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

DISCUSSION PAPER SERIES

IZA DP No. 15418

# Genetic and Socioeconomic Achievement Gaps in Elementary School

Mikkel Aagaard Houmark
Victor Ronda
Esben Agerbo
Preben Bo Mortensen
Michael Rosholm

JULY 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15418

# Genetic and Socioeconomic Achievement Gaps in Elementary School

**Mikkel Aagaard Houmark**
*Aarhus University*

**Victor Ronda**
*Aarhus University*

**Esben Agerbo**
*Aarhus University*

**Preben Bo Mortensen**
*Aarhus University and iPSYCH*

**Michael Rosholm**
*Aarhus University and IZA*

JULY 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

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

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

www.iza.org

# ABSTRACT

## Genetic and Socioeconomic Achievement Gaps in Elementary School

Socioeconomic (SES) gaps in academic achievement are well documented. We show that a very similar gap exists with respect to genetic differences measured by a polygenic score (PGS) for educational attainment. The genetic gap increases during elementary school, but only among the low SES children. Consequently, the high PGS children experience the largest achievement growth over the school years, even if they are born in socioeconomic disadvantage. While the SES gaps are partly due to selection into different environments, the high PGS children are simply better at extracting resources from a given environment because of higher conscientiousness and other predispositions.

**JEL Classification:** D10, I24, I26, J24

**Keywords:** child development, academic achievement, genetics, ses gaps, elementary schools, public investments, iPSYCH

**Corresponding author:**
Mikkel Aagaard Houmark
Department of Economics and Business Economics
Aarhus University
Nordre Ringgade 1
8000 Aarhus C
Denmark

E-mail: mhoumark@econ.au.dk

# 1 Introduction

Individuals differ in terms of their skills and abilities (cognitive and non-cognitive), and such variation explains a substantial part of the variation in later socioeconomic outcomes (Carneiro, Crawford, and Goodman, 2007; Kautz et al., 2014). In particular, whether children are born into families of high or low socioeconomic status (SES) is associated with large disparities in parental investments, which in turn has implications for a range of skills and economic preferences (Falk et al., 2021). Therefore, skill gaps tend to open up in early childhood, persist throughout life, and often increase during the school years (Heckman and Mosso, 2014).

While different childhood environments are an important cause of inequality, children also differ with respect to their innate ability to learn in a given environment. Such differences can be traced back to differences in genetic endowments (Plomin and von Stumm, 2018). It is well established that genes matter for a range of socioeconomic outcomes. Decades of twin studies have shown that practically all skills and traits are to some extent heritable (Polderman et al., 2015). More recently, advances in molecular biology have made it feasible to actually measure genetic variation at the molecular level. This has led to a number of studies showing that, just like environmental differences, differences in genetic endowments translate into differences in childhood skills, achievement in school, educational attainment, income and eventually wealth at retirement (Ward et al., 2014; Belsky et al., 2016; Cesarini and Visscher, 2017; Papageorge and Thom, 2019). It has also been shown that genetic skill gaps increase gradually during early childhood, partly because parental investments reinforce initial skill differences (Houmark, Ronda, and Rosholm, 2020).

While we know that both socioeconomic and genetic differences lead to inequality in skills and achievement, we know much less about the interrelation between these two types of (dis)advantage. Are genes and family SES independent determinants of child skill formation, or are they overlapping to the extent that the distinction is of no practical importance for the study of socioeconomic inequality? If they represent different channels, do they also work through different mechanisms? To the best of our knowledge, no research has answered these questions. And if the channels are not independent, what kind of interaction do we see between genes and SES? A few studies have found a positive gene-environment interaction for educational achievement and other outcomes later in life (Papageorge and Thom, 2019; Ronda et al., 2021). However, little is known about how this interaction emerges and whether it is stable or changes over the life course.

In this paper, we focus on the years of elementary schooling. While the childhood environment in the beginning of life is almost entirely determined by parents, this naturally

changes as children grow older. Different forms of childcare arrangements in early childhood may contribute, but generally, it is not until children enter the main educational system that parents cease to be the main providers of investments in the child's cognitive development. This shift could potentially change the dynamic relationship between genes, environments and skills. First, parents and schools may differ in their (explicit or implicit) preferences. For example, it could be that parents are more concerned with fostering the development of children that appear to be more talented[1], while schools are more concerned with preventing low-achieving children from falling too far behind. The opposite scenario is likewise conceivable. Second, entering school may alter the distribution of resources for investments. For example, children from disadvantaged homes may suddenly receive investments of the same quality as their peers from more resourceful homes, although the extent to which such an equalization of investments happens depends on the extent to which there is selection into schools of different quality.

We study variation in academic achievement throughout Danish elementary schools using a sample of individuals that have been genotyped as part of the iPSYCH study (Pedersen et al., 2018). This data can be linked with detailed register data, meaning that we can follow the children over time and obtain a range of different information. In this paper, we choose to focus on the period from 2$^{nd}$ to 8$^{th}$ grade (approximately age 9 to 15) because this allows us to measure achievement biannually using a standardized achievement test administered to all pupils enrolled in public Danish schools.

We document a strikingly similar growing achievement gap with respect to genetic endowments and socioeconomic status. Both gaps are associated with significant differences in achievement already in 2$^{nd}$ grade, and both gaps increase from 2$^{nd}$ to 8$^{th}$ grade. Although the two gaps are strikingly similar and partly overlap, we show that genes and SES represent two distinct and mostly independent sources of inequality. Children of high SES parents perform better in standardized tests partly because they enjoy more favorable environments throughout childhood by selecting into better neighborhoods and better schools. On the other hand, genetic endowments do not affect achievement through such selection. Genes are nevertheless important determinants of achievement, and their importance only increases throughout the school years.

When considering the influence of SES and genes simultaneously, we discover an important interaction: While genes are more strongly associated with 2$^{nd}$ grade achievement

[1]That children with high genetic potential receive more investments is not necessarily a result of parental preferences. It may be that these children actively seek out interactions with parents that foster cognitive development. This is known as *active* gene-environment correlation, as opposed to *reactive* gene-environment correlation where the parents are the initiators (Plomin, DeFries, and Loehlin, 1977).

among the high SES children, the increase in the importance of genes from 2nd to 8th grade is mostly driven by the low SES children. This shows that favorable genetic endowments can eventually lead to high achievement even in the face of socioeconomic disadvantage throughout childhood. At the same time, genetic and socioeconomic disadvantages compound such that children growing up in disadvantageous environments and with lower genetic scores tend to increasingly lag behind their peers as they progress through elementary school.

We provide a potential explanation for the strong influence exerted by genes on achievement. Genetic advantage is not related to selection into better schools, nor to lower school absence or any other objective school investment measures. However, the genetic endowments related to achievement also lead to children having higher levels of conscientiousness and a better experience of the school learning environment. Hence, the genetic channel is not independent of the environment, but the relationship is not one of selection but one of *extraction*, with certain genetic endowments making individuals better able to extract resources from the environment. When growing up in relatively disadvantaged families, this limits the amount of resources that can be extracted, leading to an unexploited potential among part of the low SES population. But when continuously exposed to relatively high-quality public Danish elementary schools, these children are nevertheless able to partly overcome their initial disadvantage.

## 2 Measuring genetic effects

That genes matter for a variety of socioeconomic outcomes has been demonstrated empirically by decades of research within behavioural genetics. First, a long range of twin and adoption studies have sought to compare the relative importance of genetic and environmental factors. In twin studies, the observation that monozygotic twins are more similar than dizygotic twins with respect to almost any outcome is evidence for the importance of hereditary factors (Polderman et al., 2015). Similarly, adoption studies show that for many psychological outcomes (e.g., skills and personality), adopted children are more similar to their biological parents than to their adoptive parents, confirming the important role of genetic factors (Plomin et al., 1997; Bouchard and Loehlin, 2001).

While this literature has been an important first step for research into genetic effects within the social sciences, its indirect approach relies on several assumptions, e.g. about the genetic relationship between parents (mother and father or biological and adoptive parents), which may bias these estimates of heritability. Moreover, by only comparing the relative share of variance in an outcome that can be traced back to either genes or the environment,

this approach is ill-equipped to understanding the different ways that nature and nurture may interact in human development. Consequently, these results are generally uninformative about social policy (Goldberger, 1979; Manski, 2011).

Recent advances in molecular biology have paved the way for a new research agenda to emerge. As improvements in technology have substantially reduced costs, it has become feasible for social studies to include measures of genetic endowments at the molecular level for large samples of individuals. To study the effect of genes directly, the common approach within this new literature is to summarize all of an individual's genetic endowments in a so-called polygenic score (PGS). The PGS is outcome-specific. In the present paper, we use a PGS for educational attainment (EA). The EA PGS is a linear combination of each distinct genetic endowment, weighted by the empirical association between EA and that genetic endowment at each genetic site (for more conceptual details on genetics, see Appendix A). Thus:

$$pgs_i^{EA} = \sum_{s=1}^{S} \beta_s^{EA} g_{is}, \tag{1}$$

i.e., the EA PGS of individual $i$ is a linear combination of the individual's genetic endowments, $g_{is}$, weighted by the so-called GWAS regression coefficients, $\beta_s^{EA}$. These weights are derived from a genome-wide association study, which follows an exploratory approach where the outcome of interest is regressed on each individual endowment. We use an EA PGS based on the most recent GWAS for EA by Lee et al. (2018), which includes 1.1 million individuals. Because the influence of each endowment will be very small, measurement error is important, and reducing it requires a large samples. As sample sizes have increased, so has the explanatory power of the PGS, and the present EA PGS explains around 11 percent of the variation in educational attainment in the GWAS sample. This score is thus our best estimate of the variation in propensity for educational attainment that can currently be explained by an individual's genetic endowments Lee et al. (2018).

Polygenic scores for EA have been shown to be predictive of not just educational attainment itself, but a range of socio-economic outcomes both pre- and post-educational attainment. These include childhood skills and school achievement as well as earnings and wealth (also after controlling for EA) (Ward et al., 2014; Belsky et al., 2016; Papageorge and Thom, 2019). It is not surprising that the EA PGS predicts a range of life outcomes. Many of the skills that are important for educational attainment will naturally also be relevant for these other outcomes. Importantly, this also means that the EA PGS will include a combination of genetic endowments important for both cognitive and non-cognitive skills. The

returns to genes have also been shown to depend on the context. Such a gene-by-environment interaction is found by (Papageorge and Thom, 2019), who consider heterogeneity by childhood socio-economic status (SES) and find that the EA PGS is more predictive of college graduation for high-SES children.

However, the genetic effects estimated by these studies include more than the effect of an individual's own genes. Parental genes determine child genes, and this causes the PGS of the child and its parents to be correlated. If parental genes also affect the child's EA in other ways, such as through an effect on the home environment, this will partly be captured by the child's EA PGS. This family genetic effect has been shown to be important in its own right for the EA PGS (Kong et al., 2018).

To estimate the independent effect of an individual's genes and take full advantage of the natural experiment created by the inheritance process, it is necessary to eliminate the confounding effect of parental genes. Conditional on parental genes, variation in the individual's genes is exogenous. One way to exploit this is to use genetic variation between siblings. Because the difference in the PGS between two full siblings will be independent of parental genes, this family fixed effects approach isolates the effect that is due to differences in the siblings' own genes. This approach has previously been used by, e.g., Ronda et al. (2021), who find a positive interaction effect between genes and SES on educational attainment.

A similar approach is to directly control for parental genes. This approach is used by Houmark, Ronda, and Rosholm (2020), who study childhood cognitive development and confirm that the family genetics effect is important and comparable in magnitude to the direct genetic effect. However, they also show that the family genetics channel is completely mediated by observable family characteristics, in particular, parental education. This suggests that simply controlling for some set of family SES variables is sufficient to isolate the effect of the child's own genes.

Houmark, Ronda, and Rosholm (2020) also document that part of the effect of the child's EA PGS works through parental investments, as parents invest more in children with a higher genetic propensity for education, thereby augmenting existing skill gaps. This effect, and the fact that genetic endowments also make some children better at learning new skills conditional on their current stock of skills, cause the effect of genes to materialize gradually over the first seven years of the child's life. An open question is whether this pattern continues as children grow older, or whether the effect of genes stabilizes during elementary school. Since the effect of genes depend on how investments are allocated, a different pattern may emerge as investments shift from being mainly determined by parents to being to a larger extent removed from the childhood home, i.e., happening in the school environment.

Another empirical challenge that pertains to studying genetic effects dynamically is to obtain repeated measures of the same outcome. In Houmark, Ronda, and Rosholm (2020), this is circumvented by incorporating a variety of child development measures into a dynamic factor model that identifies latent child skills and controls for measurement error. In the present paper, we are able to overcome the challenge directly because we are able to link genetic data to data on standardized reading tests taken in various grades through elementary school. This provides us with an outcome that is designed to measure the same underlying skills in different grade levels. Thus, we can interpret any change in the estimated genetic effects over time as being due to the dynamic properties of the genetic effects themselves, rather than being simply an artifact of how the skills are measured. In the next section, we describe the data used for this purpose in more detail.

# 3 Data

## 3.1 The iPSYCH sample

"The Lundbeck Foundation Initiative for Integrative Psychiatric Research"- iPSYCH - is a Danish national project aimed at understanding how interaction between hereditary factors and the environment causes mental disorders. The project has established a large case-cohort sample (iPSYCH2012) for which DNA has been extracted from neonatal blood spot samples from the Danish National Biobank. The original sample included 86,189 genotyped individuals, but was later expanded to 146,591 individuals (iPSYCH2015), born between 1981 and 2008. 93,608 of the individuals make up a selected sample with at least one psychiatric diagnosis, while the remaining individuals are a randomly and representative control sample (Pedersen et al., 2018).[2]

The iPSYCH sample can be matched to the Danish administrative population registries, which contain longitudinal information on a range of socioeconomic outcomes for the full population. This allows us to follow the individuals in our sample throughout elementary school, and at the same time include a rich set of background variables in the analysis.

## 3.2 The Danish national tests

We begin by discussing the main outcome of interest, as its availability has implications for the sampling strategy.

[2]We rely on the whole sampele for our main empirical analysis, but also test for robustness of our key results to using only the representative sample in Section 5.4.

In 2010, a new mandatory test system was introduced in all Danish public schools in order to obtain a standardized way of evaluating academic performance, both at the individual and the national level. The tests are IT-based and adaptive, with the test score being determined only from the answers the student selects on a computer. The questions are automatically adjusted in difficulty, such that if the student answers a question correctly, the next question will be slightly more difficult. Using a Rasch model algorithm, the standard error of measurement is computed throughout the test, and the test may only be terminated once this error is below a certain threshold[3].

The students are tested in seven different subjects at various times throughout school, starting from 2$^{nd}$ grade, where the children are 8-9 years old. We focus on the reading tests, as these are the most frequent, taking place in grades 2, 4, 6 and 8. The tests consist of three different "profile areas"; text comprehension, language comprehension and decoding. We create a summary score over the three areas, as is often done for assessment. This is done by first standardizing the score for each profile area by school year. The final reading score is the average of the three profile area scores and standardized again to having mean zero a standard deviation of one. In a similar fashion, we also consider the tests in mathematics, although these are only carried out in grades 3 and 6 and, from 2018, grade 8.

Even though the tests are designed to measure specific language skills, which are usually thought of as components of cognitive ability, doing well in the tests likely also depends on non-cognitive skills such as conscientiousness. The same is true for educational attainment, although this is a more general outcome. Thus, it is plausible that all of the skills that are relevant for the tests we consider are also relevant for educational attainment in general and hence will be picked up by the EA PGS. Indeed, such test scores are arguably direct precursors to actual educational attainment. Importantly, the nature of the tests (standardized and repeated) allows us to investigate the changing importance of genes during the elementary school years.

## 3.3 Sample selection and descriptive statistics

Because our main outcome only becomes available from 2010, we can only use the latest cohorts in the iPSYCH sample. Our main sample includes all individuals of European descent[4] who complete at least one test, which may happen between 2$^{nd}$ and 8$^{th}$ grade. This

[3] This feature helps to alleviate the concern that skills may be less precisely estimated at early ages.

[4] The European descendants make up 88.9% of the full sample. This selection on ancestry is necessary because the GWAS that we use is performed on European descendants only. This is a major issue of equity in the genetics literature that we hope future research will address (for a discussion on this, see Martin et al. (2017)).

includes 40,136 individuals born between 1994 and 2008 for whom we observe a total of 87,917 reading test outcomes. Of these individuals, 24,062 are from the iPSYCH psychiatric case sample, while the remaining 16,074 are randomly selected controls. In the results section, we show that our main findings are robust to using only the controls. We also utilize a smaller sample consisting of families where we observe the outcome for at least two full siblings. This allows us to interpret variation in the PGS as exogenous. The sibling sample includes 2,705 individuals (2,013 cases and 692 controls).

Table 1 displays summary statistics for child birth characteristics and family background characteristics for the different samples. The family characteristics are measured over the first eight years of the child's life, which is the period up until 2nd grade where the first outcome is observed. They include the log of average family income, maternal and paternal years of education, whether the mother and father has ever received a psychiatric diagnoses, and the fraction of time that the family has been intact (parents not divorced).

Unsurprisingly, the control sample is positively selected compared to the full sample. Their parents have somewhat longer educations, are less likely to have a psychiatric diagnosis and more likely to remain together. The children in the sibling sample, on the other hand, have generally similar family backgrounds as the children in the main sample, particularly with respect to income and education. The exception is that the siblings are more likely to be from families with psychiatric diagnoses, which is natural, given the over-sampling of psychiatric cases, which is correlated between siblings. Overall, given that we observe relatively weak selection, and it goes in different directions, if we observe the same pattern across samples it is likely representative of a more general pattern.

## 3.4 Socioeconomic status

By socioeconomic status (SES), we understand the socioeconomic qualities manifested in the home that the child is born into. Various family background characteristics may be informative about the home environment. In particular, it is well established that parental education and income is positively associated with parental investments and the quality of the childhood environment more broadly, and hence, matter for child development (Bradley and Corwyn, 2002; Bornstein et al., 2003). Measures of parental education and income have therefore traditionally been used to capture differences in family SES and childhood (dis)advantage. Following this tradition, we create an SES index by ranking all families in terms of their average income and education, taking the average of this and standardizing it to have mean zero and a standard deviation of one within our sample. This measure is used in the analysis as a proxy for home quality to investigate achievement gaps with respect to

different childhood home environments.

## 3.5 Polygenic scores

The measure we use to capture variation in the genetic propensity for education is the EA PGS. As noted in Equation 1, the EA PGS is constructed from all individual endowments based on the weights from the most recent GWAS on educational attainment (Lee et al., 2018). The study uses a combined sample of 1.1 million individuals of which, importantly, the iPSYCH sample is not part. Like the SES index, the EA PGS is standardized to have mean zero and a standard deviation of one within our main sample.

Because genetic endowments are a random draw from the parental genetic pool for each birth, we do not expect the EA PGS to be correlated with gender, birth order or birth year. Any such relationship would be due to selection. This could arise, for example, if the EA PGS of the first born child affects the probability of having a second child. Fortunately, all of these correlations are below 0.01. We include these variables in all regressions as they may increase precision, and if they induce a small change in the estimates, it is an adjustment for selection.

Contrarily, we expect the polygenic score to be moderately correlated with the parental characteristics. Not surprisingly, the EA PGS is most strongly correlated with maternal and paternal years of education ($r = 0.26$). Parents who have a strong genetic propensity for education end up with higher education, and they pass on their genetic endowments to their children. Therefore, we expect that controlling for these family characteristics will reduce the association between the child's EA PGS and the outcome of interest through eliminating this family genetics channel.

Figure 1 illustrates how the distribution of the child's EA PGS changes across our summary proxy for the home environment, the parents' SES rank. Although children of high-SES parents tend to have higher polygenic scores, there exists a large overlap of children with similar PGS who are at opposite ends of the SES scale. In other words, it is far from being the case that children with high genetic potential are also necessarily born into families with favorable socioeconomic characteristics or vice-versa. Thus, there is ample variation which can be used to study how genes interact with the childhood environment.

Figure 2 illustrates a related, but maybe more surprising point. The figure plots the distribution of the EA PGS of later born children by the decile of their first born sibling's EA PGS. Even though the polygenic scores of full siblings are the same in expectation, there is large variation in the extent to which each child inherits the genes that are associated with educational attainment. While children with higher scores naturally tend to have siblings

with higher scores as well, there are also children with high scores who have siblings with scores in the bottom decile of the distribution, and vice versa. And it is not uncommon for the within-family difference in EA PGS to have a magnitude of one standard deviation or more - in fact, this is the case for more than one fourth of all sibling pairs. Hence, being born to the same parents far from entails that one is born with similar genetic endowments.

### 3.6 School environment

We use three measures to capture differences in school environment. First, from the absence registries, we obtain the total fraction of school days that each individual has been absent. We include all absence regardless of the reason given. Second, from the education registries, we obtain information on the age at which each individual starts in 1st grade. While the default school starting age is determined by date of birth, parents have some discretion in choosing an earlier or later starting age if preferred.

Finally, to capture variation in school quality, we follow the approach of Gensowski et al. (2020) and exploit information on teachers. Using the Danish population, institution and workplace registries, we are able to link individual teachers to schools. We then obtain the following teacher characteristics: High school grade point average (GPA), age, experience, tenure, fraction of the year on sick leave, and fraction of the previous ten years spent unemployed. For each school, we calculate the average teacher characteristics by year and use them to predict the national test scores obtained at each school in a given year. The predicted test scores are then used to create a summary measure ranking each school in terms of teacher quality.

### 3.7 School experiences

Finally, we also exploit data from the compulsory national school well-being surveys. Implemented in 2015, this annual questionnaire is given to all elementary school children and asks about various aspects of their well-being and their school life. From the items, we construct six overall measures of school experience, see Table 6. We follow Niclasen, Keilow, and Obel (2018), who propose a four-factor structure capturing learning self-efficacy, school connectedness, learning environment and classroom management. We then also follow Andersen et al. (2020), who show that some of the items included in the self-efficacy factor can be used to measure conscientiousness. Conscientiousness is a non-cognitive skill that is strongly related to academic achievement through being, e.g., self-disciplined and hard-working. It is thus captured by questions relating to the ability to concentrate on a task and manage

commitments.

We instead term the first overall factor "academic thriving". This is then decomposed into two sub-factors, namely conscientiousness and achievement, with the latter including questions asking specifically about how well the child is doing academically. The second overall factor, school connectedness, is the measure most directly related to well-being, while also reflecting a sense of belonging and emotional stability (feeling safe and accepted). The third factor, learning environment, captures whether the teaching and the environment is experienced as motivating and the quality of the student-teacher interactions. Finally, classroom management captures the extent to which the classroom is conducive to learning in terms of the teacher's presence and ability to display authority in the class.

# 4 Estimation

To estimate the achievement gaps, we perform a series of regressions linking the reading test scores to the PGS and various controls (including SES). Our baseline model is described by the following equation:

$$\hat{\theta}_{ijt} = \beta_{jt} PGS_i + \delta_{jt} Z_i + \mu_{ijt} \tag{2}$$

where $\hat{\theta}_{ijt}$ is the test score of individual $i$ in subject $j$ in grade $t$, $PGS_i$ is the EA PGS of each individual, and $\mu_{ijt}$ is a random error term. $Z_i$ includes a set of individual-specific covariates. These always include dummies for year of birth, gender and birth order as well as the first ten ancestry genomic principal components. The latter are always added to estimations using polygenic scores to account for potential population stratification bias.

To consider the PGS and the SES gaps simultaneously, we include the SES index in $Z_i$. In some cases, we also include a wider set of controls and fixed effects to more fully control for the effect of the childhood environment. These include family controls for parental income and, for each parent, number of years of education and an indicator for ever receiving a psychiatric diagnosis, as well as the fraction of time the parents have been cohabiting up until the child is in 2$^{nd}$ grade. Broader controls also include the school quality index described in Section 3.6, as well as fixed effects at the municipality and school levels.

As explained, the child's genetic endowments are determined by parental genes, which are in turn correlated with the environment. The child's genetic endowments are only exogenous conditional on parental genetic endowments. However, as shown in Houmark, Ronda, and

Rosholm (2020), such family genetic effects are more or less completely mediated by differences in family characteristics, in particular, by parental education. Hence, when including the full set of family controls, we should obtain a good approximation of the direct influence of the child's own genes.

In addition, we also employ a superior strategy for identifying the independent effect of the child's genes. This is done by estimating Equation 2 on a sample of siblings who are both observed taking a specific reading test. By then adding fixed effects at the family level, this model estimates the genetic achievement gap using only within-family variation. Because the siblings share the same parents, any difference in their genetic endowments is random. This specification thus fully exploits the natural experiment created by the genetic inheritance process. However, as it comes at the cost of much lower power, it serves primarily as a robustness check of our main results.

In addition to estimating the genetic achievement gap at a particular grade level, we are also interested in whether the gap changes significantly over time. Therefore, we additionally apply the following specification (for both the between- and the within-families model, here shown only for the former), where we pool the test scores in reading over all grade levels:

$$\hat{\theta}_i = \beta_0 t_i + \beta_1 PGS_i + \beta_2 PGS_i \cdot t_i + \delta Z_i + \mu_i \tag{3}$$

This increases precision and makes $\beta_1$ and $\beta_2$ directly interpretable as estimates of effect of the child's genes close to school entry[5] and the linear approximation of how this genetic effect changes throughout elementary school.

# 5 Results

## 5.1 The SES and the PGS achievement gap

We begin by illustrating the raw achievement gaps by children with different genetic endowments and from families with different socioeconomic backgrounds. We do so by plotting the average reading test scores for each quintile of the PGS/SES distribution over time. This graphical evidence is shown in Figure 3. We see that there are significant gaps in achievement with respect to both genes and SES. In fact, the two gaps look strikingly similar. In both cases, the children in the top quintile score on average around 0.7 standard deviations higher on the 2nd grade reading tests than the children in the bottom quintile. And over time,

[5] As we do not observe skills right at school entry, the interpretation of $\beta_1$ as the genetic effect at school entry requires that the interaction effect is extrapolated somewhat outside the interval covered by the data.

the gaps increase to more than 0.8. This shows that both the PGS and the SES measure captures significant determinants of inequality in achievement.

Are we simply putting different names on the same variation? To answer this question, we include the two measures separately and in combination in a regression predicting 8th grade achievement. The estimates are displayed in Table 2. Columns 1 and 2 replicate the finding that the magnitudes of the SES and PGS gaps are strikingly similar. A one standard deviation increase in either is associated with a 0.3 standard deviation increase in the 8th grade reading test score, and either measure by itself explains 8.0% of the total variance in test scores. In column 3, including both predictors simultaneously reveals that, although they overlap, the gaps are mostly independent. Holding SES fixed, a 1 SD increase in the PGS still leads to a 0.23 SD increase in 8th grade achievement (and vice-versa). The share of the total variance in test scores explained also increases by more than 50 percent when including both measures rather than only one or the other. This shows that the SES and PGS gaps are not simply two sides of the same coin.

Even if making separate contributions to the observed inequality in achievement, SES and PGS differences may still work through similar channels. One possibility is that genetic and socioeconomic advantage leads to selection into more favorable environments. We test this in Table 2 columns 4-7 by incrementally adding a set of controls or fixed effects capturing some aspect of the environment. We start with the broadest aspect, adding fixed effects at the municipality level in column 4. If, for example, highly educated parents tend to live in certain municipalities, which in turn are more conducive to child academic development, this would explain part of the SES achievement gap. The same may be true for the PGS gap, if such selection happens on genetic potential. In Table 2, column 4, we see that both achievement gaps shrink somewhat after adding the fixed effects, but the SES gap shrinks substantially more.

Next, we control for the measure of school quality described in Section 3.6. School quality may constitute another selection mechanism. For example, children whose genes cause them to show early potential for education may be more likely to attend better schools later on. However, as shown in Table 2, column 5, we see no evidence of this, as the EA PGS-coefficient is completely unaffected by controlling for school quality. On the other hand, the SES gap again shrinks significantly. This shows that part of the reason why differences in family SES translates into differences in child achievement is through selection into better schools. In column 6, we also add fixed effects at the school level to eliminate all between-school variation. This again has a large impact on the SES coefficient but only a small impact on the PGS coefficient.

Finally, in column 7, we add the additional family controls described in Section (excluding education and income, which make up the SES measure). This also has some effect on the SES gap, but no effect on the PGS gap. Overall, we see that the 8$^{th}$ grade SES achievement gap shrinks by 25% when adding the full set of potential mediators. At the same time, the PGS achievement gap shrinks by only 4%. This result reveals that a major reason why an advantageous family environment is beneficial is that it tends to be accompanied by a favorable childhood environment more broadly. On the other hand, having a genetic advantage is only very weakly associated with the broader environment captured by these mediating variables.

## 5.2 Interactions with the PGS achievement gap

Next, we move on to considering the relationship between the EA PGS and test scores in more detail. First, we consider the change in the genetic gap over time, and then, we consider how this pattern varies across SES. In Table 3, we start by showing how the genetic achievement gap develops through grades 2-8. Panel A displays the coefficients from the baseline model, which only controls for sex, birth year, birth order and the first ten principal components of the genetic matrix (to control for population stratification). The estimates echo the graphical representation, with the major part of the PGS gap being present already in 2th grade, although the inequality in achievement increases further between grades 2 and 8, with most of the increase happening during the earlier years. In Panel B, we then show that this pattern is still present when we control for the full set of family controls. This suggests that initial differences in genetic endowments only become more important over time for explaining variation in academic achievement.

Although the family controls should capture most of the association between genes and achievement that is not directly related to the child's own genes, there may still be confounding influences of the environment included in the estimates. To perform a stronger test of whether the pattern we observe is actually due to differences in child genes, we estimate the within-families model described in Section 5.1 using siblings. These estimates, displayed in Table 3, Panel C, show that the pattern of increasing importance of child genes over time is still present. In fact, it becomes even more pronounced, although the confidence intervals are wide. Importantly, the within-families model does not reject any of the coefficients from Panel B and confirms that there is a significant increase over time in the effect of the PGS on the reading test scores. The estimates in Table 3 thus altogether confirm the graphical evidence that there are significant genetic achievement gaps, and that the gaps only increase during elementary school.

Next, we consider whether the genetic gap differs depending on the childhood environment. In Table 4, we therefore estimate the initial gap and the increase in the gap over time separately for children from low and high SES families. For simplification, we describe the increase over time by its linear approximation. The estimates reveal a strong heterogeneity in the evolution of the genetic gaps during elementary school: The increase in the importance of genes with grade is completely driven by low SES children. Their PGS matters less for reading achievement initially, but the effect increases gradually, while for the high SES children, the effect is initially large and stays at this level throughout the period. This finding does not change when we add the family controls or school fixed effects to the regressions. In Appendix Table B1, we show, first, that the difference across SES is highly significant when estimating this jointly in a three-way interaction with the PGS and the grade level (colums 1-3). Second, we show that the estimates we obtain using the within-families model are also consistent with this pattern (columns 4-5).

While differences in genetic endowments are thus initially reinforced by a favorable childhood environment, it appears that schooling compensates to some extent for childhood disadvantage, as the children with a high PGS over time come to realise this potential. This shows the importance of genetic inheritance, as favorable endowments allow some children to do well despite exposure to unfavorable environments throughout childhood (i.e., at home, in the neighborhood, and in school). At the same time, substantial SES gaps remain. This also implies that the children who are disadvantaged both genetically and socioeconomically become worse and worse off during elementary school relative to their peers.

## 5.3 Mediators of the PGS achievement gap

So far, we have shown that both genetic and socioeconomic advantage during childhood translates into higher academic achievement. We have also shown that a favorable childhood environment is associated with a favorable environment throughout childhood; in other words, the SES gap works partly through selection into better neighborhoods, better schools etc. The same is not true for the PGS gap. This leaves the question of why genetic endowments exert such a strong influence on academic achievement. In this section, we present some suggestive evidence on potential mechanisms.

While the PGS achievement gap does not appear to be mediated by selection into different environments in general, it could still influence achievement through investments at a more individual level. In Table 5, Panel A, we investigate how the PGS relates to the set of schooling investments that may vary by individual. In columns 1, 3 and 5, we see that children with a higher EA PGS tend to have lower school absence, earlier school starting

age and attend higher quality schools. However, this may simply reflect that the high PGS children are more often from high SES families. Indeed, when we estimate the causal relationship between EA PGS and the school investments using the within-families model, there is no association with either of the outcomes. This again rules out that the PGS gaps are mediated by school investments, even if the investments considered are under more individual control. For example, it is not because they have less school absence that the high PGS children have higher achievement.

So far, we have shown that the genetic achievement gap is practically orthogonal to a range of different characteristics of the environment. We now consider outcomes relating to the child's subjective experiences instead. We do so by using the items from the national well-being survey as explained in Section 3.7. The full set of items for each measure is listed in Table 6.

In Table 5, Panel B, we see the relationship between the EA PGS and the first factor, academic thriving, as well as its two sub-components, achievement and conscientiousness. Not surprisingly, there is a strong association between the EA PGS and each outcome in the between-families model (columns 1, 3 and 5). But, perhaps surprisingly, the association is even stronger for conscientiousness than for achievement. Crucially, we see that the associations are all robust to adding the family fixed effects - in fact, they all increase somewhat in the within-families model. This reveals a direct relationship between the PGS and the survey items: Children with higher PGS tend to have higher test scores, and they also feel more successful academically. They are also more conscientious, and this non-cognitive skill gap is almost as large as the achievement gap[6].

In Table 5, Panel C, we investigate the associations between the PGS and the other three well-being factors. While not as strong as for the academic factor, we also see a significant positive association between the PGS and each of the other factors in the between-families model. Again, the point estimates are also of a similar magnitude in the within-families model, although we can only point to a significant relationship between the PGS and the learning environment. This implies that the genetic endowments that lead to higher academic achievement also lead to the children being more motivated (e.g., finding the lessons more exciting), and to at least a subjective experience of receiving more teacher support. These results reveal that the genetic achievement gap overlaps with genetic gaps in conscientiousness, academic motivation and the quality of student-teacher interactions.

[6]It is very possible that higher SES also contributes independently to, e.g., higher conscientiousness. We cannot test this directly because there is no within-family variation in SES.

## 5.4 Robustness

In this section, we show that the main findings are robust to other choices regarding the sample selection process and the measurements. First, to include as many observations as possible, our main sample uses all the test scores that are available for the specific cohorts. Hence, we do not necessarily observe the same individuals taking all the tests. As a result, individuals for whom we observe the outcome in 8th grade will on average be born earlier than individuals for whom we observe 2nd grade test scores. Although we control for year of birth in all regressions, it is possible that the increasing association between genes and test scores is due to compositional changes in the sample.

Table B2 reveals that this is not the case. We still observe a significant increase in the association between the EA PGS and test scores from 2nd to 8th grade, from 0.222 to 0.276, i.e., a 0.054 increase over six grade levels, very similar to the 0.056 increase in Table 3. Furthermore, when adding family controls, all the coefficients shrink by around 30 percent, which is again very similar to the findings for the main sample. Hence, the genetic gaps do grow over time, also when we follow the same children throughout elementary school.

Another potential concern is that our main estimation sample, because of the original purpose of the iPSYCH study, over-samples psychiatric cases. However, the main findings are also similar if we use only the control part of the sample, i.e., the individuals that are randomly selected from the full population. Table B3 shows the associations between the EA PGS and each of the test outcomes. By comparing these estimates to those from Table 3, we again see that the associations between the polygenic scores and the test scores are similar in magnitude, and in both cases the associations increase over time. We also see that, again, between one fourth and one third of the association disappears once we control for the family background characteristics.

Finally, we focus on achievement tests in reading because they are the most frequent. But if our findings stem from the development of some more foundational skills, they should extend to other skill measures. Indeed, we observe the same pattern of increasing gaps with respect to achievement in mathematics, as shown in Table B4. This is again true also when controlling for family characteristics and when including family fixed effects[7]. Interestingly, these results also suggest that the genetic effects are in general even larger for math than for reading.

[7]The sample size is much smaller in grade 8 because this test was only introduced in 2018. We only include it in Panel A and B as the sample is too small to give meaningful results in the fixed effects model.

# 6 Discussion

In this paper, we show that although genetic and socioeconomic differences can be used to predict very similar gaps in academic achievement during elementary school, genetic and socioeconomic advantage make up two distinct but related causes of inequality. Being born into a favorable childhood environment is associated with more high-quality parental investments, and the environment experienced outside the childhood home only tends to reinforce such initial differences through different neighborhood quality, school quality, etc. Variation in genetic endowments has also been shown to be reinforced by parental investments (Houmark, Ronda, and Rosholm, 2020), but as we show in this paper, the link between genes and achievement is unrelated to selection into more favorable environments outside of the home. The reason that we nevertheless observe that genes only become more important throughout the school years is that they make individuals able to extract more resources from the environments that they find themselves in. This is true regardless of the quality of this environment, though children growing up in low SES families are not able to fully exploit their potential. When exposed to the high-quality Danish school system, however, their genetic potential is realized in spite of their earlier disadvantage because of the resources they reap through their psychological predispositions for learning, for example, higher conscientiousness. But childhood SES still exerts a lasting influence on achievement, and the children who are disadvantaged on both accounts face difficult odds in school.

Our results relate to several literatures. For one thing, they reveal that the apparent stability of the achievement gap throughout childhood is at least partly illusional. In particular, the skill formation literature emphasises the importance of the early childhood (Cunha and Heckman, 2007), partly because skill gaps between low- and high-SES children tend to be present both at the beginning and the end of the school years. This may suggest that schooling has little impact on inequality in ability (Heckman, 2008). Our results show that such conclusions may not be warranted, as there is significant reshuffling going on within the skill distribution. However, these gradual changes appear random until the children's genetic endowments are taken into account. Thus, neglect of genetic effects may actually cause one to underestimate the effects of schooling. This highlights how genes and the environment are closely interrelated, and how the existence of genetic effects do not imply a lesser role for investments in skill formation.

Because of this interrelation, genetic effects can manifest somewhat counter-intuitively. Since genetic endowments are realised before the start of life, it is tempting to think that genetic differences should be prominent early in life, while they should wane over time as environmental influences accumulate. In fact, our results suggest that, if anything, the

opposite is the case. Although genes are already predictive of achievement in 2nd grade, the gap in test scores that can be attributed to genetic differences is smaller than the SES gap. In 8th grade, on the other hand, we find that, while SES has neither become more nor less predictive of achievement (controlling for child genes), genes have become much more predictive. These results suggest that it is in fact genetic, rather than environmental, effects that accumulate over time[8]. The increasing importance of genes echoes early studies in behavioral genetics showing that adopted children tend to become more similar to their biological parents as they grow older (Plomin et al., 1997). It also extends recent findings that genetic effects increase gradually during early childhood (ages 0-7) (Houmark, Ronda, and Rosholm, 2020).

Our results also relate directly to Ronda et al. (2021), who also use the iPSYCH sample to study gene-environment interactions. Our finding that the PGS becomes more important among the low SES children is seemingly at odds with their finding that the relation between educational attainment and the EA PGS is *attenuated* by family SES, such that the effect of the EA PGS on educational attainment is lower in low-SES families than in high-SES families, leading to a lost potential. This apparent contradiction may have to do with the fact, mentioned by Landersø and Heckman (2017), that the low level of income inequality in Denmark makes incentives to undertake education relatively weak. If the incentives are particularly weak for children from low-SES families, it could explain the difference. This could also be related to other factors becoming more important when transitioning into the next level of the educational system, e.g., norms and traditions as well as role models and knowledge in the family of the educational system. This is a very interesting avenue for future research, as it is crucial to sustain the gains achieved in elementary school into the subsequent levels of the educational system to fully unlock the potential of all children.

# References

Andersen, Simon Calmar, Miriam Gensowski, Steven G Ludeke, and Oliver P John. 2020. "A stable relationship between personality and academic performance from childhood through adolescence. An original study and replication in hundred-thousand-person samples." *Journal of personality* 88 (5):925–939.

Belsky, Daniel W, Terrie E Moffitt, David L Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S

[8]Which is not the same as saying that the environment becomes "less important". Potentially, other environmental effects, not captured by our SES measure, could also accumulate.

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.

Bornstein, Marc H, Chun-Shin Hahn, Joan TD Suwalsky, and O Maurice Haynes. 2003. *Socioeconomic status, parenting, and child development: The Hollingshead Four-Factor Index of Social Status and The Socioeconomic Index of Occupations.* Lawrence Erlbaum Associates Publishers.

Bouchard, Thomas J and John C Loehlin. 2001. "Genes, evolution, and personality." *Behavior genetics* 31 (3):243–273.

Bradley, Robert H and Robert F Corwyn. 2002. "Socioeconomic status and child development." *Annual review of psychology* 53 (1):371–399.

Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. "The impact of early cognitive and non-cognitive skills on later outcomes." .

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

Cunha, Flavio and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov. 2006. "Interpreting the evidence on life cycle skill formation." *Handbook of the Economics of Education* 1:697–812.

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

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socioeconomic status and inequalities in children's IQ and economic preferences." *Journal of Political Economy* 129 (9):2504–2545.

Gensowski, Miriam, Rasmus Landersø, Dorthe Bleses, Philip Dale, Anders Højen, and Laura Justice. 2020. "Public and Parental Investments and Children's Skill Formation." *Rockwoolfondens Study Papers* No. 155.

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

Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science* 312 (5782):1900–1902.

———. 2008. "Schools, skills, and synapses." *Economic inquiry* 46 (3):289–324.

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

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm. 2020. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." *IZA Discussion Paper 13780* .

Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success." Tech. rep., National Bureau of Economic Research.

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

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

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

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

Martin, Alicia R, Christopher R Gignoux, Raymond K Walters, Genevieve L Wojcik, Benjamin M Neale, Simon Gravel, Mark J Daly, Carlos D Bustamante, and Eimear E Kenny. 2017. "Human demographic history impacts genetic risk prediction across diverse populations." *The American Journal of Human Genetics* 100 (4):635–649.

Niclasen, Janni, Maria Keilow, and Carsten Obel. 2018. "Psychometric properties of the Danish student well-being questionnaire assessed in¿ 250,000 student responders." *Scandinavian journal of public health* 46 (8):877–885.

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.

Pedersen, C B, J Bybjerg-Grauholm, M G Pedersen, Jakob Grove, E Agerbo, M Bækvad-Hansen, J B Poulsen, C S Hansen, J J McGrath, T D Als, J I Goldstein, B M Neale, M J Daly, D M Hougaard, O Mors, M Nordentoft, A D Børglum, T Werge, and P B Mortensen. 2018. "The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders." *Molecular Psychiatry* 23 (1):6–14.

Plomin, Robert, John C DeFries, and John C Loehlin. 1977. "Genotype-environment interaction and correlation in the analysis of human behavior." *Psychological Bulletin* 84 (2):309.

Plomin, Robert, David W Fulker, Robin Corley, and John C DeFries. 1997. "Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study." *Psychological Science* 8 (6):442–447.

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.

Privé, Florian, Julyan Arbel, and Bjarni J Vilhjálmsson. 2020. "LDpred2: better, faster, stronger." *BioRxiv, Working Paper* .

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm. 2021. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." *Scandinavian Journal of Economics* .

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.

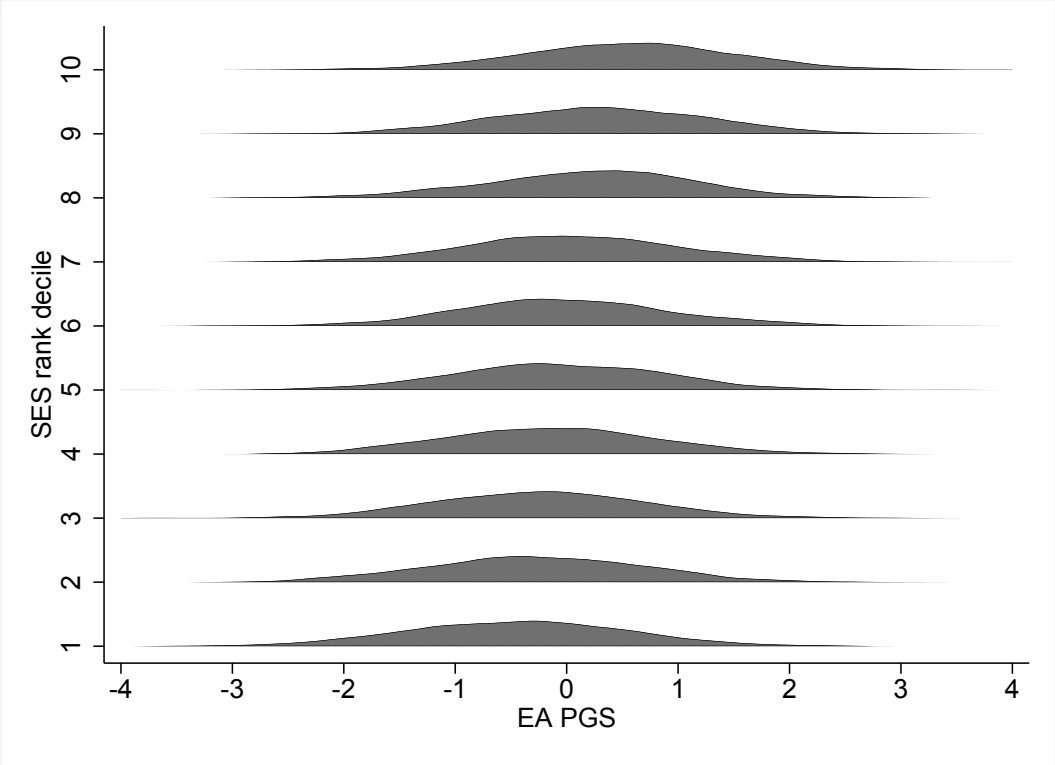
# 7 Tables and Figures

**Table 1:** Summary statistics

| | Main sample | Control sample | Sibling sample |
|---|---|---|---|
| *Birth characteristics* | | | |
| EA PGS | -0.015 | 0.054 | -0.017 |
| | (1.003) | (0.992) | (1.004) |
| Female | 0.419 | 0.487 | 0.384 |
| | (0.493) | (0.500) | (0.487) |
| Birth order | 1.712 | 1.736 | 1.813 |
| | (0.737) | (0.740) | (0.728) |
| *Family background* | | | |
| Family income | 12.946 | 12.978 | 12.963 |
| | (0.637) | (0.802) | (0.469) |
| Years of education (mother) | 13.120 | 13.503 | 13.091 |
| | (2.366) | (2.359) | (2.360) |
| Years of education (father) | 12.986 | 13.320 | 12.948 |
| | (2.358) | (2.371) | (2.346) |
| Any diagnosis (mother) | 0.089 | 0.059 | 0.147 |
| | (0.285) | (0.235) | (0.354) |
| Any diagnosis (father) | 0.056 | 0.039 | 0.091 |
| | (0.229) | (0.193) | (0.288) |
| Family intact | 0.847 | 0.895 | 0.890 |
| | (0.269) | (0.221) | (0.211) |
| SES rank | 0.492 | 0.551 | 0.489 |
| | (0.289) | (0.282) | (0.286) |
| $N$ | 40,136 | 16,074 | 2,705 |

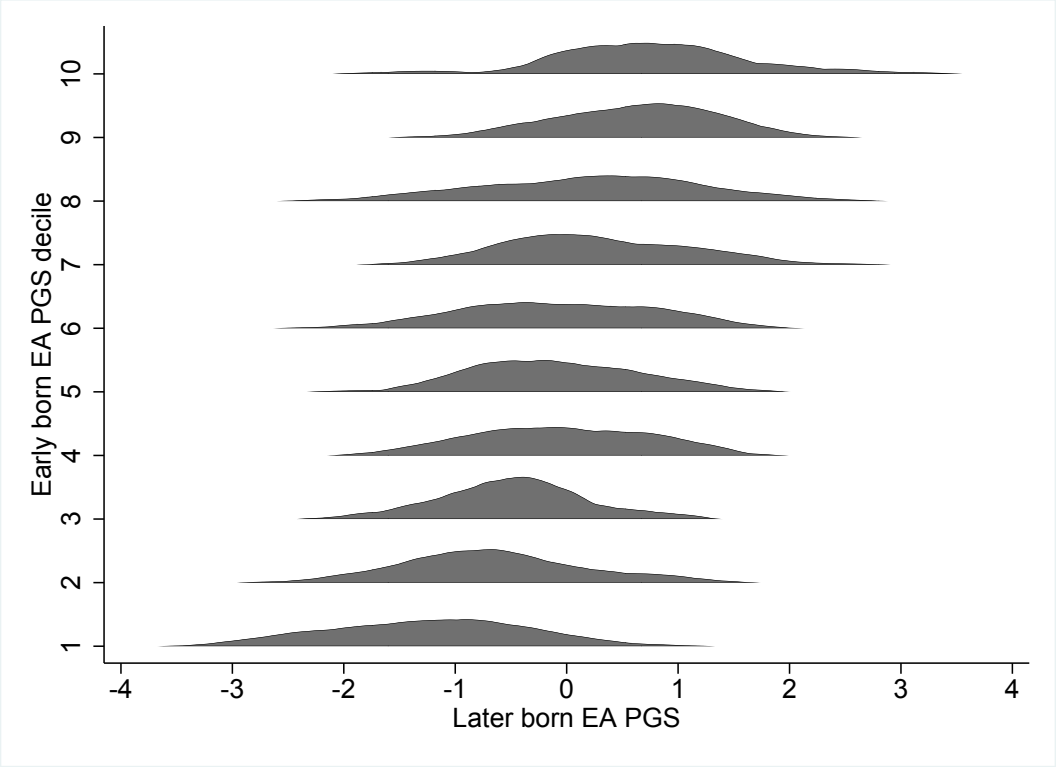
*Notes*: This table reports means and standard deviations of child birth characteristics and family background characteristics for the different estimation samples.

**Figure 1:** Distribution of the child's EA PGS by the parents' SES rank



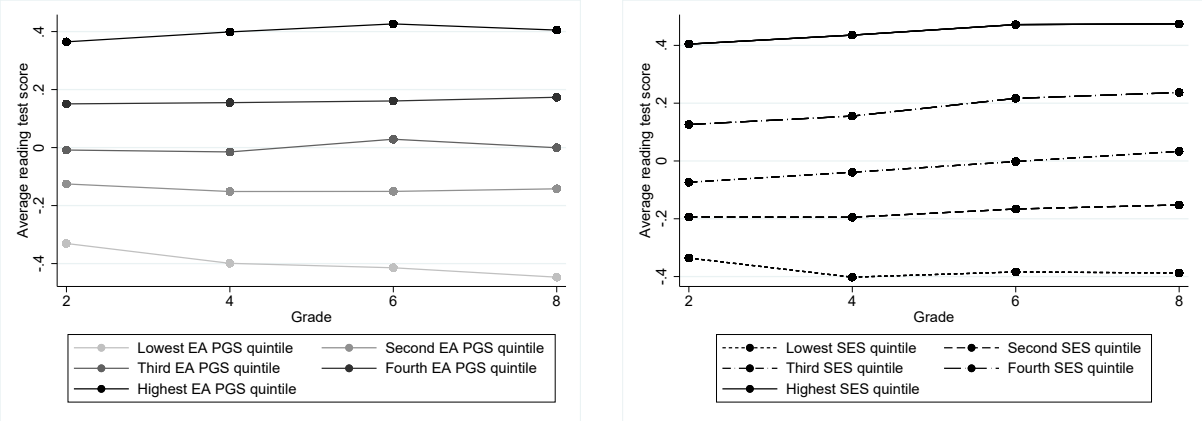

*Notes*: This figure plots the density of the child's standardized EA PGS, separately for each decile of the parental SES rank.

**Figure 2:** Siblings: Distribution of EA PGS




*Notes*: This figure plots the density of the standardized EA PGS of later born siblings, separately for each decile of the EA PGS of the earliest born sibling.

**Figure 3:** Separate achievement gaps




(a) PGS achievement gap

(b) SES achievement gap

*Notes*: These figures plot the average reading test scores from grade 2 to 8 for each quintile of the PGS and SES distributions, respectively.

**Table 2:** EA PGS, SES and School Achievement

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Test subject: | 8th grade reading | | | | | | |
| EA PGS | 0.301 | | 0.231 | 0.226 | 0.226 | 0.221 | 0.221 |
| | (0.007) | | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| SES | | 0.307 | 0.235 | 0.222 | 0.204 | 0.181 | 0.176 |
| | | (0.007) | (0.007) | (0.007) | (0.007) | (0.008) | (0.008) |
| Municipality F.E. | ( ) | ( ) | ( ) | (X) | (X) | (X) | (X) |
| School Quality | ( ) | ( ) | ( ) | ( ) | (X) | (X) | (X) |
| School F.E. | ( ) | ( ) | ( ) | ( ) | ( ) | (X) | (X) |
| Additional Controls | ( ) | ( ) | ( ) | ( ) | ( ) | ( ) | (X) |
| $R^2$ | 0.080 | 0.080 | 0.123 | 0.137 | 0.139 | 0.218 | 0.218 |
| N | | | | 25,076 | | | |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS and the SES index to test scores in 8th grade reading. In columns 1-3, we add the PGS and SES separately in combination. In columns 4-7, we include additional controls and fixed effects (F.E.). Standard errors are reported in parenthesis.

**Table 3:** EA PGS and School Achievement over time

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Test subject: | | Reading | | |
| Grade: | 2nd | 4th | 6th | 8th |
| Panel A: | | | | |
| EA PGS | 0.245 | 0.282 | 0.294 | 0.301 |
| | (0.008) | (0.007) | (0.006) | (0.006) |
| Family Controls | ( ) | ( ) | ( ) | ( ) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| N | 16,040 | 21,390 | 25,411 | 25,076 |
| Panel B: | | | | |
| EA PGS | 0.169 | 0.198 | 0.210 | 0.221 |
| | (0.008) | (0.007) | (0.007) | (0.007) |
| Family Controls | (X) | (X) | (X) | (X) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| N | 16,040 | 21,390 | 25,411 | 25,076 |
| Panel C: | | | | |
| EA PGS | 0.051 | 0.133 | 0.234 | 0.285 |
| | (0.080) | (0.057) | (0.048) | (0.056) |
| Family Controls | ( ) | ( ) | ( ) | ( ) |
| Family F.E. | (X) | (X) | (X) | (X) |
| N | 743 | 1,058 | 1,345 | 1,290 |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores in reading at different grade levels. For each outcome, we regress the test score on the EA PGS, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. In Panel B, we add controls for family income, maternal and paternal education, indicators for maternal and paternal psychiatric diagnoses and the fraction of time before 2nd grade that the child's parents have been cohabiting. In Panel C, we instead add fixed effects (F.E.) at the family level, estimating the model using only sibling pairs. Standard errors are reported in parenthesis.

**Table 4:** EA PGS and School Achievement by SES

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Family background: | | Low SES | | | High SES | |
| Grade | -0.008 | -0.009 | 0.003 | 0.003 | 0.003 | 0.011 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| EA PGS | 0.166 | 0.135 | 0.132 | 0.226 | 0.175 | 0.173 |
| | (0.014) | (0.014) | (0.015) | (0.012) | (0.012) | (0.013) |
| EA PGS x Grade | 0.014 | 0.014 | 0.013 | 0.001 | 0.003 | 0.002 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Family Controls | ( ) | (X) | (X) | ( ) | (X) | (X) |
| School F.E. | ( ) | ( ) | (X) | ( ) | ( ) | (X) |
| N | | 42,644 | | | 45,273 | |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores in reading over time using the between-families model. We regress the test outcome on the EA PGS and EA PGS interacted with grade level, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. We do this separately for low SES (columns 1-3) and high SES (columns 4-6), defined by being below/above the median on the SES index. In columns 2 and 5, we add controls for family income, maternal and paternal education, indicators for maternal and paternal psychiatric diagnoses and the fraction of time before 2nd grade that the child's parents have been cohabiting. In columns 3 and 6, we further add fixed effects (F.E.) at the school level. Standard errors are reported in parenthesis.

**Table 5:** EA PGS and Other School Outcomes

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Outcome: | School absence | | School starting age | | School quality | |
| Panel A: | | | | | | |
| EA PGS | -0.007*** | -0.001 | -0.018*** | 0.001 | 0.016*** | 0.001 |
| | (0.000) | (0.002) | (0.004) | (0.042) | (0.001) | (0.005) |
| Family F.E. | ( ) | (X) | ( ) | (X) | ( ) | (X) |
| N | 206,427 | 20,383 | 12,189 | 1,096 | 169,366 | 16,236 |
| Outcome: | Academic thriving | | Achievement | | Conscientiousness | |
| Panel B: | | | | | | |
| EA PGS | 0.108*** | 0.127*** | 0.102*** | 0.118*** | 0.146*** | 0.163*** |
| | (0.004) | (0.036) | (0.005) | (0.038) | (0.006) | (0.054) |
| Family F.E. | ( ) | (X) | ( ) | (X) | ( ) | (X) |
| N | 50,065 | 4,930 | 49,994 | 4,923 | 49,964 | 4,918 |
| Outcome: | Connectedness | | Environment | | Management | |
| Panel C: | | | | | | |
| EA PGS | 0.037*** | 0.049 | 0.026*** | 0.066** | 0.045*** | 0.037 |
| | (0.005) | (0.040) | (0.004) | (0.030) | (0.004) | (0.037) |
| Family F.E. | ( ) | (X) | ( ) | (X) | ( ) | (X) |
| N | 50,139 | 4,938 | 50,046 | 4,935 | 49,946 | 4,924 |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to various other school outcomes. Each outcome is regressed on the EA PGS, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. In columns 2, 4 and 6, we run the regressions using only sibling pairs, and also include fixed effects (F.E.) at the family level. Standard errors are clustered at the individual level and reported in parenthesis.

**Table 6:** Well-being survey measures

**Factor 1: Academic thriving**

**Achievement**

What do your teachers think of your progress in school?
Do you succeed in learning what you want in school?
How often can you find a solution to problems, if you try hard enough?
I do well in school, academically
If something is difficult for me during class, I can do something about it myself to move on

**Conscientiousness**

How often can you manage the things you set your mind to?
Can you concentrate during lessons?
If I am interrupted during class, I can quickly concentrate again

**Factor 2: School connectedness**

Do you like your school?
Do you like your class?
Do you feel lonely?
I feel that I belong at this school
Most of the students in my class are kind and helpful
Other students accept me as I am
How often do you feel safe at school?

**Factor 3: Learning environment**

Do your teachers help you learn in ways that work?
Lessons make me want to learn more
The teachers are good at supporting and helping me at school when I need it
Do you and your classmates have a say in what the class works on?
The teachers ensure that the students' ideas are used in class
Are the lessons exciting?
I like the surroundings outside my school
I like the classrooms at my school

**Factor 4: Classroom management**

If there is noise in the classroom, teachers can quickly establish quietness
Do your teachers show up for classes on time?
Is it easy to hear what the teachers say during lessons?
Is it easy to hear what the other students say during lessons?

*Notes*: This table reports the individual survey items that are used to construct the outcomes reported in Table 5.

# Appendix A Genes and socioeconomic outcomes

In this section, we first explain how we think of genetic endowments and genetic inheritance. We then use a simple conceptual framework to highlight the different channels through which genes may affect skill formation.

## Appendix A.1 Genetic endowments

The human genome consists of approximately 3 billion DNA base pairs across the 23 chromosome pairs. The base pairs are the fundamental units encoding genetic information, and a particular sequence of bases at a particular location is called a gene. Genes program cells to synthesize proteins that serve an abundance of biological functions relevant for particular inherited traits or attributes. Most of the genome is identical for two random individuals, as a typical genome differs from the "reference" genome at only 4-5 million base pair sites. The smallest unit of genetic variation is a difference in a single DNA base; a single nucleotide polymorphism (SNP).

Only four different nucleobases make up a DNA sequence. The bases are adenine (A), cytosine (C), guanine (G) and thymine (T). A SNP is thus a particular site in the DNA where the commonly found base is replaced by one of the other three bases. In practice, the vast majority of SNPs are biallelic, meaning that only two different bases are observed. For example, at a specific base site, 95% of individuals have an A while the remaining 5% have a G instead. There is then a SNP at this specific position, and the observed variants, A and G, are known as the major (common) and the minor (uncommon) allele for this site.

Thus, for any specific (biallelic) SNP, $s$, we describe the genetic endowment of individual $i$ as:

$$g_{is} \in \{0, 1, 2\} \tag{4}$$

where the value is equal to the number of minor alleles at the specific SNP.

An individual's genome is a random combination of the genomes of the individual's parents. One chromosome of each chromosome pair is inherited from each parent. Which of each chromosome is inherited is truly random as it depends on the orientation of the chromosomes at a specific point in time; analogous to a coin flip. In addition to this random assortment of chromosomes, additional genetic variation comes about because of chromosomal crossover. This happens with a small probability at any specific site on the chromosome and causes the chromosomes inherited from the mother and father to be separated at that

site and recombined with the other chromosome instead. Thus, not only are alleles inherited independently from each other across chromosomes because of random assortment. As long as the SNPs are some distance from each other on the chromosome, random crossover makes the alleles at different SNPs uncorrelated[9].

Thus, the expected value of the genetic endowment at each SNP is given as:

$$E[g_{is}] = 0.5g_{is}^{f} + 0.5g_{is}^{m} \tag{5}$$

where $g_{is}^{f}$ is the minor allele frequency for the child's father at position $s$, and similarly, $g_{is}^{m}$ is for the child's mother. Note that, because one of the two alleles is inherited from each parent, the process is deterministic when $g_{is}^{f} = g_{is}^{m} = 0$ or when $g_{is}^{f} = g_{is}^{m} = 2$. In these cases, there is no variation in parental genetic endowments at the SNP, and the child will have the same number of minor alleles as both of the parents. Similarly, if e.g., $g_{is}^{f} = 0$ and $g_{is}^{m} = 2$, it must be that $g_{is} = 1$. On the other hand, when either $g_{is}^{f} = 1$ or $g_{is}^{m} = 1$, i.e. when one parent has both the major and the minor allele at the SNP, the inheritance process is such that whether the child gets the major or minor allele is random by nature. This is a truly natural experiment potentially allowing for the identification of the causal effect of genes on any outcome.

## Appendix A.2 Genes and skill formation

The realisation of individual economic outcomes such as education and earnings depend on various personal attributes. Traditionally, economists summarized such characteristics as human capital. Human capital consists of different kinds of skills or abilities (Cunha et al., 2006). Hence, skills in the broad sense are all the psychological characteristics that influence what individuals achieve or how well they perform in a given task, ranging from cognitive ability (such as IQ) to non-cognitive skills or personality. This means that performance on a particular task may be considered a measure of particular skills (Heckman and Mosso, 2014).

In the skill formation literature, skills are characterised both by a form of permanence and impermanence. Skills are impermanent in the sense of being malleable. They are sensitive to the environment, and can be fostered through human capital investments such as education (although this sensitivity is often strongly age-dependant). On the other hand, skills at one period of time exert a permanent influence in the sense that the future stock of skills is

[9]SNPs that are close to each other and therefore not independent are said to be in linkage disequilibrium (LD). When summarizing genetic endowments, one possibility is to disregard SNPs that are in LD with each other. A preferred method, known as LDpred, allows one to use the information from all SNPs while accounting for LD (Privé, Arbel, and Vilhjálmsson, 2020)

partly determined by previous skills. Hence, skills are the product of previous skills and investments (Cunha and Heckman, 2007).

But while all skills are to some extent malleable, it is becoming evident that all skills or traits are also to some extent heritable (Polderman et al., 2015). The skill formation literature also emphasizes that skills are formed by an interaction between genetics and the environment (Heckman, 2006), but genes have generally only featured in these models indirectly through an initial stock of skills that is said to depend on unobserved genetic influences (Cunha, Heckman, and Schennach, 2010). However, the possibility to measure individuals' genetic endowments made feasible by recent scientific advances may inform the skill formation process further, as shown by Houmark, Ronda, and Rosholm (2020). Extending the traditional skill formation model to allow for genetic influences thus reveals that, just like investments, genes exert a continuing influence on skill development through early childhood (age 0-7 years).

In this paper, we likewise conceptualise skills as being determined by both previous skills, investments and genes, but we now consider the next stage of childhood, namely when children enter elementary school. Though skills are multidimensional in nature, for simplicity, we assume that skills may be described by a uni-dimensional measure. This may reflect one particular dimension, e.g., cognitive ability or reading skills, or it may reflect a summary of the various skills that are relevant for educational achievement. Hence, individual $i$'s stock of skills develop according to:

$$\theta_{it+1} = f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i}) \tag{6}$$

where $f_t^{\theta}$ is a function describing how skills are formed by an interaction between previous skills, home investments during early childhood, later investments of various types, and the child's genetic endowments. $\mathbf{g_i}$ is the vector of all genetic endowments of individual $i$, $\mathbf{g_i} = \{g_{i1}, ..., g_{iS}\}$. $I_{i0}^h$ indicates all the characteristics of the early childhood environment that affects skill development, i.e., socio-economic status (SES) in a broad sense. Thus, to the extent that early home investments affect skill formation, SES achievement gaps are captured by:

$$\frac{\partial f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial I_{i0}^h} \tag{7}$$

And the SES achievement gaps are increasing if:

$$\frac{\partial^2 f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial I_{i0}^h \partial t} > 0 \tag{8}$$

Note that this could happen for various reasons. Early investments may evoke subsequent home investments or other types of investments at a later time ($I_{it} = \{I_{it}^h, ..., I_{it}^p\}$ ), e.g. because high-SES children are sent to schools of higher quality. It may also be that early skills fostered by home investments increase the returns to subsequent investments.

However, all of these effects may also stem from differences in the child's genetic endowments. That is, the extent to which there are genetic achievement gaps is determined by:

$$\frac{\partial f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial \mathbf{g_i}} \quad \text{and} \quad \frac{\partial^2 f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial \mathbf{g_i} \partial t} \tag{9}$$

where genes may have an effect through each of the inputs in the skill production function. It may be considered a direct effect if it works either through the current stock of skills, or conditional on the current stock of skills and investments. That is, because of their genetic potential, some children will have accumulated a higher stock of skills at a certain point in time, and that may affect their ability to learn new skills. But genes may also make some children better at retaining new concepts and learn from their environment regardless of their current skills and the particular environment they are in.

In addition, genes may have an effect through investments. This is a form of gene-environment correlation, also known as genetic nurture (Kong et al., 2018). Because of their genetic endowments, some children may actively seek out environments that are more conducive to learning (active gene-environment correlation). But the environment may also change as a response to the child's genetic endowments, e.g., parents may invest differently based on the perceived potential of the child (reactive gene-environment correlation) (Plomin, DeFries, and Loehlin, 1977).

While the difference is theoretically clear, distinguishing between SES gaps and genetic gaps in achievement may be difficult in practice. More so because they share an important determinant: parental genes. As explained, a child's genetic endowments are determined by parental genes according to 5. But parental genes may also affect the childhood environment. Hence, SES achievement gaps may partly be due to genetic differences and vice-versa. Of course, differences in SES are in general predictive, or at best an imperfect measure of a causal relationship. If the same is true for genetic gaps, it may not be clear whether the two

types of inequality are qualitatively different. However, one may, for example, control for various aspects of the family background and see if the genetic gaps remain. Furthermore, if one can control for parental genes, it will be possible to exploit the natural experiment of genetic inheritance and identify the independent effect of the child's genes. If the SES and genetic gaps are not simply explained by the other, this begs the question whether there is an interaction between the two, i.e.:

$$\frac{\partial^2 f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial \mathbf{g_i} \partial I_{i0}^h} \tag{10}$$

and whether this changes over time:

$$\frac{\partial^3 f_t^{\theta}(\theta_{it}, I_{i0}^h, I_{it}, \mathbf{g_i})}{\partial \mathbf{g_i} \partial I_{i0}^h \partial t} \tag{11}$$

# Appendix B Additional results

**Table B1:** EA PGS, SES and School Achievement

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Test subject: | | | Reading | | |
| Grade | -0.002 | -0.002 | 0.008 | -0.017 | -0.017 |
| | (0.002) | (0.002) | (0.002) | (0.010) | (0.010) |
| EA PGS | 0.111 | 0.114 | 0.105 | 0.115 | 0.111 |
| | (0.020) | (0.020) | (0.022) | (0.143) | (0.143) |
| EA PGS x Grade | 0.020 | 0.020 | 0.019 | 0.025 | 0.025 |
| | (0.003) | (0.003) | (0.004) | (0.021) | (0.021) |
| EA PGS x SES | 0.099 | 0.068 | 0.079 | 0.034 | 0.040 |
| | (0.032) | (0.032) | (0.034) | (0.220) | (0.220) |
| EA PGS x Grade x SES | -0.022 | -0.021 | -0.019 | -0.025 | -0.027 |
| | (0.006) | (0.005) | (0.006) | (0.033) | (0.033) |
| SES | 0.802 | 0.341 | 0.261 | | |
| | (0.013) | (0.026) | (0.033) | | |
| Family Controls | ( ) | (X) | (X) | ( ) | (X) |
| School F.E. | ( ) | ( ) | (X) | ( ) | ( ) |
| Family F.E. | ( ) | ( ) | ( ) | (X) | (X) |
| $R^2$ | 0.138 | 0.144 | 0.110 | 0.044 | 0.046 |
| N | | 87,917 | | 4,436 | |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores in reading over time across family SES. We regress the test score on the EA PGS and EA PGS interacted with grade level as well as SES and SES interacted with each of the EA PGS terms. We further control for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. Columns 1-3 uses the between-families model and columns 4-5 uses the within-families model. In columns 2, 3 and 5, we add family controls. Further, we add fixed effects (F.E.) at the school level in column 3 and at the family level in columns 4 and 5.

**Table B2:** Balanced Sample: EA PGS and Human Capital Formation in elementary School

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Test subject: | | Reading | | |
| Grade: | 2nd | 4th | 6th | 8th |
| Panel A: | | | | |
| EA PGS | 0.222 | 0.250 | 0.276 | 0.27 |
| | (0.013) | (0.013) | (0.013) | (0.013) |
| Family Controls | ( ) | ( ) | ( ) | ( ) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| Panel B: | | | | |
| EA PGS | 0.153 | 0.180 | 0.201 | 0.201 |
| | (0.014) | (0.014) | (0.013) | (0.013) |
| Family Controls | (X) | (X) | (X) | (X) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| N | 5,171 | 5,171 | 5,171 | 5,171 |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores in reading and mathematics at different grade levels for the balanced sample. This is equivalent to Table 3, except that we only include individuals if we observe them taking all of the tests from grade 2 through 8. For each outcome, we regress the test score on the EA PGS, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. In Panel B, we add controls for family income, maternal and paternal education, indicators for maternal and paternal psychiatric diagnoses and the fraction of time before 2nd grade that the child's parents have been cohabiting. Standard errors are reported in parenthesis.

**Table B3:** Control Sample: EA PGS and Human Capital Formation in elementary School

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Test subject: | | Reading | | |
| Grade: | 2nd | 4th | 6th | 8th |
| Panel A: | | | | |
| EA PGS | 0.241 | 0.261 | 0.280 | 0.287 |
| | (0.010) | (0.009) | (0.008) | (0.009) |
| Family Controls | ( ) | ( ) | ( ) | ( ) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| Panel B: | | | | |
| EA PGS | 0.172 | 0.185 | 0.204 | 0.216 |
| | (0.011) | (0.009) | (0.008) | (0.009) |
| Family Controls | (X) | (X) | (X) | (X) |
| Family F.E. | ( ) | ( ) | ( ) | ( ) |
| N | 7,682 | 9,674 | 11,027 | 10,373 |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores in reading and mathematics at different grade levels for the control sample. This is equivalent to Table 3, except that we only include individuals if they are part of the representative iPSYCH subsample. For each outcome, we regress the test score on the EA PGS, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. In Panel B, we add controls for family income, maternal and paternal education, indicators for maternal and paternal psychiatric diagnoses and the fraction of time before 2nd grade that the child's parents have been cohabiting. Standard errors are reported in parenthesis.

**Table B4:** Mathematics: EA PGS and School Achievement

| | (1) | (2) | (3) |
|---|---|---|---|
| | Mathematics | | |
| Grade: | 3rd | 6th | 8th |
| Panel A: | | | |
| EA PGS | 0.258 | 0.298 | 0.356 |
| | (0.008) | (0.007) | (0.023) |
| Family Controls | ( ) | ( ) | ( ) |
| Family F.E. | ( ) | ( ) | ( ) |
| N | 19,188 | 25,157 | 2,155 |
| Panel B: | | | |
| EA PGS | 0.180 | 0.206 | 0.249 |
| | (0.008) | (0.007) | (0.023) |
| Family Controls | (X) | (X) | (X) |
| Family F.E. | ( ) | ( ) | ( ) |
| N | 19,188 | 25,157 | 2,155 |
| Panel C: | | | |
| EA PGS | 0.140 | 0.301 | |
| | (0.070) | (0.058) | |
| Family Controls | ( ) | ( ) | |
| Family F.E. | (X) | (X) | |
| N | 964 | 1,318 | |

*Notes*: This table reports parameter estimates from regressions used to link the EA PGS to test scores at different grade levels, equivalent to Table 3 except for mathematics instead of reading. For each outcome, we regress the test score on the EA PGS, controlling for indicators for birth year, gender, birth order and the first ten principal components of the genetics matrix. In Panel B, we add controls for family income, maternal and paternal education, indicators for maternal and paternal psychiatric diagnoses and the fraction of time before 2nd grade that the child's parents have been cohabiting. In Panel C, we instead add fixed effects at the family level, estimating the model using only sibling pairs. Standard errors are reported in parenthesis.