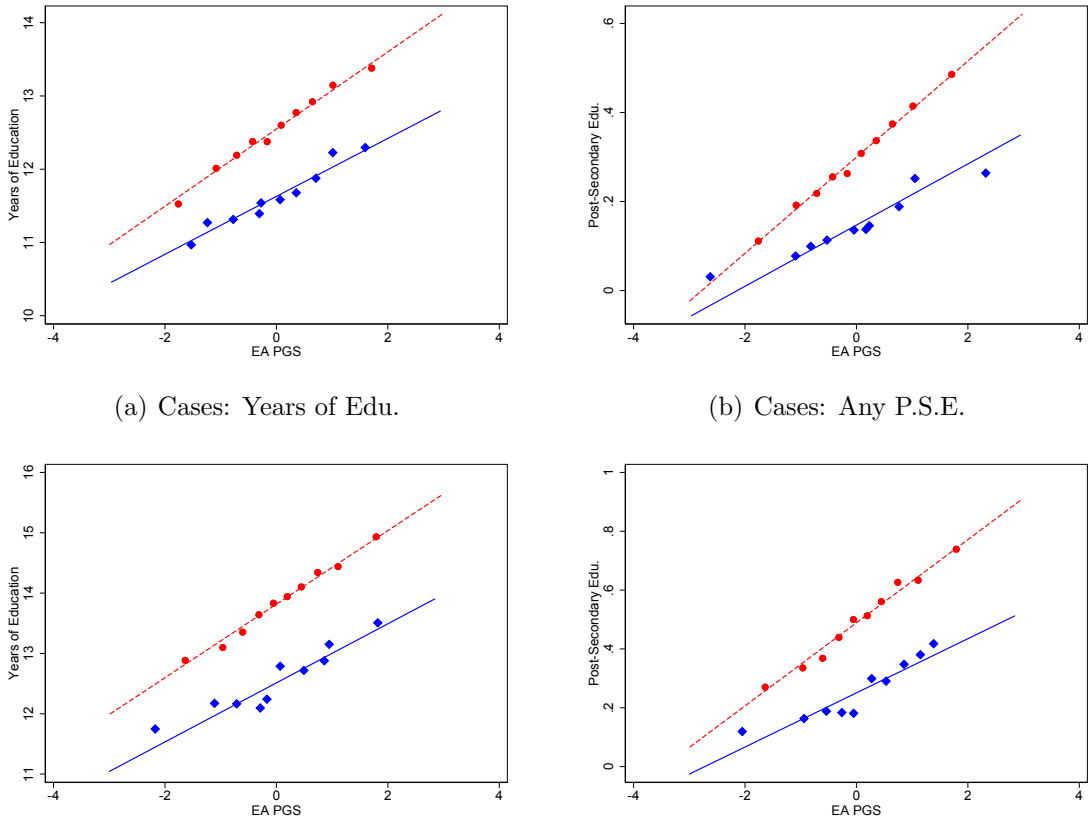
Table 6: Psychiatric Cases and Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Years Edu.		Post Sec. Edu.		Danish		Mathematics	
Group:	[Cases]	[Cont.]	[Cases]	[Cont.]	[Cases]	[Cont.]	[Cases]	[Cont.]
EA PGS	0.522	0.608	0.107	0.141	6.046	6.323	7.141	7.356
	(0.017)	(0.022)	(0.004)	(0.005)	(0.163)	(0.197)	(0.193)	(0.231)
EA PGS × Disad.	-0.138	-0.132	-0.040	-0.052	-0.955	-1.618	-1.642	-0.473
	(0.057)	(0.030)	(0.012)	(0.006)	(0.512)	(0.325)	(0.589)	(0.373)
Group Differences	-0.006		0.012		0.664		-1.169	
	(0.064)		(0.013)		(0.606)		(0.697)	
	[0.929]		[0.356]		[0.274]		[0.094]	
Family F.E.	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)
N	19,890	8,441	19,890	8,441	20,694	12,592	20,093	12,435

Notes: This table reports parameter estimates from regressions used to test for differences in the context-dependence of genetic influences across psychiatric cases and controls. The reported coefficients are between-family results for the full sample with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported parentheses and p-values in brackets.

Our main estimates include both psychiatric cases and controls in iPSYCH study. One possible concern, since the study over-samples psychiatric cases, is that there could be substantial differences between the genetic profile and the genetic effects in our sample and the remaining Danish population. For example, there could be important interactions between the genetic predisposition for mental illness and response to childhood adversity.

Given this possibility, we re-estimate the between-family model for psychiatric cases and controls separately. Since controls were randomly selected from the Danish 1981-2005 cohorts, we can test for the external validity of our results by comparing the estimates across the two-groups. We report our findings in Table 6. The case-control differences can also be



(a) Cases: Years of Edu.

(b) Cases: Any P.S.E.

(c) Controls: Years of Edu.

(d) Controls: Any P.S.E.

Figure 5: Psychiatric Cases and Controls - EA PGS and Human Capital: Figure 5 plots the relationship between the educational attainment polygenic score, years of education and the probability of attending a post-secondary program separately for the disadvantage and non-disadvantage groups, and separately for psychiatric cases and controls. Each dot corresponds to the mean value of the outcome and explanatory variable for different percentile. Blue dots and lines correspond to the disadvantage group and red dots and lines to the non-disadvantage group. Each dot corresponds to the average outcome value across EA PGS 10-percentile bins for each group.

seen graphically for years of education and any post secondary education in Figure 5, where we plot the relationship between the EA PGS and educational attainment separately for psychiatric cases and the control sample and across the two disadvantaged groups.

We find no significant differences across the two groups. The attenuation effect of childhood disadvantage on the returns of the EA PGS for educational attainment is similar in the two groups. The only group difference that is significant at the 10% significance level is on mathematics percentile scores. These results are reassuring since they show that our main findings are not driven by the iPSYCH study design.

5 Discussion

We use detailed Danish registry data combined with individual level genotypic data for a large number of Danish individuals born between 1981 to 2005 to show that a polygenic score for educational attainment explains 7-9% of the variation in educational attainment in Denmark. The large-scale nature of the data also means we are able to replicate our main findings in a smaller within-family analysis. We show that the causal effect of the polygenic scores and educational attainment is about 40-50% lower than the estimated effects using the full sample. The decrease in association in the within-family analysis is consistent with findings in Kong et al. (2018), which shows that the effects of non-transmitted parental alleles on child outcomes are about a third of that of the transmitted alleles. These results support the idea that the estimated coefficients on polygenic scores in between-family models capture both genetic and environmental influences, such as genetic nurturing. This finding emphasizes the crucial importance of within-family analyses for causal interpretations of genetic influences.

We show that the genetic effects on education are substantially attenuated by family disadvantage in Denmark, a country which is known for its comprehensive welfare state. We find that childhood disadvantage significantly attenuates the relationship between the polygenic score for education and educational outcomes and that the attenuation effect, if anything, is stronger in the within-family analysis that controls for unobserved environmental influences. These results contrast with previous findings from the behavioral genetic literature in European countries (Tucker-Drob and Bates, 2016).The attenuation effect of childhood disadvantage is particularly pronounced for boys. We also find no differences in the attenuation effect of childhood disadvantage across iPSYCH cases (individuals diagnosed with schizophrenia, ADHD, or other iPSYCH disorder) and controls (a representative sample of the Danish 1981-2005 cohorts).

A possible mechanism explaining the attenuation effect is a difference across families in parents' ability to provide high-quality interactions and a rich learning environment for their children. Interactions between young children and adults which are necessary for strengthening developing brain architecture and thereby development in general which are foundational to later educational attainment (Shonkoff et al., 2016), and the quality of these interactions is negatively affected by low SES (Hoff et al., 2002).

The Danish - and more generally, the Nordic - welfare-state specifically aims at alleviating childhood disadvantage by providing universal services to young children to supplement the home environment. The fact that most Danish children spend a large part of their day in high-quality public institutions, daycare, school, or educational institutions, was believed to

reduce the effects of family disadvantage. Hence, it is essential to answer the question of why the Danish welfare-state is less successful in this dimension than anticipated, especially when it comes to boys. Our findings complement those from three recent studies that explore this question. Landersø and Heckman (2017) find that, despite all the resources devoted overcoming this disadvantage, intergenerational mobility in education in Denmark is almost as low as in the U.S., pointing to an apparent policy failure. Rosholm et al. (2019) argue that this may to some extent be caused by a lacking evaluation culture in public institutions, such that instated policies are profoundly ineffective in helping disadvantaged children. In fact, this contention is supported by recent research showing that the interaction quality in Danish daycares is low and that public services do not meet disadvantaged children's need of high-quality interactions, thereby failing to help disadvantaged children to realize their full potentials (Slot et al., 2018).

Our findings also complement a recent literature that documents that boys are more sensitive to early life environment than girls. Bertrand and Pan (2013) demonstrate that parental inputs are lower in disadvantaged (in their case; broken) families and that this is particularly detrimental to the socio-emotional development of boys. Several studies have documented the crucial role played by socio-emotional skills in educational attainment and boys' early deficit regarding such skills, relative to girls (Almlund et al., 2011). These skills are shown in the literature to be highly malleable in children and results in Bertrand and Pan (2013) show that they are primarily affected by family inputs, while schools and daycare play a smaller role. They also note that some authors have pointed to boys being unable to reap the full benefits from school and daycare due to the more regimented nature of schools, which presumably favors girls, but they are not able to confirm this hypothesis. Our results suggest that the educational gap of disadvantaged children is to some extent driven by boys with a high genetic propensity for education, who appear more vulnerable to childhood disadvantage than similar girls. If socio-emotional skills and the genetic disposition for schooling are complementary inputs to skill formation during crucial stages of childhood development, as research by Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010) suggests, then a high genetic disposition for education will not be enough to overcome the disadvantage induced lower parental investments in socio-emotional skills.

References

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, vol. 4. Elsevier, 1–181.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics* .

Barcellos, Silvia H, Leandro S Carvalho, and Patrick Turley. 2018. "Education Can Reduce Health Differences Related to Genetic Risk of Obesity." *Proceedings of the National Academy of Sciences* 115 (42):E9765–E9772.

Barth, Daniel, Nicholas W Papageorge, and Kevin Thom. 2020. "Genetic Ability, Wealth, and Financial Decision-Making." *Journal of Political Economy* (forthcoming).

Belsky, Daniel W, Benjamin W Domingue, Robbee Wedow, Louise Arseneault, Jason D Boardman, Avshalom Caspi, Dalton Conley, Jason M Fletcher, Jeremy Freese, Pamela Herd et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31):E7275–E7284.

Belsky, Daniel W, Terrie E Moffitt, David L Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S Williams et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-course Development." *Psychological Science* 27 (7):957–972.

Berg, Jeremy J, Arbel Harpak, Nasa Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan August Boyle, Xinjun Zhang, Fernando Racimo, Jonathan K Pritchard et al. 2019. "Reduced Signal for Polygenic Adaptation of Height in UK Biobank." *Elife* 8:e39725.

Bertrand, Marianne and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5 (1):32–64.

Bulik-Sullivan, Brendan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, Laramie Duncan, John RB Perry, Nick Patterson, Elise B Robinson et al. 2015. "An Atlas of Genetic Correlations Across Human Diseases and Traits." *Nature Genetics* 47 (11):1236.

Conley, Dalton and Jason Fletcher. 2017. *The Genome Factor: What the Social Genomics Revolution Reveals About Ourselves, Our History, and the Future*. Princeton University Press.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2):191–261.

Cunha, Flavio and James J Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (4):738–782.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3):883–931.

Domingue, Benjamin W, Daniel W Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D Boardman. 2015. "Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health." *AERA Open* 1 (3):2332858415599972.

Grove, Jakob, Stephan Ripke, Thomas Damm Als, Manuel Mattheisen, Raymond Walters, Hyejung Won, Jonatan Pallesen, Esben Agerbo, Ole A Andreassen, Richard Anney et al. 2017. "Common Risk Variants Identified in Autism Spectrum Disorder." *BioRxiv* :224774.

Heckman, James J and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6 (1):689–733.

Hoff, Erika, Brett Laursen, Twila Tardif, and M Bornstein. 2002. "Socioeconomic Status and Parenting." *Handbook of Parenting Volume 2: Biology and Ecology of Parenting* 8 (2):231–52.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374):424–428.

Landersø, Rasmus and James J Heckman. 2017. "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US." *The Scandinavian Journal of Economics* 119 (1):178–230.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, M Johannesson, PD Koellinger, P Turley, PM Visscher et al. 2018. "Gene Discovery and Polygenic Prediction from a 1.1-Million-Person GWAS of Educational Attainment." *Nature Genetics* .

Lundberg, Shelly. 2013. "Educational Inequality and the Returns to Skills." *IZA Discussion Paper Series* .

Lundberg, Shelly J. 2017. "Father Absence and the Educational Gender Gap." *IZA Discussion Paper Series* .

Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R Boyko, Adam Auton, Amit Indap, Karen S King, Sven Bergmann, Matthew R Nelson et al. 2008. "Genes Mirror Geography Within Europe." *Nature* 456 (7218):98.

Okbay, Aysu, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens et al. 2016. "Genome-wide Association study identifies 74 loci associated with educational attainment." *Nature* 533 (7604):539.

Papageorge, Nicholas W, Victor Ronda, and Yu Zheng. 2019. "The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market." *National Bureau of Economic Research* .

Papageorge, Nicholas W and Kevin Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* (forthcoming).

Pedersen, Carsten Boecker, Jonas Bybjerg-Grauholm, Marianne Gioertz Pedersen, Jakob Grove, Esben Agerbo, Marie Baekvad-Hansen, Jesper Buchhave Poulsen, Christine Soeholm Hansen, JJ McGrath, Thomas D Als et al. 2018. "The iPSYCH2012 Case–Cohort Sample: New Directions for Unravelling Genetic and Environmental Architectures of Severe Mental Disorders." *Molecular Psychiatry* 23 (1):6.

Pirastu, Nicola, Mattia Cordioli, Priyanka Nandakumar, Gianmarco Mignogna, Abdel Abdellaoui, Benjamimin Hollis, Masahiro Kanai, Veera M. Rajagopal, Pietro Della et al. 2020. "Genetic analyses identify widespread sex-differential participation bias." *https://www.biorxiv.org/content/10.1101/2020.03.22.001453v1.full.pdf* .

Plomin, Robert and Sophie von Stumm. 2018. "The New Genetics of Intelligence." *Nature Reviews Genetics* 19 (3):148.

Price, Alkes L, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38 (8):904.

Rietveld, Cornelius A, Dalton Conley, Nicholas Eriksson, Tõnu Esko, Sarah E Medland, Anna AE Vinkhuyzen, Jian Yang, Jason D Boardman, Christopher F Chabris, Christopher T Dawes et al. 2014. "Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits." *Psychological Science* 25 (11):1975–1986.

Rietveld, Cornelius A, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal et al. 2013. "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment." *Science* :1235488.

Rosholm, Michael, Alexander Paul, Dorthe Bleses, Philip Dale, Anders Højen, Peter Jense, Laura Justice, and Simon Calmar Andersen. 2019. "Early Interventions Make a Difference But Contents Matter Even in the Scandinavian Welfare State." *Working Paper* .

Shonkoff, JP, J Richmond, P Levitt, SA Bunge, JL Cameron, GJ Duncan, and CA Nelson III. 2016. "From Best Practices to Breakthrough Impacts a Science-Based Approach to Building a More Promising Future for Young Children and Families." *Cambirdge, MA: Harvard University, Center on the Developing Child* .

Slot, Pauline Louise, Dorthe Bleses, Laura M Justice, Justin Markussen-Brown, and Anders Højen. 2018. "Structural and Process Quality of Danish Preschools: Direct and Indirect Associations with Childrens Growth in Language and Preliteracy Skills." *Early Education and Development* 29 (4):581–602.

Trejo, Sam, Daniel W Belsky, Jason D Boardman, Jeremy Freese, Kathleen Mullan Harris, Pam Herd, Kamil Sicinski, and Benjamin W Domingue. 2018. "Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Heath." *Sociological Science* 5:513–540.

Tucker-Drob, Elliot M and Timothy C Bates. 2016. "Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence." *Psychological Science* 27 (2):138–149.

Ward, Mary E, George McMahon, Beate St Pourcain, David M Evans, Cornelius A Rietveld, Daniel J Benjamin, Philipp D Koellinger, David Cesarini, George Davey Smith, Nicholas J Timpson et al. 2014. "Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children." *PloS One* 9 (7):e100248.

Young, Alexander I, Michael L Frigge, Daniel F Gudbjartsson, Gudmar Thorleifsson, Gyda Bjornsdottir, Patrick Sulem, Gisli Masson, Unnur Thorsteinsdottir, Kari Stefansson, and Augustine Kong. 2018. "Relatedness Disequilibrium Regression Estimates Heritability Without Environmental Bias." *Nature Genetics* 50 (9):1304.

Supplemental Material

"Family Disadvantage, Gender and the Returns to Genetic Human Capital"

Appendix A Alternative Definitions of Disadvantage

In our benchmark specification, we classify individuals in the disadvantaged group if they experienced two or more dimensions of childhood disadvantage. We recognize that this is an arbitrary classification. In this section, we test the robustness of our main findings when we use different classifications of childhood disadvantage.

First, we separate individuals into four groups according to how many dimensions of childhood disadvantage each experienced. Figure S1 plots the distribution of the EA PGS for each of the disadvantaged groups. As shown in Figure 1, worse childhood environments are associated with lower polygenic scores for educational attainment. Table S1 estimates the between-family model while allowing the returns of the EA PGS to differ according to the number of experienced dimensions of childhood disadvantage. We show that the returns of the EA PGS decreases with each additional dimension of disadvantage experienced during childhood. These results can also be seen in Figure S2.

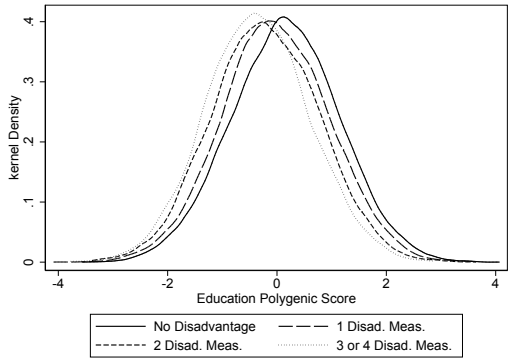


Figure S1: EA PGS by Childhood Disadvantage Groups: Figure S1 shows the estimated distribution of the educational attainment polygenic score across childhood disadvantage groups.

In addition, we also present results where we classify individuals in the disadvantaged group considering each dimension of disadvantage individually. We report these results in Table S2. We find that parental education seems to have the most substantial impact on the returns of the EA PGS, but all other measures of childhood disadvantage are related to lower returns of the EA PGS on human capital formation.

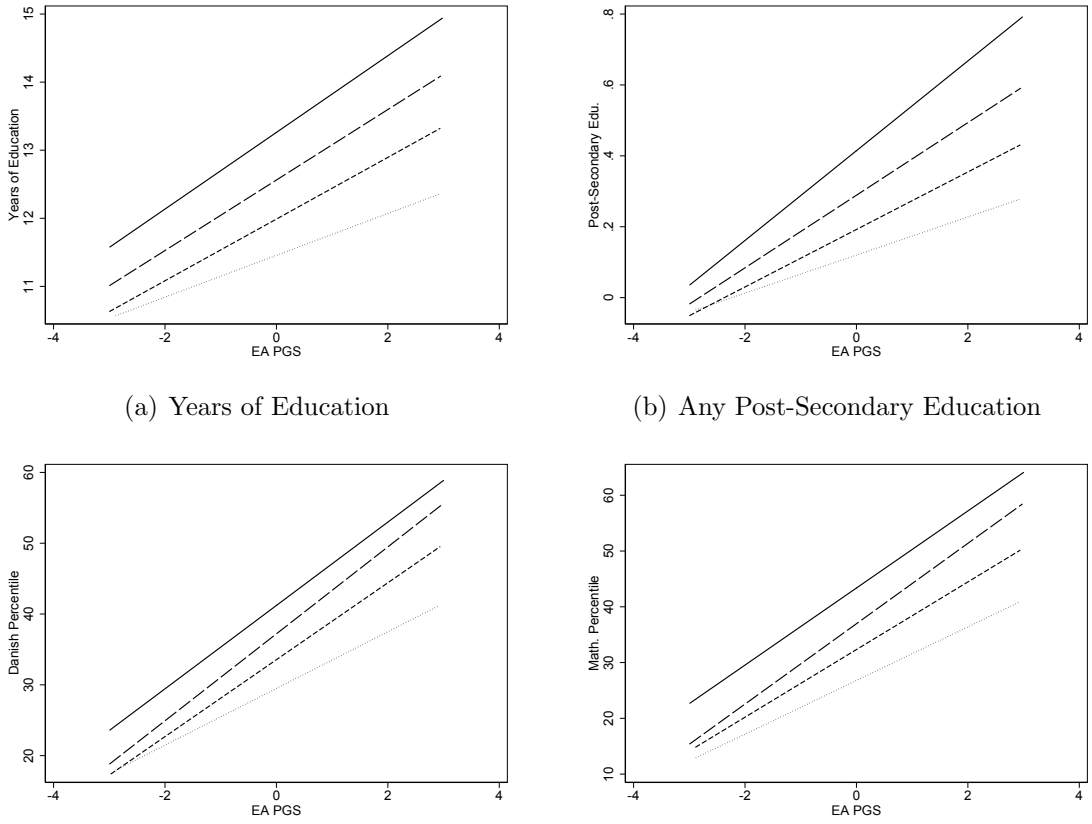
Dep. Var.	(1) Y. Edu.	(2) Any P.S.E.	(3) Danish	(4) Math.
EA PGS	0.562	0.126	5.927	6.901
	(0.019)	(0.004)	(0.158)	(0.189)
EA PGS × 1 Disad. Meas.	-0.049	-0.025	0.283	0.248
	(0.029)	(0.006)	(0.262)	(0.305)
EA PGS × 2 Disad. Meas.	-0.123	-0.047	-0.558	-0.778
	(0.035)	(0.007)	(0.344)	(0.397)
EA PGS × 3 or 4 Disad. Meas.	-0.263	-0.074	-1.914	-2.115
	(0.041)	(0.008)	(0.417)	(0.474)

Table S1: Disadvantage Levels: Attenuation Effect of Childhood Disadvantage: This table reports parameter estimates from regressions used to test for context-dependence of genetic influences on human capital formation across different levels of family disadvantage. We report between-family results for the full sample, with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.

Appendix B Within-Family Results Across Sibling-Gender Groups

In the main paper, we report gender differences using a between-family specification. However, it is possible that the polygenic score captures unobserved environmental effects that are different across gender. If this is the case, the reported gender differences could be due to environmental effects and not due to differences in the returns to genetic endowments. In this Appendix, we test for gender differences in a within-family specification.

To do so, we re-estimate the within-family model in families where all observed siblings are male and families where all observed siblings are female. Results from these analyses are reported in Table S3. Panel A reports within-family results where all families with two or more children are included in the analysis, as in the main study. In addition, in Panels B and C, we report within-family results where all observed siblings are male and female, respectively. Consistent with the between-family results described in the main paper, we find that childhood disadvantage significantly attenuates the effect of the EA PGS on educational attainment for males but not for females. On the other hand, the attenuation effect of childhood disadvantage on schooling test scores is larger for females than for males. Unfortunately, the sample sizes in these analyses are very small, and we do not have the



(a) Years of Education

(b) Any Post-Secondary Education

(c) Danish Percentile Score

(d) Math Percentile Score

Figure S2: EA PGS and Educational Outcomes by Disadvantage groups: Figures 2(a)-2(d) plot the relationship between the educational attainment polygenic score and the four educational outcomes across family disadvantage groups. Solid lines (-) correspond to individuals that did not experience any dimension of childhood disadvantage, long dash lines to individuals that experienced 1 dimension of disadvantage, medium dashed lines to individuals that experienced 2 dimensions of disadvantage, and short dashed lines to individuals that experienced 3 or 4 dimensions of disadvantage.

power to distinguish most of the gender differences.

Dep. Var.	(1) Y. Edu.	(2) Any P.S.E.	(3) Danish	(4) Math.
Panel A: Disadvantage = Parents with min. education				
EA PGS	0.562	0.119	6.047	7.215
	(0.014)	(0.003)	(0.122)	(0.144)
EA PGS × Disad.	-0.217	-0.067	-1.890	-2.936
	(0.028)	(0.005)	(0.301)	(0.337)
Panel B: Disadvantage = Family Income < 20th pct.				
EA PGS	0.606	0.125	6.326	7.345
	(0.014)	(0.003)	(0.123)	(0.145)
EA PGS × Disad.	-0.121	-0.039	-1.148	-0.926
	(0.030)	(0.006)	(0.292)	(0.332)
Panel C: Disadvantage = Parents with psychiatric diagnosis				
EA PGS	0.614	0.125	6.320	7.398
	(0.014)	(0.003)	(0.126)	(0.148)
EA PGS × Disad.	-0.106	-0.027	-0.253	-0.340
	(0.029)	(0.006)	(0.276)	(0.321)
Panel D: Disadvantage = Living in a broken family ages 0-10				
EA PGS	0.598	0.124	6.235	7.218
	(0.014)	(0.003)	(0.131)	(0.154)
EA PGS × Disad.	-0.099	-0.031	-0.338	-0.227
	(0.027)	(0.005)	(0.254)	(0.292)

Table S2: Different Measures: Attenuation Effect of Childhood Disadvantage: This table reports parameter estimates from regressions used to test for context-dependence of genetic influences on human capital formation for different measures of family disadvantage. We report between-family results for the full sample, with controls for birth year, gender, birth order and the first ten principal components of the genetic matrix. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.

Dep. Var.	Y. Edu.	Any P.S.E.	Danish	Math.
Panel A: All Families				
EA PGS	0.358	0.098	3.598	5.159
	(0.119)	(0.026)	(0.906)	(1.105)
EA PGS × Disad.	-0.177	-0.089	-2.671	-5.158
	(0.194)	(0.040)	(2.120)	(2.308)
Panel B: All observed siblings are male				
EA PGS	0.608	0.084	1.905	3.545
	(0.236)	(0.039)	(1.632)	(2.154)
EA PGS × Disad.	-0.539	-0.128	-6.001	-5.250
	(0.316)	(0.050)	(4.397)	(2.308)
Panel C: All observed siblings are female				
EA PGS	0.202	0.109	5.230	5.789
	(0.209)	(0.049)	(1.667)	(1.971)
EA PGS × Disad.	0.092	-0.095	-8.840	-19.680
	(0.459)	(0.100)	(4.321)	(3.942)

Table S3: Within Family Estimates Across Sibling Groups: This table compares within-family results across families with different sibling-gender combinations. In all panels, we report within-family results, where we control for family fixed effects. In Panel A, we include all families with at least two observed siblings. In Panel B, we only include families with at least two male siblings and where all observed siblings are male. In Panel C, we only include families with at least two female siblings and where all observed siblings are female. The polygenic score was normalized to have mean zero and standard deviation one in the full sample. Standard errors are reported in parentheses.