Genes, Parental Education, and Inequalities in Human Capital: Evidence from two British Cohorts

Gabriella Conti[1]
Ping Wu

VERY PRELIMINARY AND INCOMPLETE

Abstract

Recent advances have led to the discovery of specific genetic variants that predict educational attainment. We study how these variants, summarized as a linear index—known as a polygenic score—are associated with human capital accumulation and how they interact with parental education by using novel genetic data available in the National Child Development Study (NCDS) and in the Millennium Cohort Study. We present two main sets of results. First, we show that the polygenic score is highly predictive of child development, parental investments, educational attainment, and health and socioeconomic outcomes since childhood until adulthood; the predictive power of the polygenic score for child cognitive development persists in two cohorts born decades apart, also conditional on parental genes. Second, we find evidence that the genetic factors measured by this score interact strongly with parental education in affecting human capital in the offspring. In particular, by exploiting exogenous variation in education induced by two schooling reforms announced in the Education Acts 1944 and 1962, we find that increasing parental education raised offspring human capital only for children with high genetic endowments. Our findings uncover a potential mechanism through which the interplay of genes and environments amplifies inequalities in human capital across generations.

1. Introduction

What determines a person's human capital? Decades of economic research have investigated this question, with recent work focusing on the prenatal period and the first years of life (see for example the review by Almond et al., 2018). A key factor is believed to be the human capital of one's parents: in particular, children of more educated parents tend to have better outcomes along a number of dimensions. However, while some recent evidence has shown causal impacts of maternal schooling on child and adolescent cognitive, socioemotional, and health development (Currie and Moretti, 2003; Carneiro et al., 2013; Lundborg et al., 2014), other work has shown more limited support for intergenerational spillovers of education (Black et al., 2005). The research to date has been limited, however, in its ability to investigate the interplay between parental education and offspring genes in affecting human capital development.

In this paper we leverage the new availability of genetic data in the 1958 and in the 2000 British cohorts, together with recent advances in genetics which have led to the discovery of specific genetic variants that predict educational attainment (Lee et al., 2018), which we summarize as a linear index, known as a polygenic score (PGS, Dudbridge 2013) – henceforth

[1] Corresponding Author. Department of Economics and Social Research Institute, University College London, Institute for Fiscal Studies, CEPR and IZA. Email: gabriella.conti@ucl.ac.uk.

EA3. We present two main sets of results. First, we show that the polygenic score is highly predictive of child development, parental investments, educational attainment, and health and socioeconomic outcomes since childhood until adulthood. More specifically, in the 1958 British cohort the polygenic score explains 7%-8% of the variation in test scores for mathematics and reading comprehension at ages 11 and 16, and between 6% and 10% of the variation in completed education; its predictive power is comparable to that of maternal education; additionally, individuals with a higher polygenic score enjoy greater parental investments and better socioemotional development in childhood, and higher socioeconomic status and improved health in adulthood. In the 2000 British cohort, the polygenic score explains 3-4%of the variation in child cognitive development at ages 3 and 5, and 6% of its variation at age 7. Second, we find evidence that the genetic factors measured by this score interact strongly with parental education in affecting human capital in the offspring. More specifically, by exploiting exogenous variation in parental education induced by two schooling reforms announced in the Education Acts 1944 and 1962, we find that increasing parental education raised offspring human capital only for children with high genetic endowments. Our findings uncover both a potential source of heterogeneity which might reconcile different results in the literature on the intergenerational effects of parental education, and a potential mechanism through which inequalities in human capital might be amplified across generations.

The remainder of the paper is structured as follows. Section 2 presents the new genetic data available in the 1958 British birth cohort, and the computation of the polygenic score. Section 3 presents the results on the predictive power of the polygenic score for childhood and adulthood outcomes. Section 4 presents the analysis of the interaction between parental education and the polygenic score in the 1958 cohort. Section 5 presents the genetic data, the computation of the polygenic score and its predictive power in the 2000 British cohort. Section 6 presents the analysis of the interaction between parental education and the polygenic score in the 2000 cohort. Section 7 provides some preliminary conclusions.

2. Genetic data in the 1958 birth cohort

The National Child Development Study is a longitudinal study of about 17,000 babies born in Great Britain in the week 3-9 March 1958. The study started as the "Perinatal Mortality Survey" and surveyed the economic and obstetric factors associated with stillbirth and infant mortality. Since the first wave, cohort members have been followed up in nine other occasions: at age 7 (1965), 11 (1969), 16 (1974), 23 (1981), 33 (1991), 42 (2000), 44/45 (2002/3), 50 (2008), and 55 (2013); an age-62 sweep was in the field in January 2020, and has recently been resumed after a suspension due to the coronavirus pandemic.

Blood samples were taken during the 2002/3 Biomedical Sweep, when the cohort members were 44-45 years old. They were asked for consent to collect blood, store blood, extract DNA, and culture cells. The consent rate was very high: 90% of the cohort members who participated in the biomedical sweep consented, which increases to 96% among those for whom a blood sample was taken. Consent is not associated with gender, and it is inversely associated with social class (both at birth and 42), with cohort members of a higher social class less likely to provide consent.

From the blood samples, DNA was extracted using six different arrays[2] in different labs in different years, primarily to carry out the work of the Wellcome Trust Case Control Consortium (WTCCC), of which the NCDS constituted one of the two control cohorts (the other being the National Blood Service cohort). The WTCCC was a pioneeristic consortium funded by the Wellcome Trust with the primary purpose to accelerate efforts to identify genome sequence variants influencing major causes of human morbidity and mortality (such as type 1 and type 2 diabetes, breast cancer, hypertension, multiple sclerosis), through implementation and analysis of large-scale genome wide association studies.

To use the DNA of the NCDS cohort members in a consistent way, we first had to harmonise and combine the data collected using different partially-overlapping arrays with different genome builds. For this, we performed four different steps. First, we performed Quality Control (QC) checks on each single dataset using validated procedures established within the UCLEB consortium (Shah et al., 2013). Two datasets (Affymetrix v6 and Affymetrix 500K) were found to have low quality and were discarded, without significant sample loss, since the same blood samples were often assayed using more than one array. In a second step, the remaining four datasets (Infinium HumanHap 550K v3, Infinium HumanHap 550K v1.1, Illumina Human 660-Quad, Illumina 1.2M) were harmonized and merged, and further QC checks were carried out. The harmonised data has a sample of size n=6,382 and contains 504,924 SNPs. Plotting the first two principal components of the NCDS genetic matrix against the HapMap III data confirms that the NCDS sample is of European ancestry. In a third step, the merged data was submitted to the Michigan Imputation Server for imputation using the HRC reference panel. The imputation obtained was of a satisfactory quality. In a fourth and last step, the SNPs with top imputation quality (R-square 0.8 or above) and MAF >0.01 were retained, for a total of 7,550,647 SNPs. The polygenic score was constructed using PRSice, with the first 20 principal components of the genetic matrix already included in the scoring.[3] We constructed two versions of the score, with and without pruning (Dudbridge and Newcombe, 2015), and we standardized them (Figure 1); we did not perform p-value thresholding. The correlation between the scores with and without pruning is 0.725 (Figure 2).

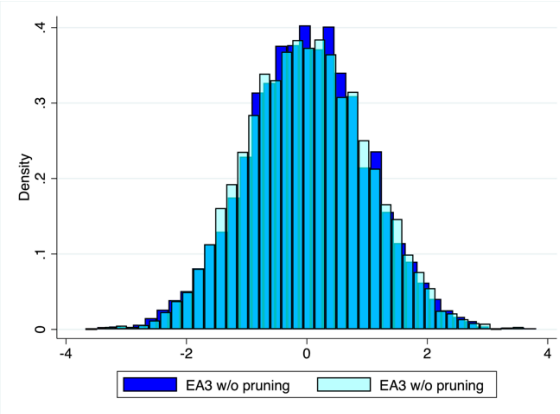

Figure 1: histogram of PGSs with and without pruning

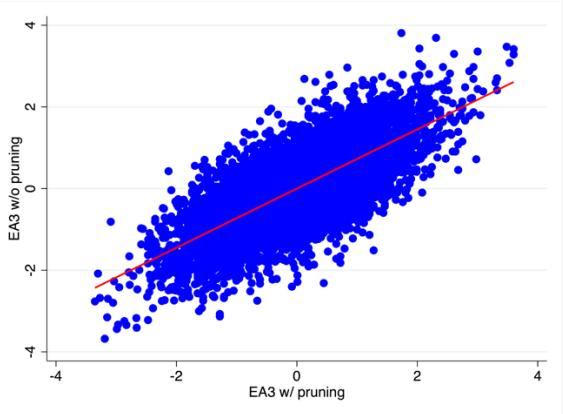

Figure 2: scatterplot of PGSs with and without pruning

[2] Infinium HumanHap 550K v3 (2,592 samples and 561,303 SNPs), Infinium HumanHap 550K v1.1 (1,436 samples and 555,174 SNPs), Illumina Human 660-Quad (871 samples, 582,892 SNPs), Illumina 1.2M (2,922 samples and 1,157,986 SNPs), Affymetrix v6 (2,997 samples and 934,967 SNPs) and Affymetrix 500K (1,502 samples and 490,032 SNPs).

[3] We are very grateful to Ayse Okbay from the University of Amsterdam for having provided us with the appropriate GWAS summary statistics (i.e. excluding the NCDS sample from the GWAS).

3. The predictive power of the polygenic score for life course outcomes

In this section we examine the predictive power of the polygenic score for various outcomes, starting from educational attainment. Lee et al. (2018) showed that the mean prevalence of college completion ranged from 10% for individuals in the bottom quintile of the PGS in both the AddHealth and the Health and Retirement Study (HRS), up to 50% in the HRS and 60% in the AddHealth for individuals in the top quintile of the PGS. We replicate these patterns perfectly in the NCDS. Figure 3 shows that the proportion of cohort members achieving a university degree or higher qualification ranges from 0.09 at the bottom quintile of the PGS to 0.42 at the top quintile; along the same lines, the proportion of cohort members who stayed on in education after the minimum compulsory school leaving age (MSLA) ranges from 0.15 at the bottom quintile of the PGS to 0.52 at the top quintile. On the other hand, interestingly there is no association between the polygenic score for education and the mean prevalence of starting school full-time before 5 years of age, which is 0.52 on average.

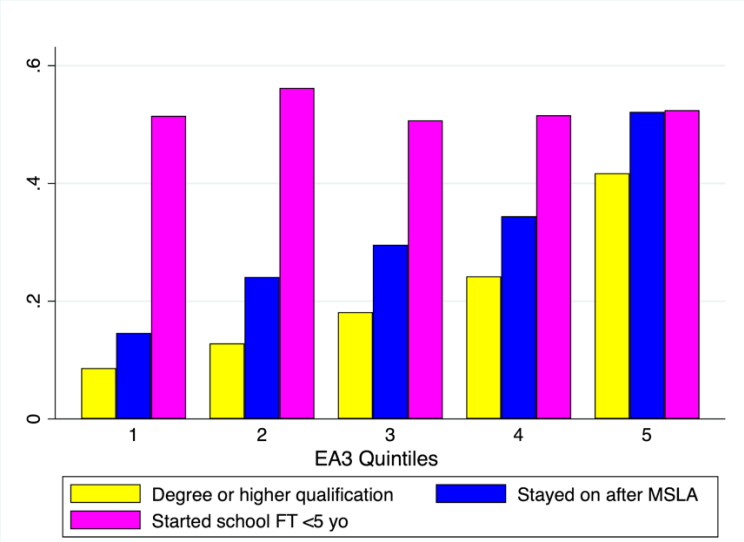

Figure 3 Education by PGS quintiles

Lee et al. (2018) also showed that the polygenic score explains between 6% and 12% of the variation in different measures of education (high school and college completion, and years of education) in both the HRS and the AddHealth. We replicate also these results in the 1958 cohort: Figure 4 shows that the PGS for education explains 6%-7% of the variation in post-compulsory education, 8%-10% of the variation in A-level or higher education, and 8% of the variation in degree or higher education. Lastly, Lee et al. (2018) showed that the polygenic score explains 2%-4% of the variation in cognition in the HRS, and 7%-12% of the variation in academic achievement in the AddHealth. Again, we are able to replicate these findings in the NCDS: Figure 5 shows that the polygenic score explains 7%-8% of the variation in test scores for mathematics and reading comprehension at ages 11 and 16. Additionally, Figure A1 in the Appendix reveals that the predictive power of the polygenic score for the Math Test at 11 is comparable to the predictive power of maternal education, and that it is substantial even on the top of significant predictors such as parental age, father social class, mother education, birth weight and smoking in pregnancy.

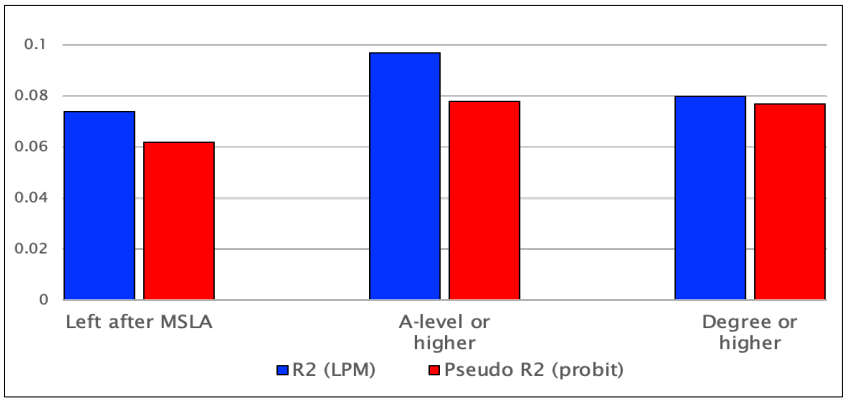

Figure 4 Predictive power of the PGS for Educational Outcomes

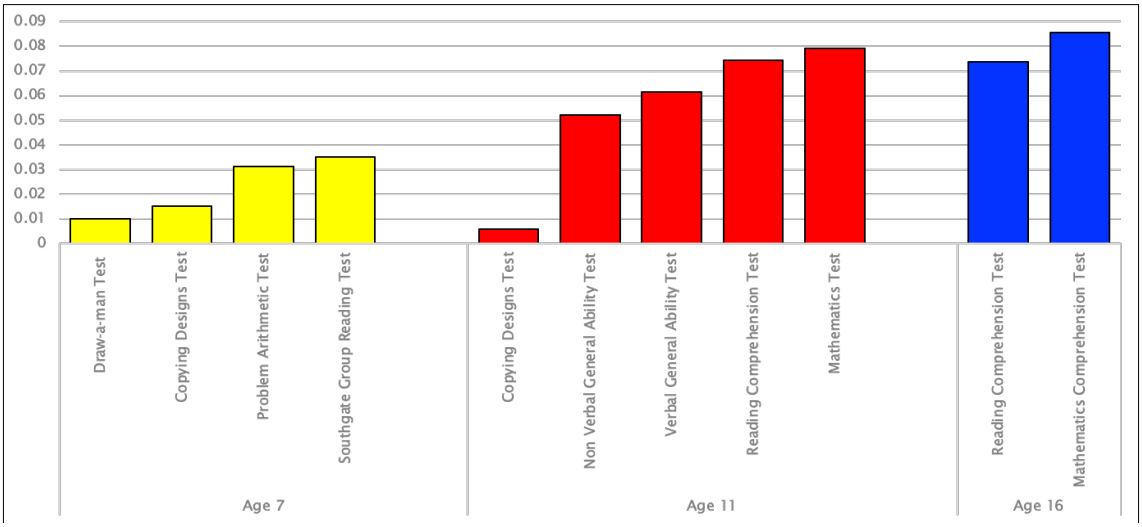

Figure 5 Predictive power of the PGS for Cognitive Development

Lastly, Figure 6 shows that the polygenic score for education is also predictive of socioemotional development (measured by the teacher-reported Bristol Social Adjustment Guide, BSAG; left panel) and parental investments (reading to the child and going out with the child, left panel) at age 11; and of health and socioeconomic outcomes since age 23 until age 55. This is the first evidence of the wide predictive power across the life course of the polygenic score for education in a UK population (see Belsky et al. 2016 for evidence based on a cohort from New Zealand).

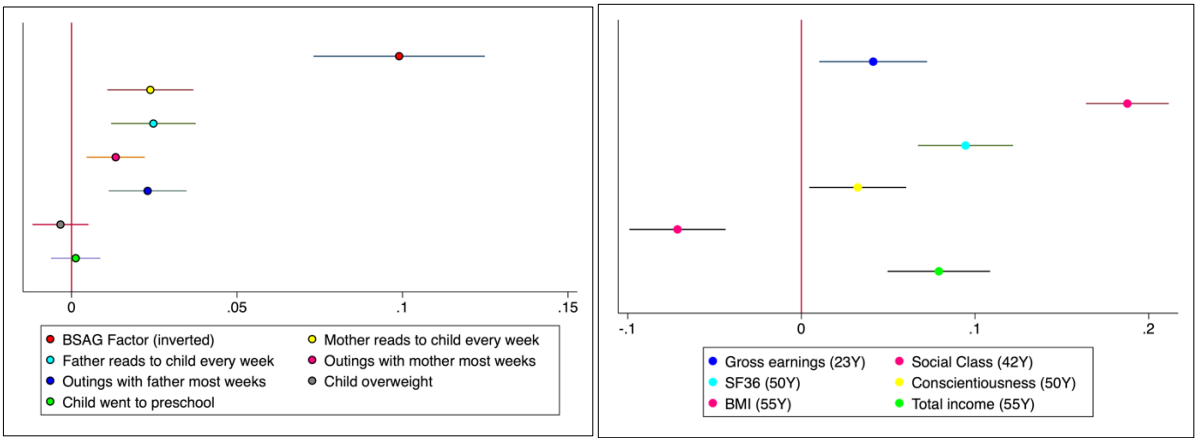

Figure 6 Predictive power of the PGS for Child (left) and Adult (right) outcomes

Note: BSAG=Bristol Social Adjustment Guide (standardized factor). All outcomes in the right panel are standardized for ease of comparison.

4. Maternal education, genes, and inequalities in offspring human capital

After having shown in the previous section that the polygenic score for education is significantly associated with a variety of outcomes since childhood until adulthood, in this section we examine its interaction with maternal education. For identifying the causal impact of parental education on child outcomes, we use the reform brought about by the Education Act of 1944, which raised the minimum school leaving age from 14 to 15 in April 1947 (Oreopoulos, 2006; Clark and Royer, 2013; Lindeboom et al., 2009). We use regression discontinuity techniques with a fuzzy design, since the school leaving age does not deterministically depend on exposure to the reform; however, exposure to the reform depends on the year of birth. Specifically, the first cohort of parents which is affected by the reform is that born in 1934; hence, we restrict the NCDS sample to mothers born in a five-year window (1931-1932 and 1934-1935), and we exclude 1933 to avoid possible misclassification.[4] In this subsample, the first stage is very strong, with mothers to the right of the cutoff ending up acquiring one year or more of education, as shown in Column 1 of Table 1. In columns 2 and 3 of Table 1 we then show that there is no association between being exposed to the reform and the child having a PGS above the median, or the mother's father being of a high social class – attenuating the concerns that there might be compositional differences at the two sides of the cutoff. Lastly, in column 4 of Table 1 we construct a "placebo reform" indicator, which takes value one if the mother is born in 1936-1937 and 0 if born in 1933-1934 (hence shifting the reform by two years): reassuringly, we find no association between this placebo reform and the age at which the mother left full-time education.

Table 1: Education Act 1944 Reform and Mother Educational Attainment

Outcome:	**Age left FT Education**	**Child high EA3 PGS**	**Mother's father is high Social Class**	**Age left FT Education**
	(1)	**(2)**	**(3)**	**(4)**
ROSLA	0.900***	-0.001	0.052	
	(0.251)	(0.043)	(0.043)	
Placebo ROSLA				0.090
				(0.224)
Controls	Yes	Yes	Yes	Yes
Observations	1,234	1,378	1,378	1,059
R-squared	0.032	0.026	0.018	0.019

Note: Sample restricted to mothers born in 1931, 1932, 1934, 1935 in columns (1)-(3), and in 1933, 1934, 1936, 1937 in column (4). Child high EA3 PGS is a binary indicator taking value 1 if the EA3 PGS is above the median. "ROSLA" (Reform of School Leaving Age) is a binary indicator which takes value 1 if the mother is born 1934-1935, 0 if born 1931-1932. "Placebo ROSLA" is a binary indicator which takes value 1 if the mother is born 1936-1937, 0 if born 1933-1934. Models also include mother's age, and binary indicators for the region of birth (there are 12 regions). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We then proceed by estimating the reduced form impacts of the reform via a linear specification by Ordinary Least Squares, whereby we regress different indicators for child human capital on the

[4] We restrict this analysis to mothers since very few fathers were born in the relevant window.

ROSLA binary indicator, controlling for maternal age at birth and region fixed effects. The main results are reported in Table 2. The top panel shows that there is no effect of parental education on three indicators of child human capital – whether the child started full-time school before age 5, a composite indicator for child cognitive development at ages 7 and 11 and a composite measure of child educational attainment at age 16 – across the whole sample. However, when we split the sample by the median of the polygenic score, we find that the children of the mothers exposed to the reform are, on average, 21.5 p.p. more likely to start school earlier, and have roughly half a standard deviation higher cognitive development across childhood and adolescence; no effects are uncovered for children at genetic risk of low education. The bottom panel of Table 2 shows that no significant impact is detected when we use instead the placebo reform.

Table 2: Reform and Child Human Capital – Main Effects, Heterogeneity by Genes & Placebo

Outcome:	**Started FT school before age 5**			**Child Cognitive Development (Ages 7 and 11)**			**Child Educational Attainment (Age 16+)**		
Sample	***All***	***Child low EA3 PGS***	***Child high EA3 PGS***	***All***	***Child low EA3 PGS***	***Child high EA3 PGS***	***All***	***Child low EA3 PGS***	***Child high EA3 PGS***
ROSLA	0.102	-0.021	0.215*	0.179	-0.161	0.484**	0.285	0.038	0.524*
	(0.081)	(0.114)	(0.117)	(0.167)	(0.233)	(0.233)	(0.189)	(0.242)	(0.277)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,441	750	691	1,301	676	625	1,030	525	505
R-squared	0.074	0.087	0.074	0.023	0.015	0.043	0.047	0.034	0.074
Placebo ROSLA	0.061	0.121	0.034	-0.177	-0.177	-0.183	0.109	0.097	0.081
	(0.086)	(0.117)	(0.128)	(0.190)	(0.255)	(0.280)	(0.202)	(0.263)	(0.302)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,220	672	548	1,083	599	484	880	479	401
R-squared	0.109	0.107	0.134	0.010	0.014	0.019	0.026	0.027	0.043

Note: Sample restricted to mothers born in 1931, 1932, 1934, 1935 in the top panel, and in 1933, 1934, 1936, 1937 in the bottom panel. Child Cognitive Development at ages 7 and 11 is a standardized Bartlett score from the factor analysis of the following seven tests: the Copy Design, Math and Reading Tests at age 7, and the British Ability Scales (Verbal and Non-Verbal), the Southgate Reading and the Math Test at age 11. Child Educational Attainment at age 16+ is a standardized Bartlett score from the factor analysis of reading and math tests at age 16, teacher's assessment of the child being of A-level and above ability (across different domains), and whether the cohort member has achieved A-level or above education. Child low/high EA3 PGS is a binary indicator taking value 1 if the EA3 PGS is below/above the median. "ROSLA" is a binary indicator which takes value 1 if the mother is born 1934-1935, 0 if born 1931-1932. "Placebo ROSLA" is a binary indicator which takes value 1 if the mother is born 1936-1937, 0 if born 1933-1934. Models also include mother's age, and binary indicators for the region of birth (there are 12 regions). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

While the results from the 1958 cohort are corroborated by the placebo reform check, they have a couple of key limitations. First, identification of the reform's impacts relies on variation across cohorts, so its effect might be confounded by other influences which might have affected cohorts born in the same years. Second, while child genes are predetermined, they are not exogenous, so the interaction with the PGS might confound interactions with other correlated traits (for example, with parental cognitive ability or PGS). Recent work has shown the importance of controlling for parental genetic influences (even non-transmitted ones) not to over-estimate the importance of child genes (see for example Kong et al., 2018). For this reason, we validate our NCDS analysis using data from the MCS, which has collected DNA from both children and their parents.

5. Genetic data and polygenic scores in the Millennium Cohort Study

The UK Millennium Cohort Study follows the lives of ~19,000 children born in the UK in 2000-01. Seven sweeps carried out so far: 9 months, 3y, 5y, 7y, 11y, 14y, 17y. Rich information on child cognitive, behavioural and health development, parental investments and family characteristics; linkage with administrative records on health and educational attainment (National Pupil Database). Saliva samples from cohort members and their biological parents were collected at the age 14y sweep. DNA extraction, assay (using Illumina GSA), cleaning, QC, and imputation were conducted at the University of Bristol (see Fitzsimons et al. 2020 for details). The DNA data available on approximately 4,500 complete trios. We constructed the EA3 PGS for both parents and children using the same scoring procedure as for the NCDS (we plot the histograms in Figure 7). Figures 8, 9 and 10 show the scatterplots with fitted regression lines of the PGSs of, respectively, the child-mother pair, the child-father pair, and the father-mother pair; the correlations of the PGSs for the first two pairs are 0.55 and 0.57 respectively, while the correlation between the mother and the father PGSs is 0.15—showing a modest degree of genetic assortative mating.

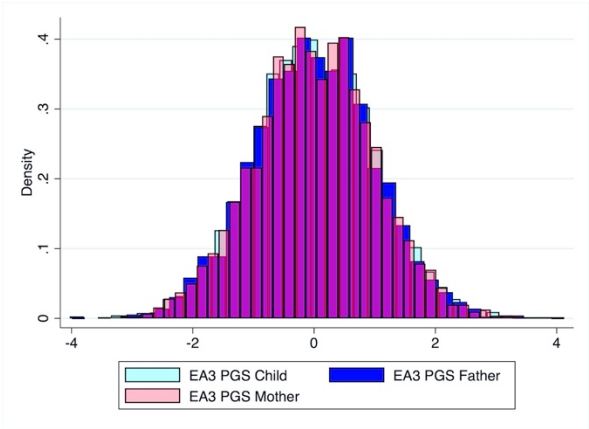

Figure 7: Histograms of EA3 PGS for MCS trio

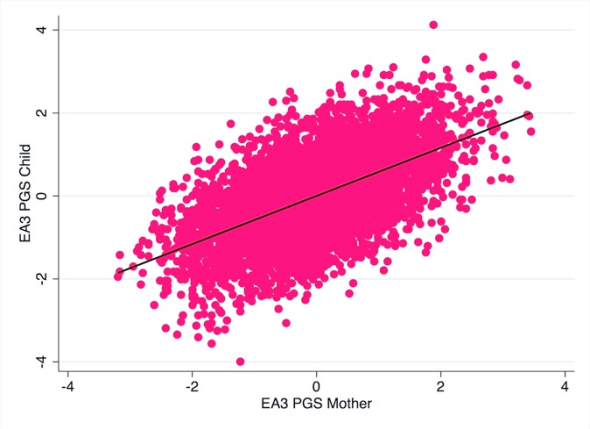

Figure 8: Scatterplot of EA3 PGS for mother-child pair

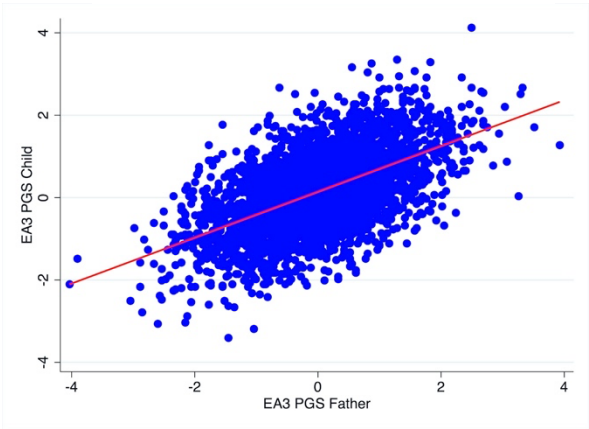

Figure 9: Scatterplot of EA3 PGS for father-child pair

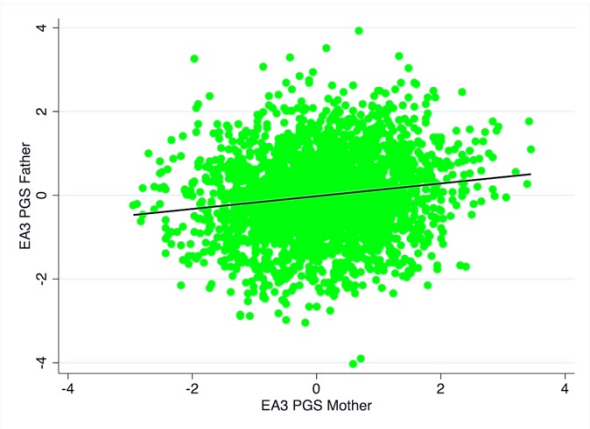

Figure 10: Scatterplot of EA3 PGS for father-mother pair

In Table 3 we show the predictive power of the EA3 PGS for child cognitive development at ages 3, 5 and 7. We construct indicators for child development via factor analysis with Bartlett scoring, using the following tests: the Bracken Colours, Shapes and Letters tests at age 3, the British Ability Scales Naming Vocabulary, Picture Similarity and Patterns Comprehension Ability Scores at age 5; and the BAS Word Reading and Patterns Comprehension Ability Scores and the NFER math scores at age 7. The results show that the EA3 PGS is strongly predictive of child cognitive

development: children with 1SD higher PGS have between 0.21 and 0.32 SD higher cognitive scores; the proportion of explained variation ranges from 4% at age 3 to 6% at age 7. Interestingly, these associations are stronger than those reported in Houmark et al. (2020), who use data from the Avon Longitudinal Study of Parents and Children (ALSPAC) – a cohort of children born in the Bristol area in 1990. Additionally, the associations between child genes and their cognitive development are still of sizable magnitude and significant after controlling for the parental PGSs, and also for parental education (Table 3), while in Houmark et al. (2020) they are reduced in magnitude and driven to insignificance when controlling for parental genes. This might be indicative of different channels through which genes might affect child development.

Table 3: Genes and Child Development in the MCS

Outcome	**Child Cognitive Development Age 3**				**Child Cognitive Development Age 5**				**Child Cognitive Development Age 7**			
Child EA3 PGS	0.259***	0.204***	0.171***	0.142***	0.212***	0.167***	0.130***	0.114***	0.324***	0.271***	0.270***	0.245***
	(0.017)	(0.023)	(0.038)	(0.037)	(0.015)	(0.019)	(0.032)	(0.031)	(0.016)	(0.021)	(0.036)	(0.036)
Mother EA3 PGS		0.080***	0.045	-0.029		0.082***	0.074***	0.011		0.086***	0.083***	0.005
		(0.022)	(0.032)	(0.033)		(0.019)	(0.027)	(0.028)		(0.021)	(0.030)	(0.030)
Father EA3 PGS			0.021	-0.033			0.006	-0.045			0.016	-0.051
			(0.032)	(0.033)			(0.027)	(0.028)			(0.032)	(0.032)
Child is a boy				-0.318***				-0.134***				-0.067
				(0.050)				(0.041)				(0.046)
Mother A-Level +				0.327***				0.301***				0.338***
				(0.056)				(0.045)				(0.053)
Father A-Level +				0.367***				0.307***				0.420***
				(0.055)				(0.044)				(0.051)
Observations	5,659	4,914	2,391	2,331	6,257	5,414	2,598	2,536	5,956	5,155	2,509	2,447
R-squared	0.040	0.041	0.027	0.088	0.033	0.036	0.028	0.080	0.064	0.065	0.071	0.128

Note: Results from ordinary least squares models, each column presents results for a separate regression. Child Cognitive Development is a standardized Bartlett score obtained from the factor analysis of: the Bracken Colours, Shapes and Letters test at age 3; the British Ability Scales Naming Vocabulary, Picture Similarity and Patterns Comprehension Ability Scores at age 5; the BAS Word Reading and Patterns Comprehension Ability Scores and the NFER math scores at age 7. Child/Father/Mother EA3 PGS are standardized with mean 0 and SD 1. A-Level + is a binary indicator which takes value 1 if the father/mother obtained an educational qualification at A-Level or above. Sample only includes individuals with European ancestry. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

6. Paternal education, genes, and inequalities in offspring human capital

For identifying the causal impact of parental education on child outcomes, we use the reform brought about by the Education Act 1962, which in Chapter 12 "School leaving dates in England and Wales" states that (i) those born from the beginning of September until the end of January have to stay at school until "the end of the appropriate spring term" (ii) those born on or after the beginning of February but before the end of the summer term have to stay at school "until the end of that summer term". This reform affected the cohorts born after 1 September 1957, and has been rarely exploited in the literature (one exception is Del Bono and Galindo-Rueda, 2004).

We select cohorts born 1958 to 1972 (as on 1 September 1972 the Minimum School Leaving Age was further increased from 15 to 16 years of age), over a 6-month window from November to April. Approximately 80% of fathers and 60% of mothers in our sample are selected; unfortunately, there is no first stage for the mothers, so we are left working with the sample of fathers. The treatment group is defined as those fathers born February-April (as they need to stay on in school until the end of the summer term), while the control group is comprised of those fathers who are born November-January (as they can leave school at the end of the spring term).

The first stage results are shown in Table 4. Columns 1 and 2 show that fathers exposed to the reform are 7.5 p.p. more likely to acquire A-level education or above (col. 1), a result which

holds when controlling for the paternal PGS (col. 2). Columns 3-5 show that exposure to the reform is not associated with any member of the trio (child, father, mother) having a high PGS

Table 4: Reform and Father Educational Attainment

Outcome:	**A-Level or Above**		**Child high EA3 PGS**	**Father high EA3 PGS**	**Mother high EA3 PGS**	**A-Level or Above**	**A-Level or Above**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Father Born Feb-Apr	0.075**	0.081**	0.033	0.022	-0.012	0.019	
	(0.033)	(0.032)	(0.032)	(0.033)	(0.034)	(0.058)	
Father EA3 PGS		0.128***					
		(0.015)					
Father Born May-July							-0.045
							(0.034)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample (geography)	England +Wales	England +Wales	England +Wales	England +Wales	England +Wales	Scotland + NI	England +Wales
Sample (birth months)	Nov-Apr	Nov-Apr	Nov-Apr	Nov-Apr	Nov-Apr	Nov-Apr	Feb-July
Observations	944	928	993	938	860	310	956
R-squared	0.051	0.110	0.036	0.035	0.052	0.055	0.041

Note: Sample restricted to cohorts born 1958-1973. Columns (1)-(6) includes fathers born November-April; column (7) includes fathers born February-July. A-Level or Above is a binary indicator which takes value 1 if the father obtained an educational qualification at A-Level or above. Child/Father/Mother high EA3 PGS are binary indicators taking value 1 if the EA3 PGS is above the median. "Father Born Feb-Apr" is a binary indicator which takes value 1 if born Feb-Apr, 0 if born Nov-Jan. Father EA3 PGS is standardized with mean 0 and SD 1. "Father Born May-July" is a binary indicator which takes value 1 if born May-July, 0 if born Feb-April. Models also include binary indicators for the year of birth and the region. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Reform and Child Human Capital – Controlling for Genes

Outcome	**Child Cognitive Development**			
Father Born Feb-Apr	0.167***	0.141**	0.183***	0.164**
	(0.064)	(0.062)	(0.070)	(0.069)
Child EA3 PGS		0.253***		0.197***
		(0.032)		(0.051)
Mother EA3 PGS			0.147***	0.050
			(0.036)	(0.044)
Father EA3 PGS			0.105***	0.009
			(0.036)	(0.044)
Controls	Yes	Yes	Yes	Yes
Observations	838	836	699	698
R-squared	0.045	0.112	0.090	0.108

Note: Sample restricted to cohorts born 1958-1973 and to fathers born November-April in England and Wales. Child Cognitive Development is a standardized Bartlett score obtained from the factor analysis of the age 3 Bracken Colours, Shapes and Letters test, the age 5 British Ability Scales Naming Vocabulary, Picture Similarity and Patterns Comprehension Ability Scores, the age 7 BAS Word Reading and Patterns Comprehension Ability Scores and the age 11 BAS Verbal Similarities Ability Scores. Child/Father/Mother EA3 PGS are standardized with mean 0 and SD 1. "Father Born Feb-Apr" is a binary indicator which takes value 1 if born Feb-Apr, 0 if born Nov-Jan. Models also include binary indicators for the year of birth and the region. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(above the median of the sample distribution); hence, like done for the NCDS, we rule out compositional effects by genes. Columns 6 and 7 carry out placebo tests, using the smaller sample for Scotland and Northern Ireland (given that the reform was only implemented in England and Wales), and constructing a placebo reform by shifting its timing by 3 months (so that fathers born February-April are now in the control group and those born May-July are in the treated group), respectively; reassuringly, we detect no significant impacts in columns 6 and 7.

The main results for the whole sample are displayed in Table 5: the children of fathers affected by the reform have higher cognitive development (a composite indicator across ages 3, 5, 7 and 11) by 0.14-0.18 SD, even conditional on child and parental genes. When we then split the sample at the median of the child PGS distribution, we see that – like for the NCDS – the effect of paternal education on child cognitive development is more than twice in magnitude and statistically significant only for the children with high genetic endowments (table 6). Lastly, a placebo test validates our results (Table 7).

Table 6: Reform and Child Human Capital – Heterogeneity by Genes

Outcome	**Child Cognitive Development**									
Sample	***Child high EA3 PGS***					***Child low EA3 PGS***				
Father Born Feb-Apr	0.179**	0.154*	0.219**	0.196**	0.274**	0.079	0.086	0.109	0.114	0.101
	(0.085)	(0.082)	(0.094)	(0.093)	(0.106)	(0.099)	(0.097)	(0.109)	(0.108)	(0.093)
Child EA3 PGS		0.261***		0.219***	0.189**		0.257***		0.168	0.201***
		(0.064)		(0.081)	(0.075)		(0.084)		(0.110)	(0.068)
Mother EA3 PGS			0.107*	0.035	0.022			0.098*	0.058	0.084
			(0.056)	(0.064)	(0.066)			(0.058)	(0.061)	(0.059)
Father EA3 PGS			0.097*	0.033	-0.026			0.006	-0.035	-0.020
			(0.057)	(0.060)	(0.095)			(0.064)	(0.069)	(0.093)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	469	469	389	389	335	367	367	309	309	363
R-squared	0.075	0.106	0.113	0.129	0.168	0.090	0.114	0.098	0.105	0.117

Note: Sample restricted to cohorts born 1958-1973 and to fathers born November-April in England and Wales. Child Cognitive Development is a standardized Bartlett score obtained from the factor analysis of the age 3 Bracken Colours, Shapes and Letters test, the age 5 British Ability Scales Naming Vocabulary, Picture Similarity and Patterns Comprehension Ability Scores, the age 7 BAS Word Reading and Patterns Comprehension Ability Scores and the age 11 BAS Verbal Similarities Ability Scores. Child high EA3 PGS is a binary indicator taking value 1 if the EA3 PGS is above the median. "Father Born Feb-Apr" is a binary indicator which takes value 1 if born Feb-Apr, 0 if born Nov-Jan. Child/Father/Mother EA3 PGS are standardized with mean 0 and SD 1. Models also include binary indicators for the year of birth and the region. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Placebo – Heterogeneity by Genes

Outcome	**Child Cognitive Development**		
Sample	***All***	***Child high EA3 PGS***	***Child low EA3 PGS***
Father Born May-July	-0.091	-0.155	-0.197
	(0.112)	(0.136)	(0.200)
Control	Yes	Yes	Yes
Observations	261	169	92
R-squared	0.061	0.097	0.316

Note: Sample restricted to cohorts born 1958-1973 and to fathers born February-July in England and Wales. Child Cognitive Development is a standardized Bartlett score obtained from the factor analysis of the age 3 Bracken Colours, Shapes and Letters test, the age 5 British Ability Scales Naming Vocabulary, Picture Similarity and Patterns Comprehension Ability Scores, the age 7 BAS Word Reading and Patterns Comprehension Ability Scores and the age 11 BAS Verbal Similarities Ability Scores. Child high EA3 PGS is a binary indicator taking value 1 if the EA3 PGS is above the median. "Father Born May-July" is a binary indicator which takes value 1 if born May-July, 0 if born Feb-April. Models also include binary indicators for the year of birth and the region. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7. Preliminary conclusions

In this paper we have leveraged on the novel availability of genetic data in the 1958 and in the 2000 British birth cohorts and in recent advances in genetics to study how specific genetic variants, summarized as a linear index—known as a polygenic score—are associated with human capital accumulation and how they interact with paternal education. We have presented two main sets of results. First, we have shown that the polygenic score is highly predictive of child development, parental investments, educational attainment, and health and socioeconomic outcomes since childhood until adulthood; the predictive power of the polygenic score for child cognitive development persists in two cohorts born decades apart, also conditional on parental genes. Second, we have found evidence that the genetic factors measured by this score interact strongly with parental education in affecting human capital in the offspring. In particular, by exploiting exogenous variation in education induced by two schooling reforms announced in the Education Acts 1944 and 1962, we have found that increasing parental education raised offspring human capital only for children with high genetic potential. Our findings have uncovered a potential mechanism through which the interplay of genes and environments amplifies inequalities in human capital across generations.

References

Almond, D., Currie, J. and Duque, V., 2018. Childhood circumstances and adult outcomes: Act II. Journal of Economic Literature, 56(4), pp.1360-1446.

Belsky, D.W., Moffitt, T.E., Corcoran, D.L., Domingue, B., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sugden, K., Williams, B.S. and Poulton, R., 2016. The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. Psychological science, 27(7), pp.957-972.

Black, S.E., Devereux, P.J. and Salvanes, K.G., 2005. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. American economic review, 95(1), pp.437-449.

Carneiro, P., Meghir, C. and Parey, M., 2013. Maternal education, home environments, and the development of children and adolescents. Journal of the European Economic Association, 11(suppl_1), pp.123-160.

Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6), 2087-2120.

Currie, J. and Moretti, E., 2003. Mother's education and the intergenerational transmission of human capital: Evidence from college openings. The Quarterly journal of economics, 118(4), pp.1495-1532.

Bono, E. D., & Galindo-Rueda, F. (2004). Do a few months of compulsory schooling matter? The education and labour market impact of school leaving rules. The Education and Labour Market Impact of School Leaving Rules (August 2004).

Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet, 9(3), p.e1003348.

Dudbridge, F. and Newcombe, P.J., 2015. Accuracy of gene scores when pruning markers by linkage disequilibrium. Human heredity, 80(4), pp.178-186.

Fitzsimons, E., Timpson, N. J., Moulton, V., Hughes, D. A., Neaves, S., Ho, K. M., Hemani, G., Calderwood, L., Gilbert, E., & Ring, S. M. (2021). Collection of genetic data at scale for a nationally representative population: the UK Millennium Cohort Study. *Longitudinal and Life Course Studies*. https://doi.org/10.1332/175795921X16223668101602

Houmark, M., Ronda, V., & Rosholm, M. (2020). The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Linnér, R.K. and Fontana, M.A., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics, 50(8), pp.1112-1121.

Lindeboom, M., Llena-Nozal, A., & van Der Klaauw, B. (2009). Parental education and child health: Evidence from a schooling reform. *Journal of Health Economics*, 28(1), 109-131.

Lundborg, P., Nilsson, A. and Rooth, D.O., 2014. Parental education and offspring outcomes: evidence from the Swedish compulsory School Reform. American Economic Journal: Applied Economics, 6(1), pp.253-78.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1), 152-175.

Shah, T., Engmann, J., Dale, C., Shah, S., White, J., Giambartolomei, C., McLachlan, S., Zabaneh, D., Cavadino, A., Finan, C. and Wong, A., 2013. Population genomics of cardiometabolic traits: design of the university college london-london school of hygiene and tropical medicine-edinburgh-bristol (UCLEB) consortium. PloS one, 8(8), p.e71345.

Appendix

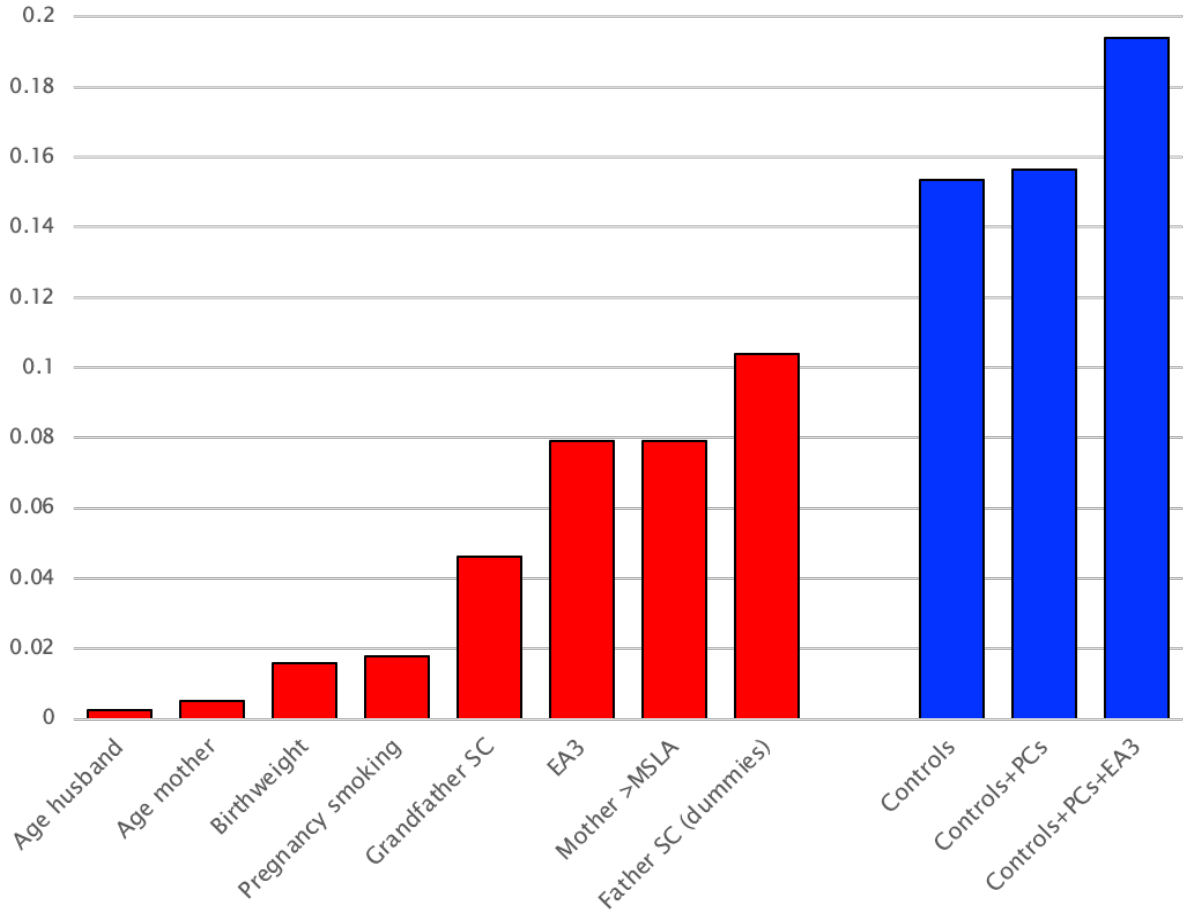

Figure A1 Predictive Power of the Education Polygenic Score for the Math Test at Age 11

Note: "Controls" include father and grandfather social class, maternal education, parental age, birth weight and smoking in pregnancy.