Online Appendix

"The Nurture of Nature and the Nature of Nurture:
How Genes and Investments Interact in the Formation of Skills"

Mikkel Aagaard Houmark, Victor Ronda and Michael Rosholm

Appendix A Additional Information on the Estimation Sample

In this section, we describe the sample selection process moving from the full ALSPAC sample to the estimation sample. We discuss the imputation of parental genotype and the representativeness of the sample.

Because we rely on specific information about the child and the parents, we select a subsample of individuals for whom this information is available. The sample selection process is illustrated in Table A1. Our ALSPAC data set includes 14,062 children. We have genetic information on 8,804 of them. Parental genotypes are often missing, especially for fathers, but using the approach described in Section 3.2, we are able to impute parental genotypes for 6,905 children. We drop individuals with missing information on sex. To ensure that the polygenic indexes are valid, we further exclude non-European individuals and individuals who are outliers because they have extreme values (above 3.5 or below -3.5) on the first three principal components of the genetic matrix. This leaves us with 6,519 individuals who may be analyzed in terms of the genetics of the mother and the child. Finally, we exclude individuals with many missing responses for the measures of skills and investments. We allow up to ten such missing values to be imputed for each summary measure, producing a final sample size of 4,510 for the main analysis. Had we not imputed the parental genotypes, we would have had a reduced sample size of only 1,267 individuals

As we need to select our sample on the above criteria to make a valid inference, our results should be interpreted for this subsample of ALSPAC individuals. Excluding non-Europeans and principal component outliers is necessary for the polygenic indexes to be valid. This causes the sample to be somewhat positively selected, as shown in Table A2. However, we also see that the selection is much less pronounced than if we had not imputed missing parental genotypes. In that case, we would have to rely on a reduced sample consisting only of individuals with both parents genotyped. At a minimum, both parents would then have to be present in the household, and this induces a much stronger positive selection, especially in terms of the father's polygenic index. Hence, imputing missing genotypes has the additional advantage that it increases the representativeness of the sample. In Tables A3 and A4, we show how the reduced form results differ depending on whether we use the imputed or non-imputed polygenic indexes, and whether we use the main or the reduced sample. This shows that we would have reached the same preliminary conclusions from the reduced form analysis if we had only used the individuals for whom we observe both parental genotypes. Using the imputed genotypes in the main analysis is thus preferable, as it increases the precision of the estimates by substantially increasing the sample size.

A major difference between these reduced form results and the main analysis is that skills and investments also enter as independent variables that cause variation in next-period skills and investments. This makes correcting for measurement error crucial. In Tables A5 and A6, we report the signal to variance ratios for each measure used to capture latent skills and investments, as explained in Section 4.1.3. It is evident from the tables that the degree of measurement error differs substantially, both across measures and across time for a given measure. In Table A7, we similarly report the signal to variance ratios for the different PGIs as explained in Section 4.2.1. Again, we see significant differences in measurement error across the PGIs. However, each PGI has approximately the same signal to variance ratio for the child, the mother, and the father.

Table A1: SAMPLE SELECTION PROCEDURE

Sample criteria	Observations
Full ALSPAC sample of children	14,062
With information on child genotypes	8,804
With imputed genotypes for child and parents	6,905
Excluding individuals with missing information on sex	6,902
Excluding individuals of non-European descent	6,580
Excluding outliers on the first three principal components	6,519
With no more than ten missing measures of skills or investments	4,510
With non-imputed parental genotypes (reduced sample)	1,267

Notes: This table shows the criteria for selecting the main sample and the associated number of individual observations.

Table A2: Summary statistics

	Full sample	Main sample	Dropped individuals	Reduced sample
Child's PGS	-0.076	0.000	-0.219	0.130
	(1.012)	(1.000)	(1.019)	(1.000)
Mother's PGS	-0.074	0.000	-0.214	0.114
	(1.018)	(1.000)	(1.037)	(1.012)
Father's PGS	-0.064	0.000	-0.186	0.203
	(0.972)	(1.000)	(0.906)	(1.300)
Female	0.496	0.490	0.507	0.468
	(0.500)	(0.500)	(0.500)	(0.499)
Birth order	1.793	1.747	1.897	1.686
	(0.914)	(0.869)	(1.000)	(0.851)
N	6902	4510	2392	1267

Notes: This table reports means and standard deviations for the full sample and for the main sample, as well as a reduced sample for whom we observe genotypes of both parents.

Table A3: EA PGI and Skills by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.043	-0.003	0.023	0.075	0.093	0.036	0.045
	(0.043)	(0.041)	(0.043)	(0.042)	(0.041)	(0.042)	(0.031)
Mother's PGS	0.042	0.028	0.047	0.073	0.049	0.050	0.048
	(0.035)	(0.034)	(0.035)	(0.034)	(0.034)	(0.034)	(0.025)
Father's PGS	-0.038	0.036	0.085	0.060	0.046	0.034	0.037
	(0.035)	(0.034)	(0.035)	(0.034)	(0.033)	(0.034)	(0.023)
R_2	0.010	0.082	0.066	0.053	0.045	0.025	0.044
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS (imputed)	0.026	-0.025	-0.001	0.045	0.076	0.036	0.026
	(0.042)	(0.040)	(0.042)	(0.041)	(0.040)	(0.041)	(0.029)
Mother's PGS (imputed)	0.008	0.020	0.043	0.065	0.057	0.032	0.038
	(0.035)	(0.033)	(0.035)	(0.034)	(0.033)	(0.034)	(0.024)
Father's PGS (imputed)	-0.007	0.019	0.065	0.041	0.035	0.036	0.031
	(0.027)	(0.026)	(0.027)	(0.026)	(0.026)	(0.026)	(0.018)
R_2	0.006	0.080	0.060	0.039	0.040	0.024	0.039
N	1267	1267	1267	1267	1267	1267	7602
Panel C:							
Child's PGS	0.015	0.007	-0.013	0.071	0.076	0.050	0.034
	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.018)
Mother's PGS	0.032	0.037	0.069	0.055	0.057	0.041	0.049
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.013)
Father's PGS	-0.006	0.019	0.061	0.028	0.034	0.027	0.027
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.015)
R_2	0.006	0.066	0.047	0.039	0.035	0.019	0.032
N	4510	4510	4510	4510	4510	4510	27060

Notes: This table reports parameter estimates from regressions used to link the polygenic index for educational attainment to children's skills across childhood. To test the effect of the EA PGI, we regress at each age the skill measure on the polygenic index, controlling for sex and the first 15 principal components of the genetic matrix. In Panel A, we use only the non-imputed indexes (and hence, the reduced sample), while Panel B and C use the imputed indexes in the reduced and the main sample, respectively. Skills have been standardized as described in the data section, with missing values set equal to the median for that measure, allowing for a maximum of ten such imputations per summary index. The polygenic indexes were constructed using the summary statistics in Lee et al. (2018) without the 23andMe information. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Table A4: EA PGI and Investments by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.020	0.024	0.000	-0.054	0.002	0.064	0.009
	(0.041)	(0.041)	(0.039)	(0.040)	(0.041)	(0.040)	(0.032)
Mother's PGS	0.067	0.156	0.110	0.162	0.126	0.110	0.122
	(0.034)	(0.034)	(0.032)	(0.033)	(0.034)	(0.032)	(0.027)
Father's PGS	0.056	0.047	0.080	0.116	0.118	0.081	0.083
	(0.034)	(0.033)	(0.031)	(0.032)	(0.033)	(0.032)	(0.025)
R_2	0.016	0.034	0.027	0.058	0.052	0.043	0.042
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS	0.036	0.033	0.020	-0.030	-0.028	0.040	0.012
	(0.041)	(0.040)	(0.038)	(0.039)	(0.040)	(0.039)	(0.030)
Mother's PGS	0.042	0.106	0.064	0.132	0.104	0.070	0.086
	(0.034)	(0.033)	(0.031)	(0.033)	(0.033)	(0.032)	(0.025)
Father's PGS	0.020	0.023	0.034	0.072	0.100	0.072	0.054
	(0.026)	(0.026)	(0.024)	(0.025)	(0.026)	(0.025)	(0.019)
R_2	0.009	0.015	0.011	0.045	0.039	0.023	0.028
N	1267	1267	1267	1267	1267	1267	7602
Panel C:							
Child's PGS	0.063	0.043	0.051	0.008	0.003	0.024	0.032
	(0.028)	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.021)
Mother's PGS	0.056	0.130	0.069	0.132	0.104	0.133	0.104
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.016)
Father's PGS	0.038	0.059	0.058	0.094	0.116	0.093	0.076
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.017)
R_2	0.021	0.039	0.019	0.043	0.039	0.041	0.034
N	4510	4510	4510	4510	4510	4510	27060

Notes: This table reports parameter estimates from regressions used to link the polygenic index for educational attainment to investments across childhood. To test the effect of the EA PGI, we regress at each age the investments measure on the polygenic index, controlling for sex and the first 15 principal components of the genetic matrix. In Panel A, we use only the non-imputed indexes (and hence, the reduced sample), while Panel B and C use the imputed indexes in the reduced and the main sample, respectively. Investments have been standardized as described in the data section, with missing values set equal to the median for that measure, allowing for a maximum of ten such imputations per summary index. The polygenic indexes were constructed using the summary statistics in Lee et al. (2018) without the 23andMe information. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Table A5: Signal to Variance Ratio: Measures of Child Skills

	Period	0	1	2	3	4	5
Measure	Age	0-2	2-3	3-4	4-5	5-6	6-7
1	Can build tower of 8 bricks	6.71	3.55	2.06			
2	Plays cards (or board games)		5.51	19.72	37.67	69.34	55.82
3	Plays peek-a-boo	9.07					
4	Can focus eyes on small object	11.09					
5	Can build tower of 4 bricks	7.06					
6	Freq. names things	8.34					
7	Combines two different words		33.13				
8	Can copy vertical line with pencil		12.08				
9	Can copy and draw a circle		14.88	14.52			
10	Uses plurals		41.67	34.81			
11	Uses possessives		40.87	26.65			
12	Adds -ing to words		46.13	34.48			
13	Adds -ed to words		30.69	34.89			
14	Can copy and draw a plus sign / cross			18.19			
15	Can copy and draw a square			15.56	16.85		
16	Can write their name				32.54		
17	Can write any numbers				38.93		
18	Knows at least 10 letters				26.01		
19	Can read simple words				40.28		
20	Can read a story with <10 words per page				36.57		
21	Can count up to 20				23.86		
22	Can read a story with >10 words per page				21.59	21.99	19.90
23	Can count up to 100				16.97	28.96	31.61
24	Can play any board games				13.10	31.01	23.71

Notes: This table reports the individual measures of child skills. The signal-to-variance ratios indicate how much information relative to measurement error is contained in each of the measures. It is defined formally in Equation 22. A signal-to-variance ratio indicates that the measure is available in that period and is used in the estimation.

Table A6: Signal to Variance Ratio: Measures of Investments

	Period	0	1	2	3	4
Measure	Age	0-2	2-3	3-4	4-5	5-6
1	Freq. goes to places of interest	16.30	20.87	10.83	16.96	13.26
2	Freq. goes to library	3.32	11.61	4.71	10.90	6.49
3	Freq. mum reads to child	20.68	26.23	9.88	5.72	3.71
4	Freq. partner sings to child	16.45	11.99	20.20	9.30	9.62
5	Freq. child taken to park	17.63	16.92	8.48		
6	Freq. mum shows child picture books	21.89		12.38		
7	Freq. partner shows child picture books	22.02		47.75		
8	Freq. partner plays with toys with child	14.36		30.22		
9	Freq. partner reads to child	36.54		34.13	11.04	11.61
10	Freq. goes to swimming pool or sports area				26.41	21.17
11	Freq. goes to special classes or clubs				8.37	7.85

Notes: This table reports the individual measures of child investments. The signal-to-variance ratios indicate how much information relative to measurement error is contained in each of the measures. It is defined formally in Equation 22. A signal-to-variance ratio indicates that the measure is available in that period and is used in the estimation.

Table A7: Signal to Variance Ratio: Measures of Genetic Factor

Measure	Child	Mother	Factor
EA PGI (w.o. 23andMe)	91.94	92.96	89.38
EA PGI (23andMe)	46.56	45.95	46.84
Cog PGI (w.o. 23andMe)	34.52	32.77	32.55

Notes: This table reports the individual measures of the genetic factor. The signal-to-variance ratios indicate how much information relative to measurement error is contained in each of the measures. It is defined formally in Equation 22.

Appendix B Polygenic Indexes and Related Literature

The importance of genetics in explaining socio-economic outcomes has been well established in the behavioral genetics literature (Polderman et al., 2015; Plomin and von Stumm, 2018; Sacerdote, 2007; Silventoinen et al., 2020; Cesarini and Visscher, 2017; Branigan, McCallum, and Freese, 2013). The fraction of the variance of educational attainment that is explained by genes (also called heritability) has been estimated at around 40% in twin studies (Polderman et al., 2015). These estimates are consistent across a variety of kinship relationships (Cesarini and Visscher, 2017).[1] While useful, kinship studies tend to rely on a variety of strong assumptions about the familial relationship, which have been noted by several critics to be unappealing (for some early and recent criticisms, see, e.g., Taubman, 1976; Goldberger, 1979; Behrman and Taubman, 1989; Björklund, Jantti, and Solon, 2005; Manski, 2011; Durlauf, Kourtellos, and Tan, 2020).

Recent advances in molecular genetics have brought a dramatic reduction in the costs of measuring genetic variation at the molecular level in humans. This has triggered a renewed interest in the role of genes in human capital formation. A new and already vast research program relies on polygenic indexes (PGI) measured at the individual level to study how genotypic variation explains behavioral and educational outcomes. These genetic factors are outcome-specific, combine information on a large number of genetic variants, and capture a large fraction of the genetic variation explaining a variety of socio-economic outcomes.

Formally, a PGI for a particular outcome, w (pgi_i^w), is a linear combination of the SNP count variables weighted by the strength of association between each SNP and the outcome of interest:

$$pgi_i^w = \sum_{s=1}^{S} \beta_j^w g_{is} \tag{1}$$

where the weights $\{\beta_j^w\}$, are obtained from a genome-wide association study (GWAS).

A GWAS follows an atheoretical approach to test the relationship between the outcome of interest and each SNP individually. In a GWAS, the outcome of interest is regressed on each SNP, one by one, along with a set of controls for population stratification. In contrast to earlier literature that relied on single genetic variants, the so-called candidate genes approach that has faced a severe replication crisis (see, e.g., Chabris et al., 2012; Charney and English, 2012; Hewitt, 2012), the GWAS approach has generated a series of robust findings. In the

[1]However, individual estimates of heritability also vary substantially, e.g. across countries (Branigan, McCallum, and Freese, 2013). Potentially, such differences could be explained by the nurture of nature effect that we find. Indeed, the extent to which genetic inequalities are reinforced by family investments will depend on the existence of egalitarian policies that provide more equal investments in children.

present paper, we use the results from the latest GWAS for educational attainment (Lee et al., 2018) to construct polygenic indexes for educational attainment and cognitive skills (EA PGI and Cog PGI for short) for the individuals in our sample as described in Section 3.2. Based on these, we can identify the underlying genetic factor, as we show in Section 4.2.

Polygenic indexes for educational attainment have been widely used in economics and other social science fields. We now know that these indexes are highly predictive of education as well as many related outcomes, see e.g., Plomin and von Stumm (2018); Lee et al. (2018). Hence, the EA PGI predicts the accumulation of early childhood 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; Ronda et al., 2020), as well as earnings, socioeconomic mobility, and wealth, over and above the direct effect of education (Papageorge and Thom, 2019; Belsky et al., 2018; Barth, Papageorge, and Thom, 2020).

In particular, using data from the Health and Retirement Study, Papageorge and Thom (2019) show that the EA PGI predicts college graduation and, moreover, that this relation has an interaction with childhood socio-economic status (SES) in the sense that the relationship is considerably stronger for children growing up in higher SES families. They also show that the EA PGI explains labor earnings even after controlling for educational attainment. Barth, Papageorge, and Thom (2020) use the same data set to show that the EA PGI also predicts wealth at the time of retirement, even after controlling for educational attainment and labor income. Investigating potential mechanisms, they point to a better understanding of complex financial decision-making as one such channel.

It is important to note, however, that, while these indexes are highly predictive of educational attainment, and are pre-determined, they are not exogenous in most models. This point has been established in two seminal papers. Using a newly developed technique for studying heritability (relatedness disequilibrium regression – RDR), Young et al. (2018) show that neglect of parental genetic influences leads to an overestimation of the importance of the child's genes.[2] Using genetic information on the child and both of its parents, Kong et al. (2018) demonstrate how an EA PGI of parents' non-transmitted genes affect their children's educational attainment; they call this 'genetic nurture.' This corresponds to the 'family genetic associations' described in the paper. They document that the size of the family genetic associations is about one-third of the direct effect of the child's own genes on

[2]They estimate SNP heritability of educational attainment to be 17%, about 75% of the conventional estimates of around 22% (Rietveld et al., 2013; Okbay et al., 2016). The difference between SNP heritability and the 40 % heritability estimated in twin studies (Branigan, McCallum, and Freese, 2013; Cesarini and Visscher, 2017) is known as the missing heritability problem (Plomin and von Stumm, 2018). Recent research shows that for many phenotypes, this missing heritability is due to rare variants (Wainschtein et al., 2021).

educational attainment, implying moreover that the latter is overestimated in most studies due to the confounding nature of parental genes.

One solution to this endogeneity problem is to exploit genetic variation between siblings. The idea is that siblings' genetic make-ups are random draws from the same parental genotypes. Thus, any genetic differences across siblings should be independent of any confounder (see the discussion in Conley and Fletcher, 2017). Sibling analyses are thus becoming more common in the literature. For example, Ronda et al. (2020) exploit genetic variation in siblings to document that the effect of genes on education is lower in low-SES families. Hence, they point to an unexploited genetic potential in particular among boys growing up in low-SES families. In comparisons of within- and between-family analyses, sibling analyses have also documented a decline in the direct effect of the child's genes once the family genetic associations are removed (see, e.g., Selzam et al. (2019) and Ronda et al. (2020)). Also, using siblings, Sanz-de Galdeano and Terskaya (2019) study how parental investment decisions depend on the child's EA PGI.

Sibling analyses, however, do not allow for the estimation of family genetic associations. Directly controlling for the parental genes is preferred since it solves the endogeneity problem while allowing for the estimation of the family genetic associations. One example of the value of observing the parental genes in combination with the child's genes is Wertz et al. (2020). They use the British E-Risk cohort study to investigate how the child's and mother's EA PGI affect parenting investments (parenting style) as well as educational achievement at age 18 years. They first find evidence that both the child's and the mother's genes affect parenting investments. Hence, the parents react to the child's genetic learning potential (the nurture of nature effect). They also confirm the presence of a family genetic association, as in Kong et al. (2018), by demonstrating that the mother's EA PGI affects the child's educational attainment after controlling for the child's own EA PGI.

The studies by Young et al. (2018), Kong et al. (2018), Ronda et al. (2020), and, to some extent, Wertz et al. (2020) highlight the importance of controlling for parental genetic influences, either directly (by incorporating the parents' EA PGI into the analyses) or indirectly (by using, e.g., sibling fixed effects designs) for identifying an effect of the child's own genes. Many earlier studies in the field did not have this possibility (and acknowledge it) due to a lack of appropriate data. While the emerging socio-genomic literature is thus beginning to reveal certain partial associations, parental genetic influences, and to some extent interactions of (own and parental) genes with the environment, our understanding of whether and how genetic endowments interact with family resources in the process of human capital formation is still lacking in the deeper theoretical sense outlined above. E.g., do parental investments depend on the child's genetic endowments or are they primarily determined by the parents' genetic endowments? Do parental genetic endowments affect the child in other ways and, if so, what are the channels and mechanisms?

Appendix C Consistency of IV Estimates

This section extends the proof of consistency in Young et al. (2020) to an IV setting. The key observation that makes the imputation also work in our setting is that both the instrument (imputed paternal PGI 1) and the instrumental variable (imputed paternal PGI 2) are imputed using the same information set (child and maternal genotypes).

Let X represent a polygenic index (PGI), which is a linear combination of genotypes and weights from a GWAS (denoted by A):

$$X = \sum_{j=1}^{J} \beta_j^A g_j$$

Similarly, let $\hat{X}$ denote the PGI from the same GWAS but derived from imputed genotypes $\hat{g}_j$, where $\hat{g}_j = E[g_j|\mathbf{g_1}, \mathbf{g_2}, A]$ and $\mathbf{g_1}, \mathbf{g_2}$ are the observed genotypes of two relatives (i.e., two children or a child and a co-parent) and A is a statistic, say the genetic correlation or IBD (identical by descent) segments shared between the different relatives.

Lemma 1: Young et al. (2020) show that:

$$Cov(X, \hat{X}) = Var(\hat{X}) \tag{2}$$

$$Cov(X_n, \hat{X}) = Cov(X_n, X) \tag{3}$$

where X_n is the PGI of one of the two relatives, so $n \in 1, 2$.

Using these two results, Young et al. (2020) show that OLS regression using imputed genotypes is consistent and unbiased (see Theorem 2 in their paper).

Now, we want to show that the consistency extends to a setting where one PGI ($\hat{Z}$) based on imputed genotypes is used as an instrument for the other PGI ($\hat{X}$) based on the same imputed genotypes but weights from a different GWAS, i.e., $\hat{X}$ is defined as before and

$$\hat{Z} = \sum_{j=1}^{J} \beta_j^B \hat{g}_j$$

Let $\mathbf{X} = [X, X_1, X_2]$ and $\mathbf{Z} = [Z, Z_1, Z_2]$ be the matrix of PGIs and $\hat{\mathbf{X}} = E[\mathbf{X}|\mathbf{g_1}, \mathbf{g_2}, A]$ and $\hat{\mathbf{Z}} = E[\mathbf{Z}|\mathbf{g_1}, \mathbf{g_2}, A]$ be the two different imputed PGIs. There are two things worth pointing out. First, $\hat{X_n} = X_n$ for $n \in 1, 2$, since $E[X_n|\mathbf{g_1}, \mathbf{g_2}, A] = X_n$. Second, while the two imputed PGIs ($\hat{X}$ and $\hat{Z}$) are constructed using different weights (B_j^A and B_j^B), they are constructed using the same imputed genotypes and information set ($\hat{g}_j = E[g_j|\mathbf{g_1}, \mathbf{g_2}, A]$).

This latter point is important for the proof below.

Theorem: Let $Y = X\theta + \epsilon$ and $X = Z\lambda + \mu$, where $\epsilon, \mu \perp Z$. Then $\hat{\theta} = (\hat{Z}^T \hat{X})^{-1} \hat{Z}^T Y$ is a consistent estimator of θ provided that $Cov(\hat{Z}, \hat{X})$ is invertible.

Proof: First note that:

$$\begin{aligned}
Cov(\hat{Z}, Y) &= Cov(\hat{Z}, \theta X) && (4)\\
&= \theta(E[\hat{Z}X^T] - E[\hat{Z}]E[X]^T) && (5)\\
&= \theta(E[E[\hat{Z}X^T|g_1, \mathbf{g_2}, A]] - E[\hat{Z}]E[E[X|g_1, g_2, A]]^T) && (6)\\
&= \theta(E[\hat{Z}\hat{X}^T] - E[\hat{Z}]E[\hat{X}]^T) && (7)\\
&= \theta Cov(\hat{Z}, \hat{X}) && (8)
\end{aligned}$$

These follow from the Law of Conditional Expectations and from the fact that $E[\hat{Z}|g_1, g_2, A] = \hat{Z}$.

This result implies that:

$$\begin{aligned}
\lim_{n \to \inf} \hat{\theta} &= Cov(\hat{Z}, \hat{X})^{-1} Cov(\hat{Z}, Y)\\
&= \theta Cov(\hat{Z}, \hat{X})^{-1} Cov(\hat{Z}, \hat{X})\\
&= \theta
\end{aligned}$$

Appendix D Alternative Specification of Empirical Model

Here, we present additional estimates of the empirical model parameters. We consider four different alternative specifications. In our benchmark specification, we correct for measurement error in the genetic factor using a combination of three different polygenic indexes. In Appendix D.1, we report results without this measurement error correction. In Appendix D.2, we report estimates using a translog specification that allows for interactions between all model parameters. Then, in Appendix D.3, we show what happens to the estimates when we gradually add a range of family controls to the model. Finally, Appendix D.4 reports estimates from a skill formation model without the genetic factors.

Appendix D.1 No Measurement Error

Table D8 compares the baseline estimates of the technology of skill formation with estimates using the raw EA PGI (without performing the measurement error correction), while Table D9 performs the same comparison for the investment function. Compared to the baseline specification, these estimates overall appear similar. This suggests that the decision regarding measurement error does not alter the overall implications of our results. Nevertheless, there are some smaller deviations between the results, in particular regarding the direct effect of the child's genes on skill formation. This is made more clear in Figure D1, which compares the relationship between the genetic factors and child latent skills at ages 2-3 and ages 6-7, using the baseline model and the raw EA PGI. As illustrated in Panel B1(a), not correcting for measurement error in the child's genetic factor causes one to underestimate the accumulated effect of child genes on skills at ages 6-7. This shows the importance of performing the measurement error correction.

Table D8: Measurement Error Correction: Technology of Skill Formation

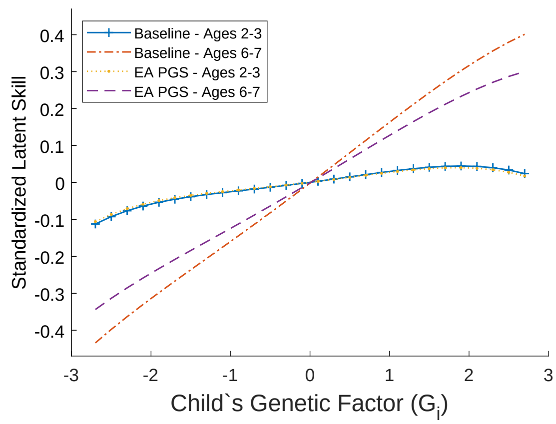
	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline						
G_i	0.002	0.001	0.001	0.030	0.022	0.008
	[-0.015, 0.019]	[-0.008, 0.008]	[-0.016, 0.017]	[0.014, 0.046]	[0.008, 0.036]	[-0.001, 0.017]
G_i^m	0.012	0.003	0.005	0.011	0.011	0.001
	[-0.001, 0.024]	[-0.002, 0.009]	[-0.005, 0.017]	[-0.000, 0.023]	[0.001, 0.021]	[-0.005, 0.008]
G_i^f	0.004	0.004	0.015	0.003	0.006	0.003
	[-0.010, 0.019]	[-0.002, 0.011]	[0.003, 0.029]	[-0.009, 0.016]	[-0.006, 0.018]	[-0.004, 0.011]
$\ln \theta_{it}$		0.224	1.844	0.583	1.009	1.988
		[0.088, 0.351]	[1.214, 2.724]	[0.340, 0.829]	[0.565, 1.486]	[1.175, 2.819]
$\ln I_{it}$		0.098	0.696	0.329	0.931	2.122
		[0.051, 0.140]	[0.386, 1.100]	[0.127, 0.542]	[0.390, 1.525]	[1.107, 3.233]
$\ln \theta_{it} \times \ln I_{it}$		-0.005	-0.216	-0.056	-0.184	-0.552
		[-0.031, 0.024]	[-0.366, -0.101]	[-0.122, 0.010]	[-0.345, -0.028]	[-0.855, -0.263]
$\ln A$	1.409	1.818	-2.375	0.936	-0.808	-3.939
	[1.391, 1.428]	[1.604, 2.038]	[-4.855, -0.652]	[0.187, 1.706]	[-2.525, 0.736]	[-7.011, -1.051]
Panel B: Model w. EA PGI						
G_i	0.006	-0.001	-0.000	0.025	0.019	0.004
	[-0.011, 0.023]	[-0.009, 0.007]	[-0.016, 0.014]	[0.011, 0.040]	[0.006, 0.033]	[-0.004, 0.012]
G_i^m	0.011	0.003	0.009	0.010	0.009	0.003
	[-0.000, 0.023]	[-0.002, 0.009]	[-0.001, 0.020]	[-0.001, 0.021]	[0.000, 0.020]	[-0.002, 0.010]
G_i^f	-0.001	0.003	0.014	0.002	0.006	0.002
	[-0.015, 0.012]	[-0.003, 0.010]	[0.003, 0.027]	[-0.010, 0.013]	[-0.005, 0.017]	[-0.004, 0.010]
$\ln \theta_{it}$		0.225	1.850	0.585	1.014	1.991
		[0.089, 0.353]	[1.223, 2.727]	[0.342, 0.828]	[0.572, 1.488]	[1.179, 2.828]
$\ln I_{it}$		0.099	0.698	0.331	0.933	2.124
		[0.052, 0.141]	[0.389, 1.100]	[0.128, 0.545]	[0.396, 1.528]	[1.107, 3.239]
$\ln \theta_{it} \times \ln I_{it}$		-0.005	-0.217	-0.057	-0.185	-0.553
		[-0.031, 0.024]	[-0.370, -0.103]	[-0.123, 0.010]	[-0.346, -0.029]	[-0.854, -0.263]
$\ln A$	1.409	1.815	-2.384	0.929	-0.814	-3.949
	[1.391, 1.428]	[1.602, 2.036]	[-4.838, -0.657]	[0.174, 1.697]	[-2.529, 0.717]	[-7.020, -1.065]

Notes: This table compares the parameter estimates for the technology of skill formation in our baseline model to the model where we do not control for measurement error in the genetic factor. That is, we use the raw EA PGI constructed with the GWAS weights from Lee et al. (2018) without the 23andMe data. 95% bootstrap confidence intervals in brackets.

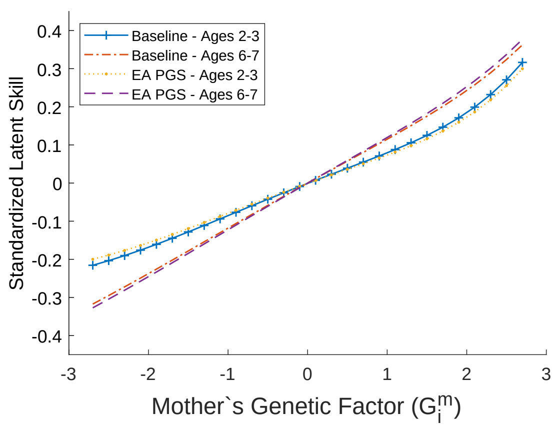
Table D9: Measurement Error Correction: Investment Policy function

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3	Ages 4-5 (4)	Ages 5-6 (5)
Panel A: Baseline					
G_i	0.026	0.022	0.012	-0.002	-0.001
	[0.006, 0.045]	[-0.005, 0.049]	[-0.002, 0.026]	[-0.008, 0.005]	[-0.007, 0.005]
G_i^m	0.038	0.079	0.030	0.019	0.013
	[0.023, 0.054]	[0.058, 0.100]	[0.020, 0.041]	[0.013, 0.025]	[0.008, 0.018]
G_i^f	0.022	0.036	0.019	0.015	0.016
	[0.004, 0.039]	[0.013, 0.058]	[0.007, 0.031]	[0.009, 0.021]	[0.009, 0.021]
$\ln \theta_{it}$	0.339	0.719	0.182	0.091	0.107
	[0.268, 0.405]	[0.597, 0.862]	[0.134, 0.230]	[0.067, 0.115]	[0.076, 0.133]
Constant	4.144	2.650	3.005	2.493	2.416
	[4.047, 4.248]	[2.269, 2.976]	[2.846, 3.159]	[2.402, 2.582]	[2.314, 2.537]
Panel B: Model w. EA PGI					
G_i	0.027	0.025	0.015	-0.000	-0.000
	[0.007, 0.047]	[-0.001, 0.051]	[0.000, 0.028]	[-0.006, 0.006]	[-0.006, 0.006]
G_i^m	0.033	0.073	0.025	0.019	0.012
	[0.019, 0.047]	[0.053, 0.094]	[0.015, 0.036]	[0.013, 0.024]	[0.007, 0.017]
G_i^f	0.019	0.032	0.016	0.013	0.014
	[0.002, 0.036]	[0.010, 0.054]	[0.004, 0.027]	[0.007, 0.019]	[0.008, 0.019]
$\ln \theta_{it}$	0.338	0.717	0.181	0.091	0.106
	[0.267, 0.405]	[0.595, 0.859]	[0.134, 0.229]	[0.067, 0.114]	[0.075, 0.133]
Constant	4.145	2.655	3.008	2.495	2.418
	[4.047, 4.249]	[2.281, 2.982]	[2.849, 3.161]	[2.404, 2.582]	[2.315, 2.537]

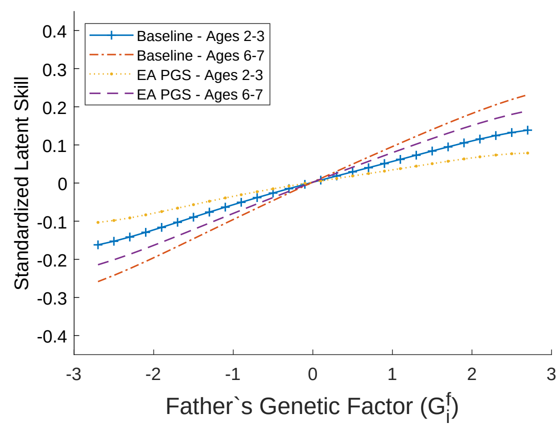
Notes: This table compares the parameter estimates for the investment policy function in our baseline model to the model where we do not control for measurement error in the genetic factor. That is, we use the raw EA PGI constructed with the GWAS weights from Lee et al. (2018) without the 23andMe data. 95% bootstrap confidence intervals in brackets.



(a) Child's Genetic Factor (G_i)



(b) Mother's Genetic Factor (G_i^m)



(c) Father's Genetic Factor (G_i^f)

Figure D1: Associations between EA PGI and Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' genetic factor and the child's latent skill at ages 2-3 and 6-7. It compares the estimated relationship between genetic factors and latent skills before and after correcting for measurement error in the genetic factors. Using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's, the mother's, and the father's genetic factors while keeping the others constant.

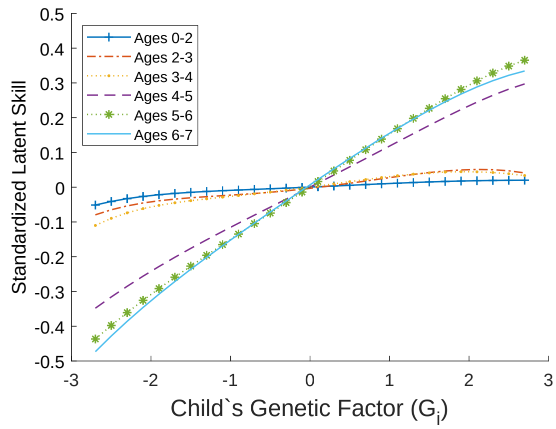
Appendix D.2 Translog

Table D10 displays the estimates of the technology of skill formation if we allow all model parameters to interact. In general, the interaction terms involving either of the genetic factors mostly turn up insignificant in this model. To get a better sense of whether this specification changes our results, we plot the simulated latent skills over time by each of the genetic factors in Figure D2. This shows that the skill development implied by the translog specification is very similar to that implied by the baseline model (illustrated in Figure 3), both in regards to the child's and each of her parents' genetic factors. Furthermore, the relative contributions of the direct genetic effect, the nurture of nature effect, and the family genetic associations are also very similar in both specifications, as evident by comparing Table D11 to Table 7. For this reason, we prefer the simpler baseline specification to illustrate our findings.

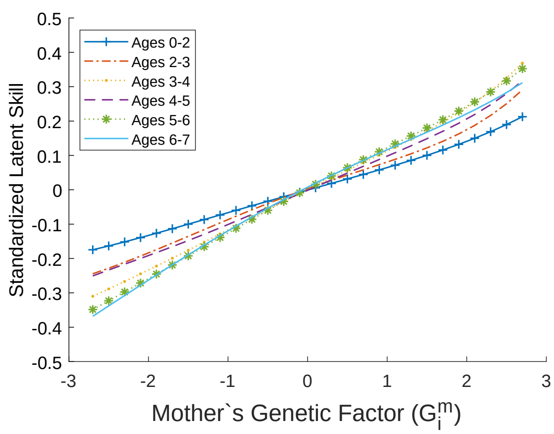
Table D10: Technology of Skill Formation

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3)	Ages 4-5 (4)	Ages 5-6 (5)	Ages 6-7 (6)
G_i	0.002	-0.089	0.067	0.299	0.074	0.154
	[-0.015, 0.019]	[-0.175, -0.004]	[-0.238, 0.373]	[0.101, 0.519]	[-0.264, 0.386]	[-0.176, 0.517]
G_i^m	0.012	0.059	0.030	-0.018	0.187	0.126
	[-0.001, 0.024]	[0.000, 0.117]	[-0.173, 0.240]	[-0.171, 0.132]	[-0.047, 0.437]	[-0.116, 0.371]
G_i^f	0.004	0.099	0.036	-0.178	-0.073	-0.097
	[-0.010, 0.019]	[0.029, 0.170]	[-0.205, 0.276]	[-0.346, -0.004]	[-0.344, 0.202]	[-0.381, 0.176]
$\ln\theta_{it}$		0.232	1.819	0.549	0.893	1.894
		[0.098, 0.363]	[1.168, 2.684]	[0.301, 0.792]	[0.439, 1.383]	[1.083, 2.770]
$\ln I_{it}$		0.100	0.681	0.298	0.796	2.005
		[0.055, 0.143]	[0.367, 1.085]	[0.089, 0.505]	[0.254, 1.397]	[0.973, 3.111]
$\ln\theta_{it}\times\ln I_{it}$		-0.006	-0.211	-0.047	-0.143	-0.519
		[-0.033, 0.023]	[-0.357, -0.096]	[-0.113, 0.020]	[-0.310, 0.016]	[-0.817, -0.228]
$G_i\times\ln\theta_{it}$		0.013	0.014	-0.057	-0.020	-0.022
		[-0.012, 0.040]	[-0.079, 0.111]	[-0.101, -0.015]	[-0.065, 0.021]	[-0.080, 0.034]
$G_i\times\ln I_{it}$		0.015	-0.022	-0.027	0.005	-0.026
		[-0.001, 0.032]	[-0.063, 0.021]	[-0.078, 0.018]	[-0.105, 0.115]	[-0.132, 0.064]
$G_i^m\times\ln\theta_{it}$		-0.005	-0.007	0.010	-0.025	0.012
		[-0.024, 0.014]	[-0.070, 0.056]	[-0.023, 0.039]	[-0.053, 0.005]	[-0.033, 0.052]
$G_i^m\times\ln I_{it}$		-0.010	-0.001	-0.000	-0.034	-0.060
		[-0.021, 0.001]	[-0.033, 0.028]	[-0.034, 0.035]	[-0.117, 0.043]	[-0.141, 0.016]
$G_i^f\times\ln\theta_{it}$		0.000	-0.011	0.039	0.005	-0.001
		[-0.020, 0.021]	[-0.085, 0.064]	[0.005, 0.073]	[-0.028, 0.042]	[-0.046, 0.041]
$G_i^f\times\ln I_{it}$		-0.020	0.002	0.018	0.023	0.037
		[-0.034, -0.006]	[-0.035, 0.035]	[-0.023, 0.058]	[-0.072, 0.111]	[-0.045, 0.126]
$G_i\times G_i^m$		0.003	-0.000	0.005	0.004	-0.003
		[-0.002, 0.008]	[-0.011, 0.010]	[-0.005, 0.015]	[-0.005, 0.013]	[-0.010, 0.003]
$G_i\times G_i^f$		0.000	-0.003	-0.004	-0.005	-0.003
		[-0.004, 0.004]	[-0.012, 0.006]	[-0.012, 0.005]	[-0.014, 0.002]	[-0.009, 0.002]
$G_i^m\times G_i^f$		0.002	0.005	0.001	-0.005	0.001
		[-0.004, 0.009]	[-0.008, 0.018]	[-0.013, 0.013]	[-0.018, 0.006]	[-0.007, 0.009]
$\ln A$	1.409	1.805	-2.304	1.042	-0.418	-3.607
	[1.391, 1.428]	[1.591, 2.020]	[-4.763, -0.521]	[0.306, 1.813]	[-2.161, 1.145]	[-6.781, -0.667]

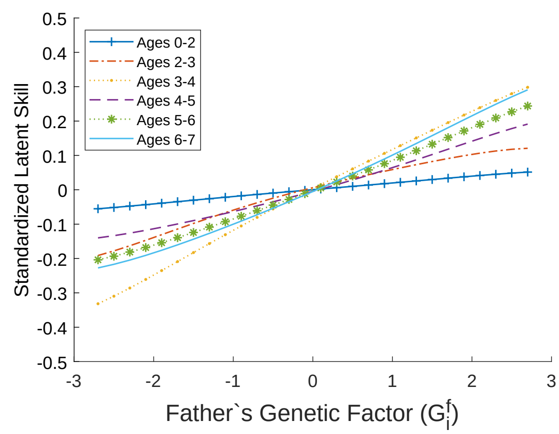
Notes: The parameter estimates for the initial skill function are reported in the first column and for the extended technology of skill formation that allows for interactions between the genetic factors in columns 2-6. 95% bootstrap confidence intervals in brackets.



(a) Child's Genetic Factor (G_i)



(b) Mother's Genetic Factor (G_i^m)



(c) Father's Genetic Factor (G_i^f)

Figure D2: Associations between EA PGI and Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' EA PGI and the child's latent skill at different ages. Using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's, the mother's, and the father's genetic factors while keeping the others constant. This figure highlights how the associations between skills and both parental and child genes increase over time.

Table D11: Mechanisms Decomposition by Age

	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
Panel A:			Child's Skills			
Direct Effects	18.09%	4.34%	3.29%	49.82%	54.87%	54.10%
Nature of Nurture	0.00%	17.75%	12.60%	7.28%	2.87%	2.89%
Family Genetic Associations	81.91%	77.91%	84.10%	42.90%	42.26%	43.01%
Panel B:			Parental Investments			
Nurture of Nature	42.86%	28.25%	37.04%	9.81%	18.54%	
Family Genetic Associations	57.14%	71.75%	62.96%	90.19%	81.46%	

Notes: This table decomposes the association between the child's polygenic index for educational attainment and child's skills (in Panel A) and parental investments (in Panel B) by the three mechanisms for the different developmental periods.

Appendix D.3 Parental Controls - Estimates

Table D12 displays the main parameter estimates (the genetic factors) in the technology of skill formation. Panel A shows the baseline estimates. In Panels B through E, we gradually add a set of parental controls to see how this changes the estimates. We should not expect the estimates on the child's genetic factor to change significantly, as any variation in the child's genetic factor should be random conditional on parental genes. Hence, this serves as a validation check of our estimation strategy. On the other hand, the estimates of the family genetic associations may change if any of the parental controls mediates the relationship between parental genes and child skill formation.

First, adding birth order (Panel B) has little effect on any of the estimates. On the other hand, adding maternal education (Panel C) pushes the coefficients on the mother's genetic factor toward zero. The estimates on the father's genetic factor are also reduced somewhat, though this becomes more clear once we also control for paternal education (Panel D). Adding additional family controls (Panel E) does little to change the estimates further. Importantly, the estimates on the child's genetic factor (the direct genetic effect) are virtually the same in Panel A and Panel E.

In Table D13, we do the same exercise for the investment policy function. We see the same pattern, though much more clearly. The parental genetic factors have a strong association with investments (the family genetic associations). However, after controlling for parental education, the coefficient basically goes to zero. This suggests that the family genetic influences are completely mediated by maternal and paternal educational attainment. Importantly, the estimates on the child's genetic factor (the nurture of nature effect) are again virtually unaffected.

Table D12: TECHNOLOGY OF SKILL FORMATION - PARENTAL CONTROLS

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3)	Ages 4-5 (4)	Ages 5-6 (5)	Ages 6-7 (6)
Panel A: Baseline						
G_i	-0.002	0.000	0.004	0.030	0.021	0.008
	[-0.021, 0.020]	[-0.009, 0.010]	[-0.014, 0.018]	[0.016, 0.046]	[0.007, 0.035]	[-0.002, 0.020]
G_i^m	0.015	0.004	0.004	0.011	0.011	0.001
	[0.000, 0.029]	[-0.002, 0.011]	[-0.009, 0.016]	[0.002, 0.023]	[0.002, 0.021]	[-0.005, 0.009]
G_i^f	0.009	0.005	0.016	0.001	0.006	0.003
	[-0.009, 0.024]	[-0.002, 0.013]	[0.003, 0.030]	[-0.011, 0.014]	[-0.004, 0.018]	[-0.006, 0.010]
Panel B: + Birth Order						
G_i	-0.000	0.000	0.005	0.031	0.021	0.009
	[-0.019, 0.013]	[-0.008, 0.008]	[-0.010, 0.022]	[0.013, 0.048]	[0.007, 0.038]	[-0.000, 0.017]
G_i^m	0.013	0.004	0.004	0.011	0.011	0.001
	[0.003, 0.026]	[-0.000, 0.010]	[-0.006, 0.016]	[0.001, 0.024]	[0.001, 0.020]	[-0.004, 0.010]
G_i^f	0.007	0.005	0.016	0.000	0.006	0.002
	[-0.007, 0.022]	[-0.003, 0.012]	[0.002, 0.027]	[-0.012, 0.013]	[-0.006, 0.018]	[-0.003, 0.010]
Panel C: + Maternal Education						
G_i	-0.001	-0.000	0.004	0.030	0.021	0.009
	[-0.020, 0.012]	[-0.008, 0.007]	[-0.012, 0.020]	[0.012, 0.045]	[0.006, 0.035]	[-0.001, 0.018]
G_i^m	0.006	0.000	-0.003	0.001	0.006	-0.002
	[-0.004, 0.019]	[-0.006, 0.008]	[-0.014, 0.009]	[-0.010, 0.014]	[-0.004, 0.018]	[-0.008, 0.006]
G_i^f	0.004	0.004	0.013	-0.003	0.004	0.002
	[-0.009, 0.020]	[-0.003, 0.012]	[0.000, 0.026]	[-0.013, 0.012]	[-0.006, 0.016]	[-0.007, 0.011]
Panel D: + Paternal Education						
G_i	-0.001	0.000	0.004	0.031	0.022	0.009
	[-0.017, 0.018]	[-0.009, 0.009]	[-0.010, 0.020]	[0.016, 0.049]	[0.007, 0.040]	[-0.002, 0.021]
G_i^m	0.005	-0.000	-0.003	-0.001	0.004	-0.002
	[-0.008, 0.022]	[-0.007, 0.006]	[-0.014, 0.006]	[-0.014, 0.010]	[-0.006, 0.014]	[-0.010, 0.005]
G_i^f	0.001	0.001	0.010	-0.009	0.000	-0.000
	[-0.014, 0.014]	[-0.006, 0.007]	[-0.001, 0.022]	[-0.020, 0.002]	[-0.010, 0.013]	[-0.008, 0.008]
Panel E: + Additional Family Controls						
G_i	-0.001	0.000	0.004	0.031	0.022	0.009
	[-0.018, 0.016]	[-0.008, 0.008]	[-0.011, 0.018]	[0.016, 0.049]	[0.008, 0.037]	[-0.000, 0.019]
G_i^m	0.006	-0.001	-0.004	-0.002	0.003	-0.003
	[-0.008, 0.019]	[-0.007, 0.006]	[-0.015, 0.007]	[-0.013, 0.009]	[-0.007, 0.013]	[-0.010, 0.004]
G_i^f	-0.000	0.001	0.008	-0.013	-0.002	-0.001
	[-0.016, 0.014]	[-0.006, 0.007]	[-0.004, 0.021]	[-0.026, -0.000]	[-0.013, 0.010]	[-0.009, 0.006]

Notes: This table compares the parameter estimates for the genetic factors in the technology of skill formation equations as we change the set of control variables X_{it}. The set of controls in Panel A includes only a sex dummy. In Panel B, we also control for birth order. Panels C and D extend the set of controls to include maternal and paternal education dummies. Panel E reports estimates with the full set of controls, which include a sex dummy, information on birth order, maternal and paternal education, maternal and paternal occupation, and family financial difficulties. 95% bootstrap confidence intervals in brackets.

Table D13: Investment Policy function - Parental Controls

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3	Ages 4-5 (4)	Ages 5-6 (5)
Panel A: Baseline					
G_i	0.024	0.015	0.009	-0.002	-0.002
	[0.002, 0.047]	[-0.016, 0.042]	[-0.003, 0.026]	[-0.008, 0.004]	[-0.008, 0.005]
G_i^m	0.038	0.083	0.034	0.019	0.013
	[0.025, 0.053]	[0.065, 0.103]	[0.024, 0.043]	[0.014, 0.024]	[0.008, 0.018]
G_i^f	0.023	0.038	0.021	0.016	0.016
	[0.005, 0.041]	[0.012, 0.061]	[0.008, 0.032]	[0.008, 0.020]	[0.009, 0.022]
Panel B: + Birth Order					
G_i	0.028	0.019	0.012	-0.001	-0.000
	[0.007, 0.049]	[-0.017, 0.043]	[-0.004, 0.024]	[-0.008, 0.005]	[-0.007, 0.004]
G_i^m	0.035	0.081	0.032	0.019	0.013
	[0.019, 0.052]	[0.059, 0.102]	[0.021, 0.045]	[0.013, 0.025]	[0.007, 0.017]
G_i^f	0.020	0.035	0.020	0.015	0.016
	[0.002, 0.036]	[0.015, 0.057]	[0.009, 0.034]	[0.010, 0.020]	[0.008, 0.021]
Panel C: + Maternal Education					
G_i	0.022	0.011	0.008	-0.003	-0.002
	[0.003, 0.041]	[-0.014, 0.035]	[-0.006, 0.023]	[-0.008, 0.005]	[-0.007, 0.006]
G_i^m	-0.000	0.034	0.008	0.008	0.003
	[-0.016, 0.014]	[0.015, 0.054]	[-0.001, 0.017]	[0.003, 0.013]	[-0.002, 0.008]
G_i^f	0.005	0.015	0.010	0.010	0.011
	[-0.014, 0.020]	[-0.010, 0.040]	[-0.002, 0.022]	[0.003, 0.016]	[0.003, 0.017]
Panel D: + Paternal Education					
G_i	0.024	0.015	0.010	-0.001	-0.000
	[0.003, 0.042]	[-0.014, 0.039]	[-0.005, 0.028]	[-0.007, 0.005]	[-0.006, 0.006]
G_i^m	-0.004	0.029	0.005	0.007	0.002
	[-0.019, 0.013]	[0.011, 0.049]	[-0.006, 0.016]	[0.002, 0.012]	[-0.004, 0.007]
G_i^f	-0.006	0.000	0.001	0.006	0.008
	[-0.020, 0.010]	[-0.017, 0.025]	[-0.012, 0.014]	[-0.000, 0.012]	[0.002, 0.012]
Panel E: + Additional Family Controls					
G_i	0.023	0.014	0.009	-0.001	-0.000
	[0.003, 0.044]	[-0.014, 0.038]	[-0.005, 0.023]	[-0.007, 0.005]	[-0.007, 0.006]
G_i^m	-0.004	0.029	0.005	0.006	0.002
	[-0.019, 0.012]	[0.009, 0.048]	[-0.005, 0.016]	[0.001, 0.011]	[-0.003, 0.006]
G_i^f	-0.008	-0.003	-0.001	0.004	0.007
	[-0.026, 0.009]	[-0.024, 0.020]	[-0.013, 0.011]	[-0.002, 0.010]	[0.000, 0.011]

Notes: This table compares the parameter estimates for the genetic factors in the investment policy function equations as we change the set of control variables X_{it}. The set of controls in Panel A includes only a sex dummy. In Panel B, we also control for birth order. Panels C and D extend the set of controls to include maternal and paternal education dummies. Panel E reports estimates with the full set of controls, which include a sex dummy, information on birth order, maternal and paternal education, maternal and paternal occupation, and family financial difficulties. 95% bootstrap confidence intervals in brackets.

Appendix D.4 No Genes

Finally, we estimate a traditional skill formation model that does not include the genetic factors. These estimates are compared to our baseline estimates in Tables D14 and D15 and to the estimates with the full set of parental controls in Tables D16 and D17. First, comparing the estimates for the technology of skill formation shows that the estimates of the self-productivity of skills and the complementarity between skills and investments are similar in both models. However, the returns to investments are somewhat overestimated in the model without genes, regardless of whether the parental controls are included or not. Comparing the estimates for the investment policy function shows that the baseline model without genes also overestimates the extent to which investments depend on previous skills (the reinforcing behavior) because it does not capture that parental genes are an influence on both skills and investments. However, including the full set of controls makes these parameters very similar.

Table D14: TECHNOLOGY OF SKILL FORMATION - NO GENES

	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline						
G_i	0.002	0.001	0.001	0.030	0.022	0.008
	[-0.015, 0.019]	[-0.008, 0.008]	[-0.016, 0.017]	[0.014, 0.046]	[0.008, 0.036]	[-0.001, 0.017]
G_i^m	0.012	0.003	0.005	0.011	0.011	0.001
	[-0.001, 0.024]	[-0.002, 0.009]	[-0.005, 0.017]	[-0.000, 0.023]	[0.001, 0.021]	[-0.005, 0.008]
G_i^f	0.004	0.004	0.015	0.003	0.006	0.003
	[-0.010, 0.019]	[-0.002, 0.011]	[0.003, 0.029]	[-0.009, 0.016]	[-0.006, 0.018]	[-0.004, 0.011]
$\ln \theta_{it}$		0.224	1.844	0.583	1.009	1.988
		[0.088, 0.351]	[1.214, 2.724]	[0.340, 0.829]	[0.565, 1.486]	[1.175, 2.819]
$\ln I_{it}$		0.098	0.696	0.329	0.931	2.122
		[0.051, 0.140]	[0.386, 1.100]	[0.127, 0.542]	[0.390, 1.525]	[1.107, 3.233]
$\ln \theta_{it} \times \ln I_{it}$		-0.005	-0.216	-0.056	-0.184	-0.552
		[-0.031, 0.024]	[-0.366, -0.101]	[-0.122, 0.010]	[-0.345, -0.028]	[-0.855, -0.263]
$\ln A$	1.409	1.818	-2.375	0.936	-0.808	-3.939
	[1.391, 1.428]	[1.604, 2.038]	[-4.855, -0.652]	[0.187, 1.706]	[-2.525, 0.736]	[-7.011, -1.051]
Panel B: Model w.o. Genes						
$\ln \theta_{it}$		0.225	1.821	0.591	1.027	1.972
		[0.091, 0.338]	[1.139, 2.682]	[0.347, 0.833]	[0.585, 1.479]	[1.137, 2.859]
$\ln I_{it}$		0.100	0.689	0.345	0.973	2.109
		[0.054, 0.137]	[0.339, 1.066]	[0.153, 0.553]	[0.442, 1.570]	[1.029, 3.261]
$\ln \theta_{it} \times \ln I_{it}$		-0.005	-0.211	-0.057	-0.188	-0.545
		[-0.029, 0.025]	[-0.350, -0.083]	[-0.119, 0.007]	[-0.344, -0.033]	[-0.857, -0.239]
$\ln A$	1.409	1.811	-2.345	0.863	-0.950	-3.916
	[1.389, 1.428]	[1.622, 2.037]	[-4.709, -0.425]	[0.069, 1.602]	[-2.655, 0.548]	[-7.134, -0.987]

Notes: This table compares the parameter estimates for the technology of skill formation in our baseline model to the model where genetic factors are excluded. 95% bootstrap confidence intervals in brackets.

Table D15: Investment Policy function - No Genes

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3)	Ages 4-5 (4)	Ages 5-6 (5)
Panel A: Baseline					
G_i	0.026	0.022	0.012	-0.002	-0.001
	[0.006, 0.045]	[-0.005, 0.049]	[-0.002, 0.026]	[-0.008, 0.005]	[-0.007, 0.005]
G_i^m	0.038	0.079	0.030	0.019	0.013
	[0.023, 0.054]	[0.058, 0.100]	[0.020, 0.041]	[0.013, 0.025]	[0.008, 0.018]
G_i^f	0.022	0.036	0.019	0.015	0.016
	[0.004, 0.039]	[0.013, 0.058]	[0.007, 0.031]	[0.009, 0.021]	[0.009, 0.021]
$\ln\theta_{it}$	0.339	0.719	0.182	0.091	0.107
	[0.268, 0.405]	[0.597, 0.862]	[0.134, 0.230]	[0.067, 0.115]	[0.076, 0.133]
Constant	4.144	2.650	3.005	2.493	2.416
	[4.047, 4.248]	[2.269, 2.976]	[2.846, 3.159]	[2.402, 2.582]	[2.314, 2.537]
Panel B: Model w.o. Genes					
$\ln\theta_{it}$	0.345	0.746	0.190	0.097	0.115
	[0.275, 0.409]	[0.628, 0.902]	[0.145, 0.243]	[0.074, 0.120]	[0.083, 0.142]
Constant	4.135	2.581	2.979	2.473	2.388
	[4.038, 4.239]	[2.167, 2.891]	[2.803, 3.124]	[2.385, 2.560]	[2.281, 2.506]

Notes: This table compares the parameter estimates for the investment policy function in our baseline model to the model where genetic factors are excluded. 95% bootstrap confidence intervals in brackets.

Table D16: Technology of Skill Formation - No Genes - All Controls

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3)	Ages 4-5 (4)	Ages 5-6 (5)	Ages 6-7 (6)
Panel A: Baseline						
G_i	-0.001	0.000	0.004	0.031	0.022	0.009
	[-0.018, 0.016]	[-0.008, 0.008]	[-0.011, 0.018]	[0.016, 0.049]	[0.008, 0.037]	[-0.000, 0.019]
G_i^m	0.006	-0.001	-0.004	-0.002	0.003	-0.003
	[-0.008, 0.019]	[-0.007, 0.006]	[-0.015, 0.007]	[-0.013, 0.009]	[-0.007, 0.013]	[-0.010, 0.004]
G_i^f	-0.000	0.001	0.008	-0.013	-0.002	-0.001
	[-0.016, 0.014]	[-0.006, 0.007]	[-0.004, 0.021]	[-0.026, -0.000]	[-0.013, 0.010]	[-0.009, 0.006]
$\ln \theta_{it}$		0.271	1.846	0.577	0.952	2.025
		[0.130, 0.393]	[1.175, 2.737]	[0.310, 0.849]	[0.518, 1.448]	[1.082, 3.056]
$\ln I_{it}$		0.102	0.696	0.254	0.789	2.099
		[0.056, 0.142]	[0.368, 1.099]	[0.026, 0.483]	[0.260, 1.397]	[0.936, 3.363]
$\ln \theta_{it} \times \ln I_{it}$		-0.015	-0.235	-0.055	-0.173	-0.566
		[-0.042, 0.015]	[-0.382, -0.109]	[-0.127, 0.020]	[-0.346, -0.021]	[-0.929, -0.236]
$\ln A$	1.414	1.774	-2.198	1.128	-0.363	-3.867
	[1.363, 1.465]	[1.580, 1.996]	[-4.639, -0.365]	[0.245, 1.966]	[-2.153, 1.150]	[-7.485, -0.609]
Panel B: Model w.o. Genes						
$\ln \theta_{it}$		0.271	1.841	0.585	0.966	2.022
		[0.062, 0.090]	[0.029, 0.065]	[0.045, 0.081]	[0.036, 0.068]	[0.008, 0.028]
$\ln I_{it}$		0.102	0.693	0.263	0.806	2.096
		[-0.007, 0.003]	[-0.035, -0.011]	[-0.028, -0.008]	[-0.020, -0.001]	[-0.018, -0.005]
$\ln \theta_{it} \times \ln I_{it}$		-0.015	-0.234	-0.057	-0.177	-0.564
		[-0.020, 0.020]	[-0.058, 0.021]	[0.020, 0.102]	[-0.037, 0.046]	[-0.027, 0.028]
$\ln A$	1.412	1.774	-2.186	1.093	-0.426	-3.863
	[1.363, 1.463]	[-0.042, 0.014]	[-0.382, -0.109]	[-0.130, 0.017]	[-0.351, -0.024]	[-0.928, -0.235]

Notes: This table compares the parameter estimates for the technology of skill formation in our baseline model with the full set of parental controls to the model where genetic factors are excluded. 95% bootstrap confidence intervals in brackets.

Table D17: Investment Policy function - No Genes - All Controls

	Ages 0-2 (1)	Ages 2-3 (2)	Ages 3-4 (3)	Ages 4-5 (4)	Ages 5-6 (5)
Panel A: Baseline					
G_i	0.023	0.014	0.009	-0.001	-0.000
	[0.003, 0.044]	[-0.014, 0.038]	[-0.005, 0.023]	[-0.007, 0.005]	[-0.007, 0.006]
G_i^m	-0.004	0.029	0.005	0.006	0.002
	[-0.019, 0.012]	[0.009, 0.048]	[-0.005, 0.016]	[0.001, 0.011]	[-0.003, 0.006]
G_i^f	-0.008	-0.003	-0.001	0.004	0.007
	[-0.026, 0.009]	[-0.024, 0.020]	[-0.013, 0.011]	[-0.002, 0.010]	[0.000, 0.011]
$\ln \theta_{it}$	0.272	0.540	0.122	0.061	0.074
	[0.210, 0.324]	[0.447, 0.650]	[0.088, 0.156]	[0.045, 0.076]	[0.051, 0.095]
Constant	4.136	2.937	3.112	2.565	2.494
	[4.035, 4.243]	[2.615, 3.193]	[2.988, 3.238]	[2.503, 2.629]	[2.411, 2.587]
Panel B: Model w.o. Genes					
$\ln \theta_{it}$	0.272	0.541	0.123	0.061	0.075
	[-0.062, 0.036]	[-0.076, 0.069]	[-0.043, 0.033]	[-0.009, 0.027]	[-0.003, 0.030]
Constant	4.132	2.923	3.108	2.562	2.490
	[-0.094, -0.063]	[-0.078, -0.042]	[-0.069, -0.043]	[-0.023, -0.013]	[-0.025, -0.014]

Notes: This table compares the parameter estimates for the investment policy function in our baseline model with the full set of parental controls to the model where genetic factors are excluded. 95% bootstrap confidence intervals in brackets.

References

Barth, Daniel, Nicholas W. Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4):1474–1522.

Behrman, Jere R and Paul Taubman. 1989. "Is schooling" mostly in the genes"? Nature-nurture decomposition using data on relatives." *Journal of Political Economy* 97 (6):1425–1446.

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.

Björklund, Anders, Markus Jantti, and Gary Solon. 2005. "Influences of nature and nurture on earnings variation." In *Unequal Chances: Family Background and Economic Success*. Princeton University Press, 145–164.

Branigan, Amelia R, Kenneth J McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* 92 (1):109–140.

Cesarini, David and Peter M Visscher. 2017. "Genetics and educational attainment." *NPJ Science of Learning* 2.

Chabris, Christopher F, Benjamin M Hebert, Daniel J Benjamin, Jonathan Beauchamp, David Cesarini, Matthijs Van der Loos, Magnus Johannesson, Patrik KE Magnusson, Paul Lichtenstein, Craig S Atwood et al. 2012. "Most reported genetic associations with general intelligence are probably false positives." *Psychological Science* 23 (11):1314–1323.

Charney, Evan and William English. 2012. "Candidate genes and political behavior." *American Political Science Review* :1–34.

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.

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.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan. 2020. "A New Approach to Measuring Genetic and Environmental Effects in Twins Studies." *Working Paper* .

Goldberger, Arthur S. 1979. "Heritability." *Economica* 46 (184):327–347.

Hewitt, John K. 2012. "Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits." *Behavior Genetics* 42 (1):1.

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.

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* .

Manski, Charles F. 2011. "Genes, eyeglasses, and social policy." *Journal of Economic Perspectives* 25 (4):83–94.

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 and Kevin Thom. 2019. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3):1351–1399.

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

Polderman, Tinca J C, Beben Benyamin, Christiaan A de Leeuw, Patrick F Sullivan, Arjen van Bochoven, Peter M Visscher, and Danielle Posthuma. 2015. "Meta-analysis of the heritability of human traits based on fifty years of twin studies." *Nature Genetics* 47 (7):702–709.

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.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm. 2020. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." *The Scandinavian Journal of Economics* .

Sacerdote, Bruce. 2007. "How large are the effects from changes in family environment? A study of Korean American adoptees." *The Quarterly Journal of Economics* 122 (1):119–157.

Sanz-de Galdeano, Anna and Anastasia Terskaya. 2019. "Sibling Differences in Educational Polygenic Scores: How Do Parents React?" *IZA Discussion Paper* .

Selzam, Saskia, Stuart J Ritchie, Jean-Baptiste Pingault, Chandra A Reynolds, Paul F O'Reilly, and Robert Plomin. 2019. "Comparing within-and between-family polygenic score prediction." *The American Journal of Human Genetics* 105 (2):351–363.

Silventoinen, Karri, Aline Jelenkovic, Reijo Sund, Antti Latvala, Chika Honda, Fujio Inui, Rie Tomizawa, Mikio Watanabe, Norio Sakai, Esther Rebato et al. 2020. "Genetic and environmental variation in educational attainment: an individual-based analysis of 28 twin cohorts." *Scientific Reports* 10 (1).

Taubman, Paul. 1976. "The determinants of earnings: Genetics, family, and other environments: A study of white male twins." *The American Economic Review* 66 (5):858–870.

Wainschtein, Pierrick, Deepti Jain, Zhili Zheng, L Adrienne Cupples, Aladdin H Shadyab, Barbara McKnight, Benjamin M Shoemaker, Braxton D Mitchell, Bruce M Psaty, Charles Kooperberg et al. 2021. "Recovery of trait heritability from whole genome sequence data." *BioRxiv* :588020.

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.

Wertz, Jasmin, Terrie E. Moffitt, Jessica Agnew-Blais, Louise Arseneault, Daniel W. Belsky, David L. Corcoran, Renate Houts, Timothy Matthews, Joseph A. Prinz et al. 2020. "Using DNA from mothers and children to study parental investment in children's educational attainment." *BioRxiv, Working Paper* .

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.

Young, Alexander I, Seyed Moeen Nehzati, Chanwook Lee, Stefania Benonisdottir, David Cesarini, Daniel J Benjamin, Patrick Turley, and Augustine Kong. 2020. "Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects." *BioRxiv* .