CESifo
# AREA CONFERENCES 2022

## Economics of Education

**Munich, 2 – 3 September 2022**

### Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments

*Benjamin W. Arold, Paul Hufe, and Marc Stoeckli*

CESifo

# Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments*

Benjamin W. Arold, Paul Hufe & Marc Stoeckli

August 1, 2022

**Abstract**

Genetic endowments are fixed at conception and matter for the educational attainment of individuals. Do investments in schooling environments mitigate or magnify the outcomes of this genetic lottery? We analyze the interdependent associations of genetic endowments, teacher quality and teacher quantity with educational attainment in the United States. Our results suggest that higher-quality teachers are substitutes for genetic endowments: a 1 SD increase in teacher quality reduces the positive association between educational attainment and a 1 SD increase in the relevant polygenic score by 20%. This increase is underpinned by relative gains in health, language ability, patience, and risk aversion.

**JEL-Codes:** I29; I21; J24
**Keywords:** Polygenic scores, School resources, Skill formation

*__Arold__: LMU Munich and ifo Institute (arold@ifo.de); **Hufe** (corresponding author): University of Bristol, CESifo and IZA, Priory Road Complex, Bristol, BS81TU, UK (paul.hufe@bristol.ac.uk); **Stoeckli**: LMU Munich and ifo Institute (marc.stoeckli@econ.lmu.de). We gratefully acknowledge funding from Deutsche Forschungsgemeinschaft (DFG) through NORFACE project "IMCHILD: The impact of childhood circumstances on individual outcomes over the life-course" (PE 1675/5-1). This paper has benefited from discussions with Pietro Biroli, Dalton Conley, Benjamin Domingue, Jeremy Freese, Stephanie von Hinke, Ingo Isphording, Hans van Kipperluis, Nicolas Papageorge, Hans Sievertsen, Ludger Woessmann, and Weilong Zhang. We are also grateful to conference participants at IIPF 2019, IWAEE 2019, SMYE 2021, the ifo Conference on Genes, Social Mobility, and Inequalities across the Life-Course, the CESifo Venice Summer Institute on Economics of Education and Equality of Opportunity, and the European Social Science Genomics Network (ESSGN) Conference. Furthermore, we have received valuable feedback from seminar audiences at Cologne, CEBI Copenhagen, Harvard, Munich, Princeton, and the University of Southern California. All remaining errors are our own.

# 1 INTRODUCTION

Education is a key determinant of life outcomes, both for individuals and for societies as a whole (Acemoglu and Autor, 2011; Hanushek and Woessmann, 2008; Krueger and Lindahl, 2001). Hence, improving equity and efficiency in education systems is a central policy goal in modern societies. To achieve such improvements, it is important to understand the role of genetic endowments in educational attainment: on the one hand, genetic endowments are strong predictors of education; in heritability studies they account for 40% of the variation in years of education (Branigan et al., 2013; Lee et al., 2018). On the other hand, the importance of genetic endowments varies with social environments like families, neighborhoods, and schools (Cesarini and Visscher, 2017; Koellinger and Harden, 2018). Therefore, the link between genetic endowments and life outcomes may be modified by policy interventions. This observation raises important questions: can school reforms moderate the link between genetic endowments and educational outcomes? If so, which domains of the school environment are particularly effective in doing so? Answers to these questions are of utmost importance to address equity and efficiency concerns related to current education systems. Despite this importance, current evidence is scant.

In this paper, we study the interaction of genetic endowments and school environments in the production of educational attainment. We focus on two dimensions of school environments that have been studied extensively in the literature on education economics: teacher quality and class size (Angrist and Lavy, 1999; Angrist et al., 2019; Chetty et al., 2014a,b; Fredriksson et al., 2013; Jackson, 2019; Leuven and Løkken, 2020; Rivkin et al., 2005; Rockoff, 2004). Importantly, these dimensions can be directly influenced by policy interventions that apply to all children and do not presuppose any form of genetic screening (Martschenko et al., 2019).

We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to study the interaction of genetic endowments and school environments in a between-family design. Add Health is a 5-wave panel study that follows a representative sample of US high school students from 1994/95 to the present. To the best of our knowledge, Add Health is the only (publicly available) data set that offers detailed information on schooling environments from both survey and administrative sources for a genotyped sample of reasonable size.

To measure genetic endowments, we leverage recent advances in molecular biology and use a polygenic score for educational attainment ($PGS^{EA}$, Dudbridge, 2013; Lee et al., 2018). $PGS^{EA}$ is an individual measure of the genetic propensity to attain education.[1] It offers important advantages over traditional proxies for "innate ability", such as student test scores and IQ tests (Brinch and Galloway, 2012; Hanushek and Woessmann, 2008, 2012; Heckman et al., 2010).

[1] In addition, $PGS^{EA}$ has been shown to be highly predictive of a number of life outcomes that are closely related to educational attainment. These outcomes include earnings, wealth and (non-)cognitive skills (Barth et al., 2020; Buser et al., 2021a; Demange et al., 2021; Houmark et al., 2020; Lee et al., 2018; Muslimova et al., 2020; Papageorge and Thom, 2020).

More specifically, it is a fine-grained DNA-based measure of innate ability that is fixed at conception and cannot be modified by environmental interventions thereafter. To measure the quality of school environments, we use information from headmaster surveys and administrative data sources such as the Common Core of Data and conduct a principal component analysis of the following school-level characteristics: teacher experience, teacher turnover, teacher education, teacher diversity, as well as class sizes and student-teacher ratios. From this analysis, we extract two factors that are indicative of the quality of teachers ($I_{Quality}$) and the quantity of teachers relative to the number of students ($I_{Quantity}$), respectively.

Causal identification of gene-environment interactions is challenging. In this study, we rely on a between-family comparison in which we control for an extensive set of predetermined family background characteristics. Between-family designs identify causal effects under strong identifying assumptions. We discuss these identification assumptions in detail and provide tests for their satisfaction. First, genetic endowments are correlated with other family characteristics that co-determine educational attainment. Therefore, even conditional on observables our parameters of interest may be confounded by *gene-environment correlation* or *genetic nurture effects*. In response, we show that the relevant point estimates from the between-family design can be replicated in a smaller sibling sample that allows us to control for genetic nurture by including family fixed effects. Second, school characteristics may be correlated with other family characteristics that co-determine educational attainment. Therefore, even conditional on observables our parameters of interest may be confounded by *selection effects*. Thus, we calculate different summary statistics to quantify the potential magnitude of confounding by unobservables (Cinelli and Hazlett, 2020; Oster, 2019). The results suggest that any residual confounding influence is small and unlikely to overturn our main results. Lastly, gene-environment interactions can only be identified if genetic endowments and the environmental variable of interest are distributed independently of each other. In response, we show that we cannot reject the equality of $PGS^{EA}$ distributions in various school environments. In addition to extensive tests for the relevant identifying assumptions, we note that our results withstand a series of empirical tests for competing mechanisms, that we discuss in detail below.

Our results can be summarized as follows. First, genetic endowments and teacher quality are highly predictive of years of education: a one-standard-deviation increase in $PGS^{EA}$ (teacher quality) increases educational attainment by $\approx$ 0.37 (0.22) years.[2] Second, genetic endowments and teacher quality act as substitutes in the production of educational attainment: a one-standard-deviation increase in teacher quality reduces the positive association between educational attainment and $PGS^{EA}$ by $\approx$ 20%. This result implies that improvements in teacher quality may reduce the genetic gradient in educational attainment. Furthermore, it suggests that teacher quality may offset the effects of family socio-economic status—an environmental characteristic that tends to magnify the genetic gradient in educational attainment (Papa-

[2]These increases correspond to 17% (10%) of a standard deviation.

george and Thom, 2020; Ronda et al., forthcoming).[3] Third, unlike teacher quality, teacher quantity is not associated with educational attainment—a null result that does not vary across the $PGS^{EA}$ distribution.

We perform a series of robustness checks to evaluate whether our results are conflated by competing mechanisms. We first show that our measures for teacher quality and teacher quantity do not pick up the effects of other school characteristics that may correlate with student outcomes. These characteristics comprise school peer characteristics, school-level policies such as sanctions for academic misconduct, and overall school value-added. Next, we demonstrate that our results are not driven by gene-environment interactions that reflect family instead of school environments. To that end, we run a models that controls for all possible interactions between $PGS^{EA}$, $I_{Quality}$, $I_{Quantity}$, and a broad set of parental background characteristics (Keller, 2014).

We also analyze the mechanisms underlying the substitutability of genetic endowments and teacher quality. Educational attainment summarizes information from various educational stages, with each stage requiring a different mix of skills (Cunha et al., 2006, 2010). Therefore, we repeat our analysis by replacing total educational attainment with binary variables indicating whether respondents have obtained a given educational degree. We find a substitutability of genetic endowments and teacher quality in the production of high school and college degrees, while there is no substitutability for post-graduate degrees. These results are in notable contrast to Papageorge and Thom (2020), who find a growing complementarity of parental background characteristics and genetic endowments as individuals progress through the educational system. To uncover which type of skills drives our results, we analyze the associations of $PGS^{EA}$ and teacher quality with a range of intermediate outcomes, including subjective and objective health, cognitive skills, economic preferences, and personality measures. We find that the substitutability of genetic endowments and teacher quality with respect to subjective health, verbal intelligence, risk-aversion, and patience underpins our main result.

Our study contributes to three strands of literature. First, we contribute to the literature on gene-environment interactions. Existing evidence shows that the association between socio-economic outcomes and genetic endowments varies with parental socio-economic status (Houmark et al., 2020; Papageorge and Thom, 2020; Ronda et al., forthcoming). Evidence on gene-environment interactions regarding school environments is more sparse. Barcellos et al. (2021) use a compulsory schooling reform to show that returns to schooling are lower for genetically advantaged students. However, they focus on the length of schooling and not the quality of school environments. Trejo et al. (2018) show a stronger genetic gradient in schools with better educated parents. However, with endogenous sorting, the composition of schools is a difficult target for policy intervention. Therefore, we focus on margins that can be directly influenced by policy-makers: the quality and quantity of teachers.

[3] See also our replication of their findings in section 5.

Second, we contribute to the literature on teacher quality. The positive effects of teacher quality on short- and long-term outcomes of students are well-documented (Chetty et al., 2014a,b; Jackson, 2019; Rivkin et al., 2005; Rockoff, 2004). However, the literature is far less conclusive regarding the equalizing effects of teacher quality on different student subgroups. For example, Aaronson et al. (2007) find that low-achieving students benefit more from high-quality teachers. In contrast, Chetty et al. (2014b) show that students from minority and low-income backgrounds benefit less. While previous studies have evaluated heterogeneities along dimensions that conflate genetic and social factors, we are able to measure the genetic predisposition for educational success as fixed at conception. We show that the positive effects of higher-quality teachers on educational attainment are concentrated among students with lower genetic endowments.

Third, we contribute to the literature on class size. Here, the average effects on students' outcomes are subject to academic debate. On the one hand, experimental studies on class size reductions tend to show positive effects on student achievement (Chetty et al., 2011; Krueger, 1999). On the other hand, quasi-experimental analyses exploiting maximum class-size rules tend to find mixed results, even when analyzing similar settings (Angrist and Lavy, 1999; Angrist et al., 2019; Fredriksson et al., 2013; Leuven and Løkken, 2020). The equalizing effects of class-size reductions are also controversial. For example, Krueger (1999) shows that class size reductions are more beneficial to students from a minority and low-income background. In contrast, Fredriksson et al. (2013) document that wage increases following a reduction in class size are more pronounced for students from high income backgrounds. Our study is the first to evaluate heterogeneities along the genetic dimension. We show that teacher quantity is not associated with gains in educational attainment irrespective of genetic endowment.

Our results are policy relevant. First, we show that higher teacher quality promotes educational attainment in the lower tail of the $PGS^{EA}$ distribution, but does not compromise achievement in the upper tail. This finding suggests that policymakers do not face an equity-efficiency trade-off when investing into teacher quality. Second, in contrast to teacher quality, we find no effect of teacher quantity on the educational outcomes of students, regardless of genetic endowments. This finding suggests that policymakers who are willing to address the equity and efficiency concerns related to genetic endowments do not face a trade-off between investments into teacher quality and teacher quantity. The latter finding is economically relevant as teacher salaries and employee benefits are by far the largest cost factor in the US school system, accounting for about half of US public primary and secondary schools expenditures (Figure B.1).

The remainder of this paper is structured as follows. In section 2, we provide an introduction to the measurement of genetic endowments. In section 3, we detail our empirical strategy. After introducing our data sources in section 4, we present results in section 5. Section 6 concludes the paper.

## 2 MEASURING GENETIC ENDOWMENT

The "First Law of Behavior Genetics" states that all human behavioral traits are heritable (Turkheimer, 2000). That is, genetic endowments explain the expression of each trait, at least to some extent. The empirical challenge is to identify the specific sequences in the genome that are related to the traits of interest.[4] Recent advances in molecular genetics have enabled a novel method of genetic discovery: genome-wide association studies (GWAS). GWAS exploit the most common type of genetic variation between humans, so-called single-nucleotide polymorphisms (SNP). SNPs occur when a single nucleotide—the basic building block of DNA molecules—differs at a specific position in the genome. Humans have around ten million SNPs. GWAS estimate separate linear regressions that relate a SNP of individual $i$ at genome location $j$ to an outcome of interest $y$:

$$y_i = \psi_j^y SNP_{ij} + \delta C_i + \varepsilon_i. \tag{1}$$

$SNP_{ij} \in \{0, 1, 2\}$ is a count variable and indicates the number of minor alleles that individual $i$ possesses at location $j$. Minor alleles are the less frequent genetic variation within a population. As humans inherit one of each chromosome from each parent, they possess either zero, one, or two minor alleles at each location $j$. $C_i$ is a vector of control variables to filter out spurious correlations due to non-biological differences across population groups. A particular SNP coefficient $\psi_j^y$ is considered genome-wide significant if the null hypothesis of non-association is rejected at a level of $p < 5 \times 10^{-8}$ (Chanock et al., 2007). The $p$-value is deliberately set low to account for multiple hypothesis testing.

The association of a single SNP with $y$ is minuscule, but jointly they can explain a substantial share of the observed outcome differences between individuals (Lee et al., 2018). In particular, the estimated SNP coefficients can be used to construct polygenic scores (PGS). A PGS is a scalar measure of an individual's genetic pre-disposition to an outcome of interest relative to the population. Formally, individual $i$'s PGS for outcome $y$, $PGS_i^y$, is constructed by linear aggregation of all $SNP_{ij}$ using $\psi_j^y$ as weighting factors:

$$PGS_i^y = \sum_j \widehat{\psi}_j^y SNP_{ij}, \tag{2}$$

where $\widehat{\psi}_j^y$ is the estimated SNP coefficient from equation (1). To avoid overfitting, equation (1) is estimated in a discovery sample, whereas the PGS is constructed in a hold-out sample (Wray et al., 2014).

The predictive power of a PGS is largely determined by two factors: the heritability of the outcome, which serves as an upper bound of the variance the PGS can explain, and the size

[4]Human genetic information is stored in 23 chromosome pairs that consist of deoxyribonucleic acid (DNA) molecules. These chromosomes, in turn, contain 20,000 to 25,000 genes—specific DNA sequences that provide instructions for building proteins. More than 99% of the sequences are identical in all humans.

of the discovery sample (Dudbridge, 2013). All else equal, the greater the heritability of the outcome, or the larger the discovery sample used to estimate the aggregation weights $\widehat{\psi}_j^y$, the higher the predictive accuracy of the PGS. For example, the heritability of educational attainment is around 40% (Branigan et al., 2013). The PGS for educational attainment constructed by Lee et al. (2018) is based on information from 1.1 million individuals and explains 12.7% of the variance in educational attainment.

The interpretation of PGS is not trivial. First, PGS are not purely measures of biological influence. In particular, GWAS coefficients may capture environmental factors such as population stratification across geographic regions (Abdellaoui et al., 2019). To this address this concern, we follow standard practice and always control for the first 20 principal components of the genetic data in our empirical analysis.[5] Second, the explanatory power of PGS depends on the context of its application. If a PGS is applied in one context, while the underlying GWAS was estimated in a completely different context, the predictive power of the PGS will be attenuated. In our context, this concern is limited: we apply PGS to a sample from the United States, while the underlying GWAS predominantly draws on samples from other industrialized countries with comparable education systems. Third, PGS are noisy measures of genetic endowments. Due to current GWAS sample sizes, they do not capture all of the genetic variation relevant to the outcome of interest. As a direct consequence, alternative PGS are still predictive of educational attainment over and above $\text{PGS}^{\text{EA}}$. However, in Appendix Table A.4, we show that $\text{PGS}^{\text{EA}}$ has significantly better predictive power than any plausible alternative PGS. Therefore, it is the best among other noisy measures for genetic endowments.

PGS are now available for a variety of outcomes. These include, for example, the body mass index and height (Yengo et al., 2018), attention deficit hyperactivity disorder (Demontis et al., 2019), major depressive disorder (Howard et al., 2019), intelligence (Savage et al., 2018), smoking (Liu et al., 2019), and sleep duration (Jansen et al., 2019). For our analysis, we rely on the PGS for educational attainment from Lee et al. (2018).

## 3 EMPIRICAL STRATEGY

### *3.1 Empirical Model*

Consider a model in which the skills $\theta$ of child $i$ at age $a$ are determined by prior skill levels $\theta_{ia-1}$, parental investments $I_{ia}^P$, school investments $I_{ia}^S$, and genetic endowments $G_i$.[6] There are

[5] The first principal components of the full matrix of genetic data capture most of the geographic variation in allele frequencies (see Mills et al., 2020, chapter 9.4, for a discussion). Therefore, they control for the geographic correlation between allele frequencies and socio-economic status.

[6] For the sake of parsimony, we abstract from other actors in the child development process, such as grandparents or providers of early childhood education.

three phases of skill accumulation:

$$\theta_{ia} = \begin{cases} f_a(G_i) & \text{, for child age } a = -1, \\ f_a(I_{ia}^P, \theta_{ia-1}, G_i) & \text{, for child age } a = 0, ..., 5, \\ f_a(I_{ia}^S, I_{ia}^P, \theta_{ia-1}, G_i) & \text{, for child age } a = 6, ..., A. \end{cases} \tag{3}$$

The skills determined at conception are defined by genetic endowments only. For children ages $a = 0, ..., 5$, i.e. in the period after conception and prior to attending school, parents are the only source of investments into skills. Parental investments include health behaviors during pregnancy, monetary investments such as buying toys or books, and time investments such as playing with or reading to the child. For $a = 6, ..., A$, schools are an additional source of investments into skills. School-based investments include instruction by teachers or interactions with peers.

Furthermore, assume the completed education $Y$ to be a function of individual skills accumulated by the end of childhood at age $a = A$:

$$Y_i = g(\theta_{iA}). \tag{4}$$

By recursively substituting equations (3) and (4) across child ages $a = 1, ..., A$, we obtain a model in which educational attainment is determined by initial genetic endowments, the history of family inputs, and the history of schooling inputs:

$$Y_i = h(I_{iA}^S, ..., I_{i6}^S, I_{iA}^P, ..., I_{i1}^P, G_i). \tag{5}$$

We are interested in the complementarity of schooling inputs and genetic endowments at a particular child age $a$:

$$\kappa = \frac{\partial^2 h(I_{ia}^S, \overline{I_{ia-1}^S}, ..., \overline{I_{i6}^S}, \overline{I_{ia}^P}, \overline{I_{ia-1}^P}, ..., \overline{I_{i1}^P}, G_i)}{\partial I_{ia}^S \partial G_i}. \tag{6}$$

If $\kappa < 0$, genetic endowments and school investments at age $a$ are *substitutes* in the production of educational attainment, i.e. school investments are more productive for individuals with comparatively disadvantageous genetic endowments. Reversely, if $\kappa > 0$, genetic endowments and school investments at age $a$ are *complements* in the production of educational attainment, i.e. school investments are more productive for individuals with comparatively advantageous genetic endowments.

In this study, we focus on school investments during high school ($14 \leq a \leq 18$). We estimate the complementarity parameter $\kappa$ using a linear regression model with an interaction term:

$$Y_i = \alpha G_i + \beta I_{ia}^S + \kappa (G_i \times I_{ia}^S) + \mathbf{X}_i(a)\gamma + \epsilon_i, \tag{7}$$

where $\mathbf{X}_i(a)$ denotes a vector of control variables to condition on the history of family and schooling inputs up to age $a = 14$.

### *3.2 Identifying Assumptions*

The parameter of interest $\kappa$ is identified if the following conditions are met: (i) exogenous variation in $G_i$, (ii) exogenous variation in $I_{ia}^S$, and (iii) independent variation in $G_i$ and $I_{ia}^S$ (Almond and Mazumder, 2013; Johnson and Jackson, 2019; Nicoletti and Rabe, 2014). In the following, we discuss each of these conditions, potential threats to their fulfillment, and how we address them in the context of this paper.

**(i) Exogenous variation in $G_i$.** Genetic endowments are not exogenous to family characteristics as the genetic endowments of children are drawn from the genetic pool of their biological parents.[7] As a consequence, $G_i$ is a function of maternal and paternal genetic endowments that may correlate with parental investments $I_{i1}^P, ..., I_{ia}^P$. Hence, when estimating equation (7), $\alpha$ and $\kappa$ may be confounded by *gene-environment correlation* or *genetic nurture effects* (Kong et al., 2018). Genetic nurture can be controlled for either by estimating a sibling fixed effects model that relies on within-family variation in $G_i$ only (Houmark et al., 2020; Kweon et al., 2020; Selzam et al., 2019); in a non-transmitted genes design, in which both maternal and paternal genetic endowments are included in control vector $\mathbf{X}_i(a)$; or in an adoption design, in which offspring are biologically unrelated to their parents.[8] All approaches, however, are very data demanding. For example, the sibling design requires a large sample of siblings with individual measurements of $G_i$. Therefore, it can only be applied in a limited set of existing data sets.

In this study, we estimate a between-family model using an extensive set of pre-determined family background characteristics to control for genetic nurture effects. This approach is standard in the literature and aims to approximate condition (i) while maximizing statistical power to detect the sought-after gene-environment interaction (Domingue et al., 2020). Furthermore, we formally assess the potential for residual confounding by genetic nurture effects by comparing the estimates of $\alpha$ from the between-family model to a sibling fixed effects model that we estimate for a subset of our data ($N = 525$, Appendix Table A.2). Reassuringly, the point

[7] During meiosis the chromosomes of the father and the mother are re-combined to produce genetically distinct offspring. Therefore, singleton children of the same parents are never genetically identical to their siblings. Furthermore, conditional on the parents' genome, the offspring's set of genes is randomly distributed.

[8] See Demange et al. (2020) for a detailed comparison of all three approaches.

estimates are very close to each other. This result suggests that after conditioning on $\mathbf{X}_i(a)$, residual genetic nurture is low and very unlikely to overturn our main findings.

**(ii) Exogenous variation in $I_{ia}^S$.** School characteristics are not exogenous to family characteristics as parents choose schools for their children (Altonji et al., 2005; Beuermann et al., forthcoming). As a consequence, $I_{ia}^S$ is a function of family and child characteristics that may correlate with parental investments $I_{i1}^P, ..., I_{ia}^P$. Hence, in estimation model (7), $\beta$ and $\kappa$ may be confounded by *selection effects* (Altonji et al., 2005; Altonji and Mansfield, 2018; Biasi, forthcoming). Selection into schools can be controlled in (quasi-)experimental settings, e.g. using variation based on admission lotteries (Angrist et al., 2016; Cullen et al., 2006), or the geographic design of catchment areas (Laliberté, 2021). Existing data sets that avail such variation, however, do not contain sequenced DNA data needed to measure $G_i$ at the individual level.

In this study, we use an extensive set of pre-determined family background characteristics to control for selection into schools based on observables. Furthermore, we formally assess the sensitivity of our results to residual confounding by calculating different summary statistics for selection on unobservables (Cinelli and Hazlett, 2020; Oster, 2019). Reassuringly, these summary statistics are consistent and point to low potential for selection on unobservables (Appendix Table A.2 and Appendix Figure B.2). These results suggest that after conditioning on $\mathbf{X}_i(a)$, residual selection into schools is low and very unlikely to overturn our main findings.

**(iii) Independent variation in $G_i$ and $I_{ia}^S$.** Conditions (i) and (ii) must be met in a way that $G_i$ and $I_{ia}^S$ are distributed independently of each other. Strong correlation between $G_i$ and $I_{ia}^S$ implies little variation in $G_i$ at different levels of $I_{ia}^S$ and vice versa. As a consequence, there may not be sufficient variation to identify $\alpha, \beta,$ and $\kappa$ separately from each other.

To verify that condition (iii) is satisfied, we present empirical evidence that $G_i$ and $I_{ia}^S$ are indeed distributed independently of each other. This conclusion holds both unconditionally and conditional on $\mathbf{X}_i(a)$ (Figure 3 and Appendix Table A.3).

In summary: in an ideal setting, one would estimate the complementarity parameter $\kappa$ by combining a sibling fixed effects model with experimental variation in school characteristics among children of the same biological parents. We are not aware of any data set that simultaneously includes genetic data at the individual level, a large set of siblings, and quasi-experimental within-family variation in school assignment. Therefore, we approximate the ideal-type conditions with the best data available to us. Within this setting, one must make rather strong identifying assumptions to give our estimates of $\alpha, \beta,$ and $\kappa$ a causal interpretation. However, in our empirical analysis we show extensive evidence that these assumptions are reasonably well met in our setting.

Erring on the side of caution, we speak of *associations* instead *causal effects* in the remainder of the paper. What is the direction of biases in case of violations of our identifying assumptions? Any residual confounding by genetic nurture and selection into schools will bias our estimates of $\alpha$ and $\beta$ upwards. As a consequence, bias in our estimate for $\kappa$ will also be positive (see Supplementary Material A). Therefore, in the presence of residual confounding due to genetic nurture and selection into schools, we identify a lower (upper) bound of the substitutability (complementarity) of genetic endowments and school environments.[9]

## 4 DATA

We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a 5-wave panel study that focuses on the determinants of health-related behaviors and health outcomes. Add Health is a nationally representative sample of adolescents enrolled in grades 7–12 in 1994/95. Initial information (wave 1, $N = 20,745$) was collected from a stratified sample of 80 high schools across the US as well as from their associated feeder schools. In addition to in-depth interviews with adolescents, questionnaires were administered to school representatives, parents, and roughly 90,000 students of the sampled schools. Follow-up in-home questionnaires were collected in 1996 (wave 2, $N = 14,738$), 2001/02 (wave 3, $N = 15,179$), and 2008/09 (wave 4, $N = 15,701$). In the most recent wave (2016/18, $N = 12,300$), Add Health respondents are between 33 and 43 years old.

In the following, we describe our main variables of interest. Detailed descriptions of all variables used in our analysis are disclosed in Supplementary Material B.

**Outcomes.** We measure educational attainment $Y_i$ by the total number of years of education after age 27. In each wave, respondents were asked about their highest level of education at the time of the interview. For each individual, we use the most recent information and transform education levels into years of education, following the mapping suggested by Domingue et al. (2015).[10]

To analyze the mechanisms behind our headline results, we additionally use a series of mea-

[9] In theory, this conclusion does not hold if either $\alpha$ or $\beta$ were downward biased. In practice, this case is very unlikely. First, the patterns of positive genetic nurture and positive sorting into schools are widely-documented (Altonji and Mansfield, 2018; Kong et al., 2018). Second, consistent with this evidence, we document positive bias for $\alpha$ and $\beta$ when controlling for observable covariates $\mathbf{X}_i(a)$ in our data (see Figure 2).

[10] Numeric values in parentheses: eighth grade or less (8), some high school (10), high school graduate (12), GED (12), some vocational/technical training (13), some community college (14), some college (14), completed vocational/technical training (14), associate or junior college degree (14), completed college (16), some graduate school (17), completed a master's degree (18), some post-baccalaureate professional education (18), some graduate training beyond a master's degree (19), completed post-baccalaureate professional education (19), completed a doctoral degree (20).

sures for academic degrees, health, and (non-)cognitive skills. First, academic degrees allow us to investigate at which educational stage our results emerge. We focus on whether respondents finished high school, obtained a college degree, or completed a post-graduate degree. Second, measures for health and (non-)cognitive skills serve as proxy variables for $\theta_{iA}$ and allow us to analyze the dimensions of skill development that drive the main findings on educational attainment. We proxy health by quality-adjusted life years (QALY), that we derive from self-assessed health measures as well as a summary index of diagnosed health conditions. We proxy cognitive skills using the Picture Vocabulary Test (PVT), a test of receptive hearing vocabulary that is a widely-used measure for verbal ability and scholastic aptitude. We proxy non-cognitive skills by self-reported measures of general risk aversion and patience (Falk et al., 2018) as well as self-reported information on the Big Five personality traits (Almlund et al., 2011).

**Genetic endowments.** Add Health obtained saliva samples from consenting participants in wave 4. After quality control procedures, genotyped data is available for 9,974 individuals and 609,130 SNPs. Add Health uses this data to calculate different PGS using summary statistics from existing GWAS. We use a PGS for educational attainment, referred to as $\text{PGS}^{\text{EA}}$, that is based on the GWAS by Lee et al. (2018).[11]

Lee et al. (2018) perform a meta-analysis of 71 quality-controlled cohort-level GWAS. Their meta-analysis produced association statistics for around 10 million SNPs, of which 1,271 reached genome-wide significance. Genes near these genome-wide significant SNPs are relevant for the central nervous system, and many of them encode proteins that carry out neurophysiological functions such as neurotransmitter secretion or synaptic plasticity. They are relevant for brain-development processes before and after birth.

$\text{PGS}^{\text{EA}}$ is highly predictive of educational attainment and has been widely used in existing studies. Lee et al. (2018) suggest that $\text{PGS}^{\text{EA}}$ is a better predictor for years of education than household income. Including the score in a regression of years of education on a set of controls yields an incremental $R^2$ of 0.127 in the Add Health sample. Among other uses, $\text{PGS}^{\text{EA}}$ has been used to study the formation of early childhood skills (Belsky et al., 2016), educational attainment (Domingue et al., 2015; Houmark et al., 2020), earnings (Papageorge and Thom, 2020), wealth accumulation (Barth et al., 2020), and social mobility (Belsky et al., 2018).

We standardize $\text{PGS}^{\text{EA}}$ in our analysis sample so that is has a mean of zero ($\mu = 0$) and a standard deviation of one ($\sigma = 1$).

[11] Lee et al. (2018) construct $\text{PGS}^{\text{EA}}$ for two prediction cohorts, Add Health and the Health and Retirement Study (HRS). $\text{PGS}^{\text{EA}}$ is based on results from the meta-analysis in which these two cohorts were excluded from the discovery sample. $\text{PGS}^{\text{EA}}$ was generated from HapMap3 SNPs using the software LDpred—a Bayesian method that weights each SNP by the posterior mean of its conditional effect given other SNPs.

**School investments.** In waves 1 and 2, Add Health administered detailed questionnaires to headmasters of Add Health schools. The schools are also linked to administrative data from the Common Core of Data (CCD) and the Private School Survey (PSS). We use these sources to construct indicators for $I_{ia}^{S}$ using a principal component analysis that includes the following school-level information: (i) average class size, (ii) average student-teacher ratio, (iii) share of teachers with a master degree, (iv) share of new teachers in the current school year, (v) share of teachers with school-specific tenure of more than five years, and Herfindahl indices measuring teacher diversity with respect to (vi) race and (vii) Hispanic background.

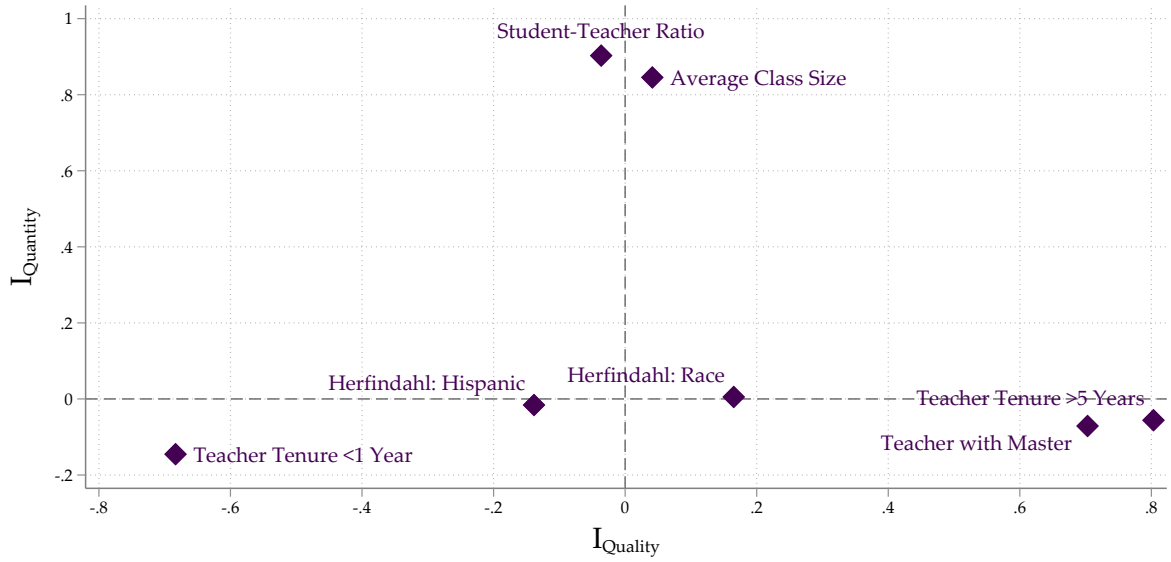
Many of these characteristics have been shown to predict teacher value-added. For example, Hanushek et al. (2016) and Ronfeldt et al. (2013) show that a high teacher turnover, which we proxy by the share of new teachers, impairs teaching quality and student achievement. Papay and Kraft (2015) and Rockoff (2004) show that teaching experience, which we proxy by the share of teachers with more than five years of tenure, correlates with teacher performance. Finally, Clotfelter et al. (2010) and Jacob et al. (2018) show that academic credentials, which we proxy by the share of teachers with a master's degree, are positively associated with teacher effectiveness.

Figure 1 shows the rotated loadings on the first two principal components. The first component loads almost exclusively on average class size and average student-teacher ratio. Hence, we interpret this component as an indicator of the "quantity" of teachers, denoted by $\mathrm{I}_{\mathrm{Quantity}}$. The second component loads positively on the percentage of teachers with a master's degree and the share of teachers with more than five years of tenure; it loads negatively on the share of new teachers in the current school year. We interpret this component as an indicator of the "quality" of teachers, denoted by $\mathrm{I}_{\mathrm{Quality}}$. Both factors are coded such that higher values indicate higher school investments, i.e., higher investments in teacher "quantity" (smaller classes) and higher teacher "quality" (better teachers), respectively. The calculated factors are orthogonal to each other by construction and standardized to $\mu = 0$ and $\sigma = 1$.[12] Furthermore, note that $\mathrm{I}_{\mathrm{Quantity}}$ and $\mathrm{I}_{\mathrm{Quality}}$ are time-invariant measures for each school.

**Control variables.** Add Health provides extensive information on the environments to which respondents were exposed during childhood. We approximate the identification pre-requisites discussed in section 3 by choosing a vector of pre-determined variables $\mathbf{X}_i(a)$ to control for genetic nurture effects and selection into schools. Specifically, we control for family background characteristics by including maternal and paternal education (in years), family religious affiliation (Christian/non-Christian), parental birth place (US/non-US), and maternal age at birth

[12] Intuitively, one may expect a negative correlation between teacher quality and quantity: given a budget, a school administrator may prefer to invest in teacher quality at the expense of average class size, or vice versa. However, this is not what we observe in the data. If quality and quantity were substitutes, we would expect the loadings on the two principal components to pull in diametrically opposed directions. To the contrary, we find that the variables capturing the quality and quantity dimensions are orthogonal to each other and almost exclusively load on only one principal component.

**FIGURE 1 – Rotated Loadings on Factors for School Characteristics**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows the rotated factor loadings on $I_{Quality}$ and $I_{Quantity}$. The principal component analysis is conducted using the following school-level information: (i) average class size, (ii) average student-teacher ratio, (iii) share of teachers with a master's degree, (iv) share of new teachers in the current school year, (v) share of teachers with school-specific tenure of more than five years, and Herfindahl indices to measure teacher diversity with respect to (vi) race and (vii) Hispanic background.

(in years). Furthermore, we include the mean and standard deviation of potential wages for both mothers and fathers across child ages 0–14.[13] At the level of children, we control for a firstborn indicator, age in months, biological sex, as well as their interaction. We follow standard practice in the literature and account for population stratification in genetic endowments by including the first 20 principal components of the full matrix of genetic data. Lastly, all estimations include a vector of state fixed effects.

Note that we focus on pre-determined variables—variables that are fixed prior to the period of observation—to avoid smearing through "bad controls" (Angrist and Pischke, 2009). However, in robustness analyses we expand the vector of controls to include potentially endogenous parental investments and family income. Our results remain unaffected.

**Analysis sample.** We apply the following sample selection criteria. First, we restrict our sample to genotyped respondents of European descent.[14] This is common practice in the literature

[13]Note that Add Health contains information on actual income. However, actual income may be a bad control as it reflects parental responses to both $G_i$ and $I_{ia}^S$. Therefore, we follow the procedure of Shenhav (2021) and combine data from the 1970 Census and the March Current Population Survey (1975–2000) to construct potential wages for gender/education/census region/race/ethnicity cells and match these potential wages to parents of children aged $a = 1, ..., 14$.

[14]The ancestry groups in Add Health are identified by principal component analysis on all unrelated members of the full Add Health genotyped sample.

because GWAS are predominantly conducted on this ancestry group. As a consequence, there is a lack of statistical power to account for population stratification between ancestry groups and estimates of genetic influence would be biased without this restriction (Martin et al., 2017; Ware et al., 2017).

Second, we retain the subsample of individuals who (i) visited an Add Health high school or an associated feeder school in wave 1, and (ii) for whom the high school exit record indicates that they graduated from the same school. These sample selection criteria strike a balance between sample size and the accuracy of matching individuals with our measures for schooling environments. For example, when applying criterion (i), we assume that individuals do indeed transfer from feeder schools to designated Add Health schools. While this increases our sample, it is possible that the information on $I_{ia}^{S}$ is incorrectly attributed to individuals transferring to high schools outside the Add Health universe. Reversely, when we apply criterion (ii), we exclude individuals that may have moved to other high schools during grades 9–12. In this way, we reduce our sample size but minimize the risk that information on $I_{ia}^{S}$ is misattributed to transfer students. We note that neither strengthening (i) by excluding individuals from feeder schools, nor relaxing (ii) by assuming that individuals remain in the same school through grades 9–12 overturns our main conclusions (Appendix Table A.6).

Third, we drop all observations with missing information in $Y_i$, $G_i$, $I_{ia}^{S}$, and $\mathbf{X}_i(a)$ by list-wise deletion.

Applying these restrictions, we obtain a sample of $3,075$ individuals from 77 high schools across the US, for which we provide summary statistics in Table 1. 55% are female, and the average age measured at wave 1 equals $\approx$ 16 years (194 months). The average educational attainment in our sample is 14.8 years which exceeds the average educational attainment in the parental generation by $\approx$ 1.1 years. Almost all individuals graduate from high school, which is not surprising given that our sample is restricted to individuals of European descent who remained in the same high school in grades 9–12. The college completion rate equals $\approx$ 50%.

To assess sample representativeness, we compare our analysis sample to the 1974–1983 birth cohorts of non-Hispanic Whites in the American Community Survey (ACS) and the Current Population Survey (CPS) (Appendix Table A.1). This comparison shows a slight over-representation of females and children of young mothers in our sample. Otherwise, our sample is by and large comparable to the corresponding groups in the ACS and CPS. In robustness analyses, we reweight our analysis sample to match the ACS and CPS with respect to gender composition, educational attainment of parents, and the age of mothers at birth. Our results remain unaffected (Appendix Table A.6).

**TABLE 1 – Summary Statistics**

| | N=3,075; Siblings=525; High Schools=77 | | | |
|---|---|---|---|---|
| | Mean | SD | Min | Max |
| *Educational Attainment* | | | | |
| Years Education | 14.81 | 2.25 | 8.00 | 20.00 |
| High School Degree | 0.97 | 0.18 | 0.00 | 1.00 |
| 2-year College Degree | 0.53 | 0.50 | 0.00 | 1.00 |
| 4-year College Degree | 0.42 | 0.49 | 0.00 | 1.00 |
| Post-Graduate Degree | 0.15 | 0.36 | 0.00 | 1.00 |
| *Variables of Interest* | | | | |
| $PGS^{EA}$ | 0.00 | 1.00 | -4.18 | 3.36 |
| $I_{Quality}$ | 0.00 | 1.00 | -3.41 | 1.92 |
| $I_{Quantity}$ | 0.00 | 1.00 | -3.22 | 3.20 |
| *Child Background Characteristics* | | | | |
| Female | 0.55 | 0.50 | 0.00 | 1.00 |
| Firstborn | 0.47 | 0.50 | 0.00 | 1.00 |
| Age in Months (Wave 1) | 193.63 | 19.77 | 144.00 | 256.00 |
| Maternal Age at Birth | 25.49 | 4.83 | 16.00 | 44.33 |
| Christian | 0.82 | 0.38 | 0.00 | 1.00 |
| Education Mother (in Years) | 13.63 | 2.50 | 8.00 | 19.00 |
| Education Father (in Years) | 13.67 | 2.68 | 8.00 | 19.00 |
| Foreign-born Mother | 0.03 | 0.17 | 0.00 | 1.00 |
| Foreign-born Father | 0.03 | 0.16 | 0.00 | 1.00 |
| Potential Wage/Hour Mother | 12.62 | 1.38 | 9.45 | 14.27 |
| Potential Wage/Hour Father | 15.48 | 1.31 | 11.14 | 17.11 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows summary statistics for the core analysis sample. The sample is restricted to genotyped individuals who (i) are of European descent, (ii) attended an Add Health high school or an associated feeder school in wave 1, and (iii) graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion.

## 5 RESULTS

We present our results in four steps. In section 5.1, we discuss the association of educational attainment, genetic endowments, and school investments in light of the identifying assumptions discussed in section 3. In section 5.2, we present our estimates for the complementarity parameter $\kappa$. After a robustness analysis in section 5.3, we conclude with an analysis of mechanisms in section 5.4.

### 5.1 *The Association of Educational Attainment with Genetic Endowments and School Investments*

Figure 2 visualizes the association of educational attainment with our measures for genetic endowments $G_i$ and school investments $I_{ia}^S$. In the left column we show raw correlations that do not account for the control variables $\mathbf{X}_i(a)$. In the right column, we show associations conditional on $\mathbf{X}_i(a)$.

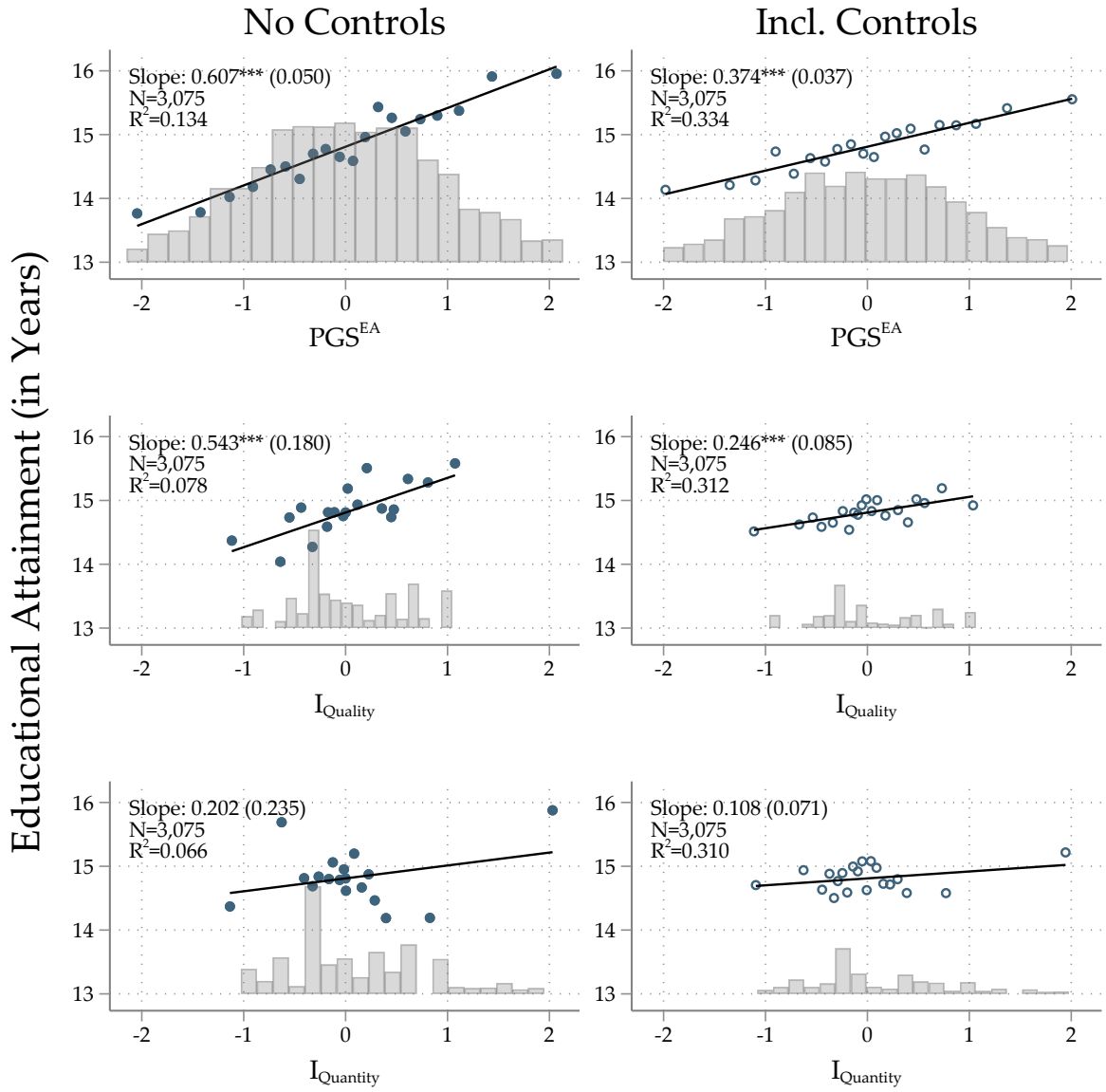
First, $\text{PGS}^{\text{EA}}$ is highly predictive of educational attainment. Without controls, a one-standard-deviation (1 SD) increase in $\text{PGS}^{\text{EA}}$ is associated with an increase in educational attainment of 0.607 years. This association does not have a causal interpretation as it may be confounded by genetic nurture effects. When we control for pre-determined child and family characteristics, a 1 SD increase in $\text{PGS}^{\text{EA}}$ is associated with an increase in educational attainment of 0.374 years. This decrease is consistent with sibling studies showing that genetic nurture effects usually account for 40–50% of the raw association between $\text{PGS}^{\text{EA}}$ and educational attainment (Kweon et al., 2020; Muslimova et al., 2020; Ronda et al., forthcoming; Selzam et al., 2019).

Is $\mathbf{X}_i(a)$ sufficient to control for genetic nurture effects? We test whether there is residual confounding due to genetic nurture by comparing estimates of the between-family model with a sibling fixed effects model that we estimate on a subsample of our data ($N = 525$). The within-family comparison allows us to perfectly control for genetic nurture effects. Therefore, a strong divergence of between- and within-family estimates would suggest that there is residual genetic nurture that is not picked up by $\mathbf{X}_i(a)$. In Appendix Table A.2 we show that this is not the case. The within-family comparison yields a point estimate of 0.394 that is significant at the 1%-level. This point estimate is very close to the between-family estimate after controlling for $\mathbf{X}_i(a)$ and lends further credence to our research design.

Second, $\text{I}_{\text{Quality}}$ is highly predictive of educational attainment. Without controls, a 1 SD increase in $\text{I}_{\text{Quality}}$ is associated with an increase in educational attainment of 0.543 years. This association does not have a causal interpretation as it may be confounded by selection effects. When we control for pre-determined child and family characteristics, a 1 SD increase in $\text{I}_{\text{Quality}}$ is associated with an increase in educational attainment of 0.246 years. This 55% decrease reflects positive selection into schools based on "teacher quality"—a pattern that has been thoroughly documented in the existing literature for the US (Biasi, forthcoming). Nevertheless, even when accounting for selection, the association of $\text{I}_{\text{Quality}}$ and educational attainment remains strong and positive. This result is consistent with prior literature that has repeatedly demonstrated positive effects of teacher quality on students' educational success (Chetty et al., 2014a; Hanushek and Rivkin, 2010).

Is $\mathbf{X}_i(a)$ sufficient to control for selection into schools? We test whether there is any residual confounding due to selection effects by re-running our estimation, replacing educational attain-

**FIGURE 2 – Association of Educational Attainment with $PGS^{EA}$, $I_{Quality}$, and $I_{Quantity}$**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure visualizes the correlation of completed years of education with $PGS^{EA}$, $I_{Quality}$, and $I_{Quantity}$, respectively. We bin scatterplots using 20 quantiles of the variable of interest. Gray bars indicate density distributions of the (residualized) variable of interest. Black lines are fitted from linear regressions of educational attainment on the variable of interest. In the left-column, we only control for state fixed effects. In the right column, we introduce the full set of control variables. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the school level.

ment with predicted educational attainment. In particular, we predict educational attainment using a fully interacted model of an indicator for whether the child was breastfed, log family income, and maternal and paternal time investments.[15] On the one hand, breastfeeding,

[15]We collect information on a series of activities that the child has engaged in with their mother or father over

parental income, and time investments are important predictors of children's educational outcomes (Agostinelli and Sorrenti, 2018; Kosse, 2016).[16] On the other hand, we do not use these predictors in our estimates but treat them "as if unobservable." Therefore, we can interpret the association between predicted educational attainment and $I_{Quality}$ as a summary measure of selection on unobservables: a large and positive association between predicted educational attainment and $I_{Quality}$ would suggest that there is residual selection into schools that is not picked up by $\mathbf{X}_i(a)$. In Appendix Table A.2 we show that this is not the case. The point estimate of 0.076 is small and insignificant at conventional levels of statistical significance. Therefore, this test does not point to the presence of strong selection effects after conditioning on $\mathbf{X}_i(a)$. Alternatively, we can assess any residual selection effects by assuming that changes in the coefficient of $I_{Quality}$ due to the introduction of $\mathbf{X}_i(a)$ provide information about the extent of confounding due to unobservables (Altonji et al., 2005; Cinelli and Hazlett, 2020; Oster, 2019). We follow Cinelli and Hazlett (2020) and assess the strength of the association that unobserved confounders would need to have with $I_{Quality}$ and educational attainment to change our conclusions. In Appendix Figure B.2, we show that $I_{Quality}$ would remain positive and statistically significant at the 5%-level even if the partial $R^2$ of unobserved confounders with $I_{Quality}$ and educational attainment were more than six times higher than the corresponding partial $R^2$ of paternal education. Given the decisive role of parental education in school choices, and its strong predictive power for educational outcomes of children, this results lends further confidence that our results are genuine and not a mere reflection of selection into schools based on family background.

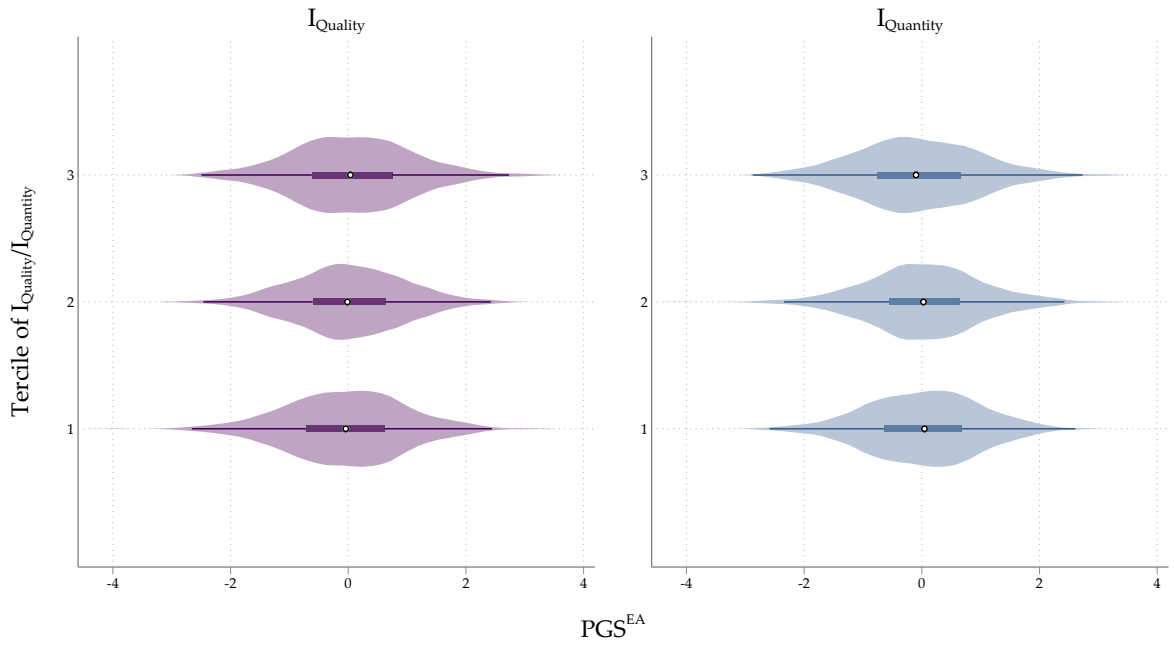
Third, $I_{Quantity}$ is not significantly associated with educational attainment. The weak positive correlation is imprecisely estimated and does not attain statistical significance at conventional levels. Furthermore, this result does not change when accounting for selection effects by introducing control vector $\mathbf{X}_i(a)$. This finding is in line with prior literature which has not been able to establish consistent effects of teacher quantity on students' educational success (Angrist et al., 2019; Fredriksson et al., 2013; Leuven and Løkken, 2020). However, this average association may mask heterogeneity among students with different genetic endowments—a hypothesis that we test in the following subsection.

In addition to genetic nurture effects and selection effects, a high correlation between $G_i$ and $I_{ia}^S$ would pose another threat to the identification of the gene-environment interaction. Figure 3 shows that this threat does not exist in our setting. In this figure, we plot the unconditional $PGS^{EA}$ distribution by tercile of $I_{Quality}$ and $I_{Quantity}$, respectively. Visual inspection suggests that the $PGS^{EA}$ distributions are almost congruent to each other within each tercile of the two

the past four weeks. These activities include shopping, playing sports, going to church, talking about dates, going to the movies and similar events, talking about personal problems, arguing, talking about school work, working together on school work, and talking about other things at school. Following Anderson (2008) and Kling et al. (2007) we standardize each response dimension to $\mu = 0$ and $\sigma = 1$ and sum them linearly by parent to obtain aggregate indexes of time investment. See Supplementary Material B for details.

[16] The $R^2$ of the prediction regression is 0.185.

**FIGURE 3 – $PGS^{EA}$ Distribution by $I_{Quality}$ and $I_{Quantity}$**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows unconditional $PGS^{EA}$ distribution by terciles of both $I_{Quality}$ and $I_{Quantity}$. The central point indicates the median, and the bar indicates the interquartile range. The density distributions represent estimated Epanechnikov kernel densities.

indicators. In Appendix Table A.3, we present formal statistical tests for this observation. In particular, we residualize $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ using control vector $\mathbf{X}_i(a)$. We then perform two-sample Kolmogorov-Smirnov tests for the equality of $PGS^{EA}$ distributions within the terciles of $I_{Quality}$ and $I_{Quantity}$, respectively. Only one of the comparisons is borderline significant at conventional levels of statistical significance. Hence, we conclude that $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ are indeed independently assigned.

## 5.2 *The Interplay of Genetic Endowments and School Investments in the Production of Educational Attainment*

Table 2 shows our baseline estimates for the interaction of genetic endowments and school investments. In all regressions, we include the vector $\mathbf{X}_i(a)$ to control for genetic nurture and selection into schools.

In column (1), we focus on the teacher quality indicator $I_{Quality}$. The point estimates for $PGS^{EA}$ and $I_{Quality}$ replicate the findings from Figure 2 and show a strong and positive association of $PGS^{EA}$ and $I_{Quality}$ with educational attainment. A 1 SD increase in $PGS^{EA}$ ($I_{Quality}$) increases educational attainment by $\approx 0.37$ ($\approx 0.22$) years.

**TABLE 2 – Association of $PGS^{EA}$ and School Environments with Years of Education**

| Outcome: Years of Education | (1) | (2) | (3) |
|---|---|---|---|
| $PGS^{EA}$ | 0.368*** | 0.370*** | 0.365*** |
| | (0.033) | (0.037) | (0.033) |
| $I_{Quality}$ | 0.219*** | – | 0.214*** |
| | (0.078) | | (0.079) |
| $PGS^{EA} \times I_{Quality}$ | -0.075** | – | -0.074** |
| | (0.033) | | (0.033) |
| $I_{Quantity}$ | – | 0.078 | 0.078 |
| | | (0.063) | (0.053) |
| $PGS^{EA} \times I_{Quantity}$ | – | 0.037 | 0.027 |
| | | (0.037) | (0.033) |
| Child Controls | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ |
| N | 3,075 | 3,075 | 3,075 |
| $R^2$ | 0.338 | 0.335 | 0.339 |
| Outcome Mean | 14.811 | 14.811 | 14.811 |
| Outcome SD | 2.249 | 2.249 | 2.249 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level. Standard errors of bias adjusted treatment effects are bootstrapped based on 200 draws.

$PGS^{EA} \times I_{Quality}$ is our estimate for the complementarity parameter $\kappa$. The negative coefficient of the interaction term indicates that genetic endowments and teacher quality act as *substitutes* in the production of educational attainment. A 1 SD increase in teacher quality reduces the positive association of educational attainment and $PGS^{EA}$ by $\approx 20\%$ (= 0.08/0.37). This result is in notable contrast to the existing literature investigating the gene-environment interaction between $PGS^{EA}$ and parental socio-economic status, which tend to act as *complements* in the production of educational attainment (Papageorge and Thom, 2020; Ronda et al., forthcoming). Furthermore, we highlight that the estimated substitutability is likely a lower bound: positive bias in the effects of $PGS^{EA}$ and $I_{Quality}$ due to genetic nurture and selection into schools not captured by $\mathbf{X}_i(a)$, results in upward bias in the corresponding gene-environment interaction (Appendix A). Therefore, in the presence of residual confounding due to genetic nurture and selection into schools, our estimate of $\kappa$ would be biased towards zero.

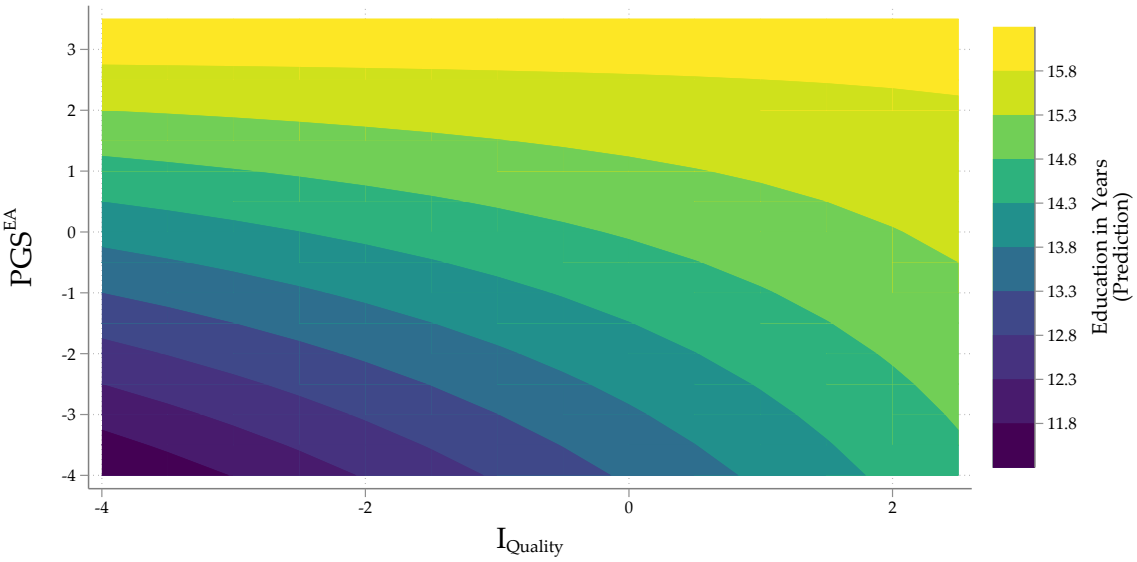
In column (2), we focus on $I_{Quantity}$. The point estimate for $I_{Quantity}$ is again statistically indistinguishable from zero. The estimate for $PGS^{EA} \times I_{Quantity}$ indicates that this null result is not driven by heterogeneity along the $PGS^{EA}$ distribution. Our estimate for the complementarity

parameter $\kappa$ is small and not statistically different from zero.

In column (3), we estimate both complementarity parameters in the same model and show that our results remain virtually unchanged. This stability is to be expected since $I_{Quality}$ and $I_{Quantity}$ are distributed independently of each other by construction.

What drives the substitutability of $PGS^{EA}$ and $I_{Quality}$? In principle, the negative gene-environment interaction could be due to low $PGS^{EA}$ students benefiting from higher-quality teachers, or high $PGS^{EA}$ students losing from higher-quality teachers. In Figure 4, we provide evidence for the former, but not for the latter. In this figure, we show years of education as predicted from the

**FIGURE 4 – Association of $PGS^{EA}$ with Years of Education by $I_{Quality}$**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows a prediction of completed years of education by $PGS^{EA}$ and $I_{Quality}$ cell. Predictions are calculated using the model estimated in column (3) of Table 2.

estimates in column (3) of Table 2. Moving horizontally from left to right at a given $PGS^{EA}$ level, we see that predicted education increases sharply in the lower parts of the $PGS^{EA}$ distribution. To the contrary, in the upper ranges of the $PGS^{EA}$ distribution, predicted education remains virtually unchanged regardless of the quality of teachers at a given school. This pattern is encouraging as it suggests that investments into teacher quality mitigate inequality in educational outcomes without compromising the attainment of genetically advantaged students.

In Appendix Table A.5, we decompose this effect into the different school characteristics underlying the construction of $I_{Quality}$. We show that the effects of $I_{Quality}$ are especially driven by the share of teachers with school-specific tenure of more than five years. This measure combines

information about teaching experience with information about teacher turnover. On the one hand, teachers have taught for at least five years in their lives and do not suffer from a lack of basic experience. On the other hand, longer tenure at a given school increases the likelihood of student-teacher re-matches in consecutive school years. Both features tend to have positive effects on student learning and are therefore plausible candidates to rationalize the relative educational gains of students in the lower tail of the $PGS^{EA}$ distribution (Hill and Jones, 2018; Hwang et al., 2021; Jackson and Bruegmann, 2009; Rockoff, 2004).

### *5.3 Robustness Analysis*

We check the robustness of our results in two steps. First, we investigate whether $I_{Quality}$ and $I_{Quantity}$ pick up the effects of other school characteristics that may correlate with student outcomes. Second, we test whether our estimates of the complementarity parameter $\kappa$ are confounded by the heterogeneity of genetic effects in different family environments.

**Other school characteristics.** First, in Figure 2 we document positive sorting into schools based on $I_{Quality}$. As a consequence, students in schools with high-quality teachers may additionally be exposed to a more favorable composition of their peer group. It has been widely documented in the literature that skill formation is influenced by school peers (Bietenbeck, 2019; Isphording and Zölitz, 2020; Sacerdote, 2014).[17] Hence, our results for $I_{Quality}$ may reflect both the teacher quality and peer group composition. To test this hypothesis, we use Add Health's in-school questionnaire that elicits background information from a total of 90,000 students in the sampled schools. Based on this questionnaire, we calculate proxy indicators for the quality of the students' peers. In particular, we calculate (i) average years of paternal education, (ii) the share of single parent families, and (iii) students' average self-assessment of the probability of obtaining a college degree.[18] Then, we incorporate these indicators as well as their interaction with $PGS^{EA}$ into our estimation model.

Table 3 displays the results. Column (1) replicates our baseline estimates. In columns (2)–(4), we sequentially introduce the peer quality indicators as well as their interaction with $PGS^{EA}$. Each proxy for peer quality is highly predictive of educational attainment. For example, a 1 SD increase in the average paternal education of peers is associated with a 0.26 increase in years of education. Importantly, however, for all considered peer quality indicators, our conclusions regarding $I_{Quality}$, $I_{Quantity}$, and their interaction with genetic endowments remain unaffected.

Second, $I_{Quality}$ and $I_{Quantity}$ may be correlated with school rules and sanction policies. Exist-

[17] Sotoudeh et al. (2019) show genetic endowments of peers are also associated with individual outcomes.

[18] To avoid mechanical relationships between individual characteristics and peer group composition we calculate leave-one-out (jackknife) indicators. A detailed description of these variables is disclosed in Supplementary Material B.

**TABLE 3 – Robustness to Additional School Characteristics**

| | Baseline | + School Peer Characteristics | | | + School Sanction Policies | | | + School VA (GPA) |
|---|---|---|---|---|---|---|---|---|
| Outcome: Years of Education | (1) | Educ. Father (2) | Single Parents (3) | College Aspir. (4) | Drugs (5) | Social (6) | Acad. (7) | (8) |
| $PGS^{EA}$ | 0.365*** | 0.355*** | 0.363*** | 0.355*** | 0.364*** | 0.361*** | 0.362*** | 0.360*** |
| | (0.033) | (0.035) | (0.035) | (0.034) | (0.035) | (0.035) | (0.035) | (0.035) |
| $I_{Quality}$ | 0.214*** | 0.142** | 0.210*** | 0.202*** | 0.199** | 0.188** | 0.220** | 0.164** |
| | (0.079) | (0.068) | (0.077) | (0.070) | (0.086) | (0.084) | (0.088) | (0.080) |
| $PGS^{EA} \times I_{Quality}$ | -0.074** | -0.078** | -0.074** | -0.078** | -0.073** | -0.075** | -0.069** | -0.073** |
| | (0.033) | (0.036) | (0.033) | (0.034) | (0.035) | (0.035) | (0.035) | (0.035) |
| $I_{Quantity}$ | 0.078 | -0.014 | 0.006 | 0.000 | 0.071 | 0.051 | 0.050 | 0.033 |
| | (0.053) | (0.055) | (0.052) | (0.052) | (0.058) | (0.054) | (0.057) | (0.062) |
| $PGS^{EA} \times I_{Quantity}$ | 0.027 | 0.018 | 0.030 | 0.025 | 0.027 | 0.027 | 0.022 | 0.032 |
| | (0.033) | (0.029) | (0.033) | (0.031) | (0.034) | (0.033) | (0.033) | (0.035) |
| School Characteristic | – | 0.257*** | -0.204*** | 0.209*** | 0.003 | -0.106 | 0.059 | 0.115* |
| | | (0.053) | (0.046) | (0.044) | (0.052) | (0.074) | (0.072) | (0.061) |
| $PGS^{EA} \times$ School Characteristic | – | -0.048 | 0.030 | -0.049 | 0.005 | 0.025 | 0.037 | -0.020 |
| | | (0.043) | (0.039) | (0.035) | (0.041) | (0.030) | (0.037) | (0.031) |
| Child Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 3,075 | 2,959 | 2,959 | 2,959 | 2,993 | 2,993 | 2,993 | 2,768 |
| $R^2$ | 0.339 | 0.347 | 0.347 | 0.346 | 0.341 | 0.342 | 0.342 | 0.319 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.

**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. We control for additional school characteristics and their interaction with $PGS^{EA}$. The relevant school characteristics are indicated in the column header. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0$, $\sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

ing literature suggests that school rules may promote educational attainment (Bacher-Hicks et al., 2019). For example, the success of charter schools has been attributed to strict "no excuses" policies (Angrist et al., 2013). Hence, our results for $I_{Quality}$ may reflect both the quality of teachers and the effects of school rules. To test this hypothesis, we exploit information from headmaster questionnaires and conduct a principal component analysis on various school policies.[19] We extract three components that reflect the school's strictness regarding (i) drug use, (ii) social misconduct, and (iii) academic misconduct.

[19] In wave 1, headmasters were asked about the school's policy in the following domains of behavior: cheating, fighting with or injuring another student, alcohol or drug possession, drinking alcohol or using illegal drugs, smoking, verbally or physically abusing a teacher, and stealing school property. Possible measures are (i) no action, (ii) verbal warning, (iii) minor action, (iv) in-school suspension, (v) out-of-school suspension, and (vi) expulsion. A detailed description of these variables is disclosed in Supplementary Material B.

In columns (5)–(7) of Table 3, we sequentially introduce the strictness indicators as well as their interaction with $PGS^{EA}$. None of the indicators is predictive of educational attainment, nor is there an interaction with genetic endowments. Our conclusions with respect to $I_{Quality}$, $I_{Quantity}$, and their interaction with genetic endowments remain unaffected.

Third, there may be unobservable school characteristics that drive the relationship between $I_{Quality}$, $I_{Quantity}$, and educational attainment. To address this concern, we use transcript records from grades 9–12 of roughly 12,000 Add Health respondents to calculate cohort-specific measures of school value-added in GPAs for Science, Math, and English. In the extant literature, value-added measures are mostly calculated with respect to test scores that are unaffected by evaluation biases of teachers. To the contrary, GPAs capture student progress in cognitive and behavioral outcomes as well as teacher perceptions (Jackson, 2019). In spite of these intricacies, GPAs are highly predictive of long-term student outcomes (Borghans et al., 2016; Kirkebøen, 2021). Therefore, GPA-based value-added measures are a good way to capture the quality of schooling environments beyond the measures reported in headmaster surveys and administrative data. Specifically, we follow the indirect calculation procedures proposed by Chetty et al. (2014a) and Jackson et al. (2020): we residualize subject-specific GPAs from lagged GPAs in English, Math, and Science, lagged and contemporaneous measures of tracks in these subjects, and a rich set of individual background characteristics. In turn, we sum the residuals to calculate school-times-cohort fixed effects. To avoid mechanical relationships between individual outcomes and cohort-specific school effects, we calculate leave-cohort-out predictions, giving greater weight to neighboring cohorts. We calculate these measures separately for each subject, but summarize the school-specific information by extracting the first principal component from the three value-added measures (see Supplementary Material B for details).

In column (8) of Table 3, we introduce school value-added as well as its interaction with $PGS^{EA}$ as additional controls. While school value-added is indeed predictive of educational attainment, there is no effect heterogeneity across the $PGS^{EA}$ distribution. Furthermore, the associations of $I_{Quality}$, $I_{Quantity}$, and $PGS^{EA}$ with educational attainment remain unaffected. Hence, we find no evidence that our relationships of interest are confounded by unobservable school characteristics.

**Family environments and behavioral responses.** In our baseline analysis, we control for a rich set of parental background characteristics to control for genetic nurture effects and selection into schools. However, even if we were able to perfectly control for these confounding factors, the complementarity parameter for genetic endowments and school investments may still be susceptible to bias from the following sources: (i) heterogeneity in behavioral adjustments that arise in response to observed genetic endowments and schooling environments (Biroli et al., 2022), and (ii) gene-environment interactions with family characteristics and investments (Domingue et al., 2020; Keller, 2014). For example, highly educated parents may try to compensate for poor schooling environments of their children by helping with homework,

providing additional educational stimuli etc. Since parental education correlates with $PGS^{EA}$, similar parental responses could also explain the negative gene-environment interaction between $PGS^{EA}$ and $I_{Quality}$.

Table 4 presents results that address such concerns. Column (1) replicates our baseline estimates. In column (2), we test for potential confounding due to behavioral responses. In particular, we follow Biroli et al. (2022) and include a set of higher-order polynomials for $PGS^{EA}$, $I_{Quality}$, $I_{Quantity}$ and their mutual interactions to absorb heterogeneity in behavioral responses. Despite a drop in precision, our results remain unaltered. In column (3) we test for potential confounding due to gene-environment interactions with family characteristics and investments. In particular, we follow Keller (2014) and extend our estimation model by interacting genetic endowments and school environments with the full control vector $\mathbf{X}_i(a)$. In doing so, we allow for the possibility that family socio-economic status interacts with both genetic endowments and school investments. Our conclusions with regard to $I_{Quality}$, $I_{Quantity}$, and their interaction with $PGS^{EA}$ remain unaffected.

Lastly, in column (4) of Table 4, we again estimate a fully interacted model but also incorporate controls for family environments that are potentially endogenous to $PGS^{EA}$ and schooling environments. In particular, we include indexes for breastfeeding, maternal and paternal time investments, and the log of annual family income. Despite a decrease in sample size and the associated loss in precision, our results remain unaffected.

Overall, these results provide us with confidence that our estimates for the complementarity parameter $\kappa$ are not confounded by the heterogeneity in genetic effects across family environments.

**Further robustness checks.** In Supplementary Figures B.3-B.5, we show the results of additional robustness analyses. First, we check whether our results are driven by outlier schools. Therefore, we re-run our analysis 77 times, excluding one school from the sample per iteration. Reassuringly, the results are very close to our benchmark estimates in each iteration (Supplementary Figure B.3). Second, we check whether our findings are driven by ceiling effects in educational attainment. To this end, we re-run our analysis, artificially censoring educational attainment stepwise from above. If ceiling effects were driving our results, we would expect the interaction between $I_{Quality}$ and $PGS^{EA}$ to increase across parts of the censoring interval. However, this is not the case. Instead, the corresponding coefficient decreases monotonically (Supplementary Figure B.4). Third, we conduct a placebo analysis. To this end, we re-run our analysis for 20 additional outcomes for which we do not expect a differential impact of teacher quality depending on genetic endowments. For only one out of these 20 outcomes do we detect a statistically significant interaction effect at the 10%-level (Supplementary Figure B.5). In summary, these additional checks further support the conclusion that our main findings are genuine.

**TABLE 4 – Robustness to Family Environments and Behavioral Responses**

| Outcome: Years of Education | Baseline (1) | Higher-Order Polynomials (2) | Full Interaction (3) | Endogenous Controls (4) |
|---|---|---|---|---|
| $PGS^{EA}$ | 0.365*** | 0.486*** | 0.364*** | 0.370*** |
| | (0.033) | (0.060) | (0.031) | (0.044) |
| $I_{Quality}$ | 0.214*** | 0.189* | 0.202** | 0.272*** |
| | (0.079) | (0.108) | (0.079) | (0.093) |
| $PGS^{EA} \times I_{Quality}$ | -0.074** | -0.070** | -0.095*** | -0.093** |
| | (0.033) | (0.035) | (0.035) | (0.043) |
| $I_{Quantity}$ | 0.078 | -0.115 | 0.082 | 0.072 |
| | (0.053) | (0.082) | (0.060) | (0.079) |
| $PGS^{EA} \times I_{Quantity}$ | 0.027 | 0.031 | 0.024 | 0.072 |
| | (0.033) | (0.034) | (0.031) | (0.079) |
| Child Controls | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ |
| Higher-Order Polynomials | × | ✓ | × | × |
| Full Interaction | × | × | ✓ | ✓ |
| Endogenous Controls | × | × | × | ✓ |
| N | 3,075 | 3,075 | 3,075 | 2,109 |
| $R^2$ | 0.339 | 0.342 | 0.359 | 0.385 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. In column (2) we control for second- and third-order polynomials of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$, and interactions of their second order polynomials. In column (3) we control for all possible interactions between $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ and the control variables. In column (4) we introduce potentially endogenous control variables. Endogenous control variables include an indicator for breastfeeding, indexes for maternal time investments, paternal time investments, and log family income. We also allow for all possible interactions between $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ and all (endogenous) control variables. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

### *5.4 Mechanisms*

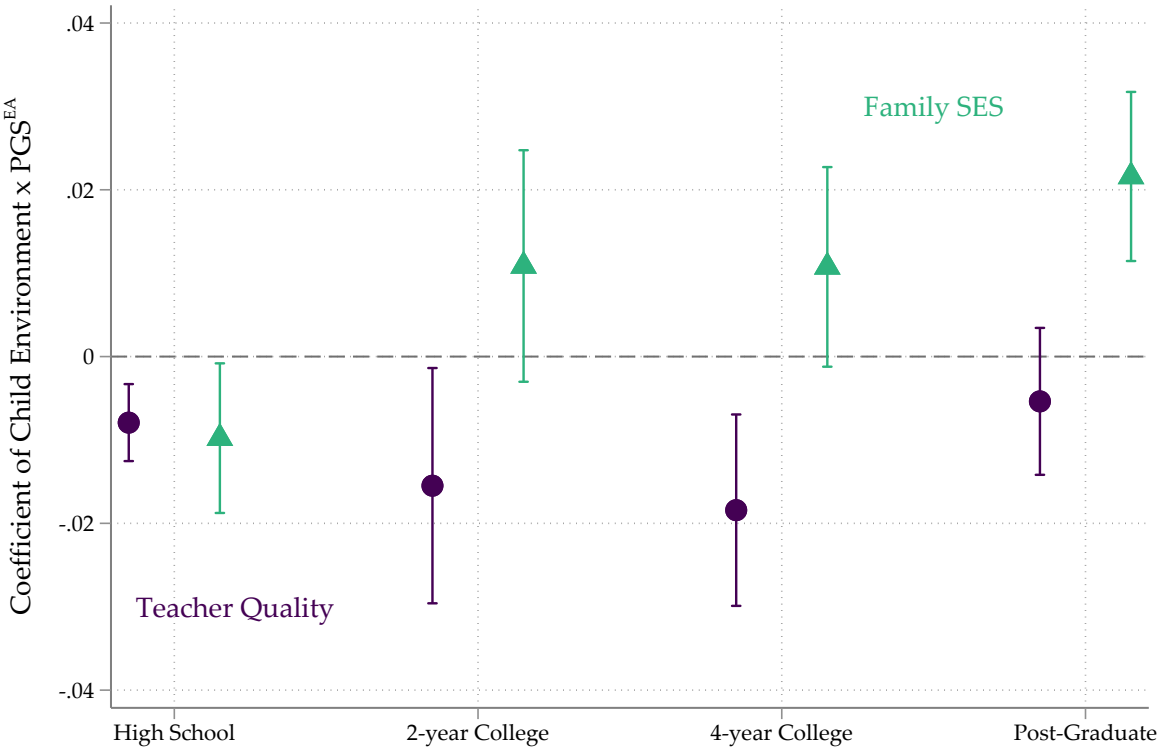
In this section, we analyze the mechanisms underlying the substitutability of genetic endowments and teacher quality. We abstract from $I_{Quantity}$ given its robust non-association with educational outcomes (see sections 5.1–5.3).

**Educational degrees.** Total years of education summarizes information from various educational stages, with each stage requiring a different mix of skills $\theta_i$ (Cunha et al., 2006, 2010). Therefore, we repeat our analysis by replacing total years of education with binary variables for whether respondents obtained (i) at least a high school degree or GED, (ii) a 2-year college

degree, (iii) a 4-year college degree, or (iv) a post-graduate degree.

In Figure 5, we present the resulting point estimates for the complementarity parameter $\kappa$ and the associated confidence bands.

**FIGURE 5 – Association of $PGS^{EA}$ and School/Family Environments with Degree Attainment**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows point estimates and 90% confidence bands for the interaction between $PGS^{EA}$ and school/family environments during childhood and their association with completed levels of education. Estimates follow the specification of equation (7). *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. Standard errors are clustered at the school level.

The circular series suggests that the compensating effect of teacher quality follows has a U-shaped pattern throughout the educational life cycle. The probability of dropping out of high school decreases slightly, followed by an increase in substitutability for 2-year and 4-year college degrees. The substitutability of high-quality teachers and genetic endowments levels off at the post-graduate level. This pattern is consistent with the following interpretation. High school graduation is a relatively "inclusive" educational outcome that is accessible to most, including low $PGS^{EA}$ students in low-quality schooling environments. Evidence of this effect

is the high school graduation rate of 97% in our sample (Table 1). In contrast, post-graduate education is a relatively "exclusive" educational outcome that is more accessible to students with a advantageous genetic endowments *and* who experienced conducive environments. In both cases, there is limited opportunity for high-quality teachers to make a difference for low $PGS^{EA}$ students. We interpret this pattern as suggestive since the confidence bands are wide and we cannot statistically distinguish among the point estimates for different educational degrees.

The triangular series shows that the gene-environment interplay is very different for schooling environments and family environments. Consistent with Buser et al. (2021a) and Papageorge and Thom (2020), the complementarity between genetic endowments and family socioeconomic status (SES) increases over the educational life-cycle of individuals.[20] On the one hand, the differential patterns of school investment and family SES point to the complexity of the education production function where genetic endowments and different investments interact in distinct and time-variant ways over the life-cycle of individuals. On the other hand, the patterns suggest that investments into school environments may be able to offset the inequality-increasing interplay between genetic endowments and family investments.

**Skill formation.** In section 3, we formulated educational attainment $Y_i$ as a function of children's skills $\theta_i$ at the end of childhood. The skills that influence educational attainment are multidimensional and comprise a broad set of health indicators and (non-)cognitive skills (Almlund et al., 2011; Heckman and Mosso, 2014). Furthermore, there is evidence in the literature that each of these skill dimensions is shaped, in part, by genetic influence (Buser et al., 2021a; Demange et al., 2020, 2021).

We evaluate these potential channels by analyzing the associations of $PGS^{EA}$ and $I_{Quality}$ with a set of intermediate outcomes. In terms of health outcomes, we focus on subjective health, measured by quality-adjusted life years (QALY), and objective health, measured by an index that comprises information on whether the respondent is obese, has first-stage hypertension, or has high cholesterol. In terms of cognitive skills, we use the Picture Vocabulary Test (PVT) as a measure of verbal intelligence. Lastly, we focus on personality and preferences as two distinct conceptualizations of non-cognitive skills (Becker et al., 2012; Humphries and Kosse, 2017). In particular, we focus on risk aversion and patience, and the Big Five personality traits. All measures were collected in waves 3 and 4 of Add Health, that is, after respondents had left high school but potentially before they had completed their highest level of education (see Supplementary Material B for details).

Health, cognitive skills, risk aversion, and patience have been shown to be strong predictors of

[20] In particular we use the "social origins score" from Belsky et al. (2018) measured in wave 1. Results for alternative measures of family SES, such as family income or potential wages of either parent, are similar and available upon request.

educational attainment (Burks et al., 2015; Castillo et al., 2018a,b; Jackson, 2009). Furthermore, openness and emotional stability—the opposite of neuroticism—are positively associated with educational attainment (Becker et al., 2012; Buser et al., 2021b). Based on this evidence, we expect positive associations of both $PGS^{EA}$ and $I_{Quality}$ with each of these intermediate outcomes. The sign of the gene-environment interaction is a priori unclear. However, given the substitutability of $PGS^{EA}$ and $I_{Quality}$ in the production of educational attainment, we expect similar substitutability patterns for a subset of these intermediate outcomes as well.

Table 5 summarizes the results. In column (1)–(2) of Panel (a), we focus on health outcomes. As expected, our results show a positive association of $PGS^{EA}$ with both subjective and objective health. A 1 SD increase in $PGS^{EA}$ increases subjective (objective) health by 0.067 SD (0.042 SD). Furthermore, the negative coefficient on the interaction of $PGS^{EA}$ and $I_{Quality}$ suggests that this increase is particularly pronounced for low $PGS^{EA}$ students: a 1 SD increase in teacher quality reduces the positive association of subjective health with the $PGS^{EA}$ by $\approx 40\%$ $(= 0.027/0.067)$.

In column (3) of Panel (a), we focus on the PVT as a measure of cognitive skills. As expected, our results show positive associations of both $PGS^{EA}$ and $I_{Quality}$ with the PVT. A 1 SD increase in $PGS^{EA}$ ($I_{Quality}$) is associated with a 0.179 SD (0.104 SD) increase in the PVT. Furthermore, both factors are substitutes for each other. A 1 SD increase in teacher quality reduces the positive association of PVT and $PGS^{EA}$ by $\approx 20\%$ $(= 0.035/0.179)$.

In columns (4)–(5) of Panel (a), we focus on economic preferences. As expected, we find strong positive associations of $PGS^{EA}$ with both risk aversion and patience. A 1 SD increase in $PGS^{EA}$ is associated with a 0.038 SD (0.073 SD) increase in risk aversion (patience). Furthermore, $PGS^{EA}$ and $I_{Quality}$ are substitutes for each other. A 1 SD increase in $I_{Quality}$ reduces the positive associations of risk aversion and patience with the $PGS^{EA}$ by $\approx 121\%$ $(= 0.046/0.038)$ and $\approx 62\%$ $(= 0.045/0.073)$, respectively.

In Panel (b), we focus on personality traits. As expected, we find a positive association of $PGS^{EA}$ with openness and a negative association of $PGS^{EA}$ with neuroticism. However, $I_{Quality}$ is not predictive of any of the Big Five dimensions. Furthermore, we find no evidence of complementarity between $PGS^{EA}$ and $I_{Quality}$ in the production of personality traits.

To summarize: we find negative gene-environment interactions between genetic endowments and teacher quality in the production of subjective health, cognitive skills, risk aversion and patience. Given their predictive power for educational attainment, these intermediate outcomes are plausible channels for explaining the substitutability of genetic endowments and teacher quality in the production of educational attainment.

**TABLE 5 – Association of $PGS^{EA}$ and School Environments with Skill Measures**

| *Panel (a)* | Health | | Cognitive | Preferences | |
|---|---|---|---|---|---|
| | Subjective (1) | Objective (2) | PVT (3) | Risk (4) | Patience (5) |
| $PGS^{EA}$ | 0.067*** | 0.042** | 0.179*** | 0.038** | 0.073*** |
| | (0.017) | (0.018) | (0.017) | (0.015) | (0.017) |
| $I_{Quality}$ | 0.016 | 0.027 | 0.104*** | 0.047 | 0.046 |
| | (0.043) | (0.037) | (0.039) | (0.030) | (0.038) |
| $PGS^{EA} \times I_{Quality}$ | -0.027** | -0.000 | -0.035* | -0.046*** | -0.045*** |
| | (0.014) | (0.020) | (0.018) | (0.015) | (0.013) |
| Child Controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 3,075 | 3,075 | 2,995 | 3,071 | 3,071 |
| $R^2$ | 0.080 | 0.057 | 0.222 | 0.114 | 0.098 |

| *Panel (b)* | Personality | | | | |
|---|---|---|---|---|---|
| | Open-ness (1) | Conscient-iousness (2) | Extra-version (3) | Agree-ableness (4) | Neuro-ticism (5) |
| $PGS^{EA}$ | 0.071*** | -0.018 | -0.006 | 0.037* | -0.084*** |
| | (0.017) | (0.017) | (0.019) | (0.020) | (0.019) |
| $I_{Quality}$ | 0.030 | -0.033 | -0.044 | 0.055 | -0.020 |
| | (0.031) | (0.036) | (0.030) | (0.037) | (0.035) |
| $PGS^{EA} \times I_{Quality}$ | 0.010 | -0.007 | -0.001 | -0.008 | 0.022 |
| | (0.013) | (0.015) | (0.023) | (0.019) | (0.019) |
| Child Controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 3,053 | 3,073 | 3,069 | 3,071 | 3,071 |
| $R^2$ | 0.092 | 0.042 | 0.031 | 0.135 | 0.094 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with health, cognitive skills, preferences, and personality. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

## 6 CONCLUSION

The question of how natural endowments and environmental factors determine life outcomes has a long history of inquiry in philosophy and science (Darwin, 1859; Descartes, 1641; Lamarck, 1838; Locke, 1690). The assumption that life outcomes are the result of genetic and environ-

mental factors initially led to the so-called "nature versus nurture" debate. However, current research has moved beyond this simplistic dichotomy and recognizes that individual life outcomes are the result of a complex interplay between nature and nurture. This insight highlights that the importance of genetic endowments for life outcomes is not immutable. Instead, it opens a path for policy interventions that shape the relevant environment.

In this paper, we contribute to this research agenda by studying the interplay of genetic endowments and schooling environments in the production of educational outcomes. Making use of recent advances in molecular genetics, we link an individual-level index of genetic predispositions for educational success with measures for teacher "quality" and "quantity" during high school. In turn, we investigate whether the importance of genetic endowments varies with the quality of the high school environment.

Our findings suggest that school investments have the potential to mitigate the genetic gradient in educational attainment. However, this conclusion depends on the particular type of investment. On the one hand, increases in "teacher quality" offset genetic disadvantages. On the other hand, we find no substitutability with respect to "teacher quantity." Our findings furthermore suggest that higher gains in educational attainment for students with lower genetic endowments are mediated by gains in subjective health, language skills, risk aversion, and patience.

The use of genetics in education research has an ugly history. Therefore, many people are wary of the emergence of genetic markers in this context, especially when these markers are used for genetic screening (Martschenko et al., 2019). We emphasize that our results do neither presuppose nor endorse the use of genetic screening for educational interventions. Instead, our results suggest that universal policy reform that increases the quality of teachers for *all students* may provide an important step to level the playing field regardless of a student's draw in the genetic lottery.

## References


AARONSON, D., L. BARROW, and W. SANDER (2007). "Teachers and Student Achievement in the Chicago Public High Schools". *Journal of Labor Economics* 25 (1), pp. 95–135.

ABDELLAOUI, A., D. HUGH-JONES, L. YENGO, K. E. KEMPER, M. G. NIVARD, L. VEUL, Y. HOLTZ, B. P. ZIETSCH, T. M. FRAYLING, N. R. WRAY, et al. (2019). "Genetic Correlates of Social Stratification in Great Britain". *Nature Human Behaviour* 3 (12), pp. 1332–1342.

ACEMOGLU, D. and D. AUTOR (2011). "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings". Ed. by D. CARD and O. ASHENFELTER. Vol. 4. Handbook of Labor Economics. Elsevier, pp. 1043–1171.

AGOSTINELLI, F. and G. SORRENTI (2018). "Money vs. Time: Family Income, Maternal Labor Supply, and Child Development". *HCEO Working Paper Series* 2018-017.

ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, and T. KAUTZ (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education*. Ed. by E. A. HANUSHEK, S. MACHIN, and L. WOESSMANN. Vol. 4. Elsevier, pp. 1–181.

ALMOND, D. and B. MAZUMDER (2013). "Fetal Origins and Parental Responses". *Annual Review of Economics* 5 (1), pp. 37–56.

ALTONJI, J. G., T. E. ELDER, and C. R. TABER (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools". *Journal of Political Economy* 113 (1), pp. 151–184.

ALTONJI, J. G. and R. K. MANSFIELD (2018). "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects". *American Economic Review* 108 (10), pp. 2902–46.

ANDERSON, M. L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American Statistical Association* 103 (484), pp. 1481–1495.

ANGRIST, J. D., S. R. COHODES, S. M. DYNARSKI, P. A. PATHAK, and C. R. WALTERS (2016). "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice". *Journal of Labor Economics* 34 (2), pp. 275–318.

ANGRIST, J. D. and V. LAVY (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement". *Quarterly Journal of Economics* 114 (2), pp. 533–575.

ANGRIST, J. D., V. LAVY, J. LEDER-LUIS, and A. SHANY (2019). "Maimonides' Rule Redux". *American Economic Review: Insights* 1 (3), pp. 309–324.

ANGRIST, J. D., P. A. PATHAK, and C. R. WALTERS (2013). "Explaining Charter School Effectiveness". *American Economic Journal: Applied Economics* 5 (4), pp. 1–27.

ANGRIST, J. D. and J.-S. PISCHKE (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

BACHER-HICKS, A., S. B. BILLINGS, and D. J. DEMING (2019). "The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime". *National Bureau of Economic Research Working Paper Series* 26257.

BARCELLOS, S. H., L. S. CARVALHO, and P. TURLEY (2021). "The Effect of Education on the Relationship between Genetics, Early-life Disadvantages,and Later-life SES". *National Bureau of Economic Research Working Paper Series* 28750.

BARTH, D., N. W. PAPAGEORGE, and K. THOM (2020). "Genetic Endowments and Wealth Inequality". *Journal of Political Economy* 128 (4), pp. 1474–1522.

BECKER, A., T. DECKERS, T. DOHMEN, A. FALK, and F. KOSSE (2012). "The Relationship Between Economic Preferences and Psychological Personality Measures". *Annual Review of Economics* 4 (1), pp. 453–478.

BELSKY, D. W. et al. (2016). "The Genetics of Success: How Single-nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-course Development". *Psychological Science* 27 (7), pp. 957–972.

BELSKY, D. W. et al. (2018). "Genetic Analysis of Social-class Mobility in Five Longitudinal Studies". *Proceedings of the National Academy of Sciences* 115 (31), E7275–E7284.

BEUERMANN, D., C. K. JACKSON, L. NAVARRO-SOLA, and F. PARDO (forthcoming). "What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output". *Review of Economic Studies*.

BIASI, B. (forthcoming). "School Finance Equalization Increases Intergenerational Mobility". *Journal of Labor Economics*.

BIETENBECK, J. (2019). "The Long-term Impacts of Low-achieving Childhood Peers: Evidence from Project STAR". *Journal of the European Economic Association* 18 (1), pp. 392–426.

BIROLI, P., T. J. GALAMA, S. von HINKE, H. van KIPPERSLUIS, C. A. RIETVELD, and K. THOM (2022). "The Economics and Econometrics of Gene-Environment Interplay". *arXiv* 2203.00729.

BORGHANS, L., B. H. H. GOLSTEYN, J. J. HECKMAN, and J. E. HUMPHRIES (2016). "What Grades and Achievement Tests Measure". *Proceedings of the National Academy of Sciences* 113 (47), pp. 13354–13359.

BRANIGAN, A. R., K. J. MCCALLUM, and J. FREESE (2013). "Variation in the Heritability of Educational Attainment: An International Meta-analysis". *Social Forces* 92 (1), pp. 109–140.

BRINCH, C. N. and T. A. GALLOWAY (2012). "Schooling in Adolescence Raises IQ Scores". *Proceedings of the National Academy of Sciences* 109 (2), pp. 425–430.

BURKS, S. V., C. LEWIS, P. A. KIVI, A. WIENER, J. E. ANDERSON, L. GÖTTE, C. G. DEYOUNG, and A. RUSTICHINI (2015). "Cognitive skills, personality, and economic preferences in collegiate success". *Journal of Economic Behavior & Organization* 115, pp. 30–44.

BUSER, T., R. AHLSKOG, M. JOHANNESSON, P. KOELLINGER, and S. OSKARSSON (2021a). "Using genes to explore the effects of cognitive andnon-cognitive skills on education and labor market outcomes". *Tinbergen Institute Discussion Paper* 2021-088.

BUSER, T., M. NIEDERLE, and H. OOSTERBEEK (2021b). "Can Competitiveness predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures". *National Bureau of Economic Research Working Paper Series* 28916.

CASTILLO, M., J. L. JORDAN, and R. PETRIE (2018a). "Discount Rates of Children and High School Graduation". *Economic Journal* 129 (619), pp. 1153–1181.

CASTILLO, M., J. L. JORDAN, and R. PETRIE (2018b). "Children's Rationality, Risk Attitudes and Field Behavior". *European Economic Review* 102, pp. 62–81.

CESARINI, D. and P. M. VISSCHER (2017). "Genetics and Educational Attainment". *npj Science of Learning* 2 (4).

CHANOCK, S. J., T. MANOLIO, M. BOEHNKE, E. BOERWINKLE, D. J. HUNTER, G. THOMAS, J. N. HIRSCHHORN, G. R. ABECASIS, D. ALTSHULER, J. E. BAILEY-WILSON, et al. (2007). "Replicating Genotype-phenotype Associations". *Nature* (447), pp. 655–660.

CHETTY, R., J. N. FRIEDMAN, N. HILGER, E. SAEZ, D. W. SCHANZENBACH, and D. YAGAN (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star". *Quarterly Journal of Economics* 126 (4), pp. 1593–1660.

CHETTY, R., J. N. FRIEDMAN, and J. E. ROCKOFF (2014a). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates". *American Economic Review* 104 (9), pp. 2593–2632.

– (2014b). "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood". *American Economic Review* 104 (9), pp. 2633–79.

CINELLI, C. and C. HAZLETT (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82 (1), pp. 39–67.

CLOTFELTER, C. T., H. F. LADD, and J. L. VIGDOR (2010). "Teacher Credentials and Student Achievement in High School: A Cross-subject Analysis with Student Fixed Effects". *Journal of Human Resources* 45 (3), pp. 655–681.

CULLEN, J. B., B. A. JACOB, and S. LEVITT (2006). "The Effect of School Choice on Participants: Evidence from Randomized Lotteries". *Econometrica* 74 (5), pp. 1191–1230.

CUNHA, F., J. J. HECKMAN, L. LOCHNER, and D. V. MASTEROV (2006). "Interpreting the Evidence on Life Cycle Skill Formation". Ed. by E. HANUSHEK and F. WELCH. Vol. 1. Handbook of the Economics of Education. Amsterdam: Elsevier. Chap. 12, pp. 697–812.

CUNHA, F., J. J. HECKMAN, and S. M. SCHENNACH (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". *Econometrica* 78 (3), pp. 883–931.

DARWIN, C. R. (1859). *On the Origin of Species*. London: John Murray.

DEMANGE, P. A. et al. (2020). "Parental influences on offspring education: indirect genetic effects of non-cognitive skills". *bioRxiv* 2020.09.15.296236.

DEMANGE, P. A. et al. (2021). "Investigating the Genetic Architecture of Non-Cognitive Skills Using GWAS-by-Subtraction". *Nature Genetics* 53, pp. 35–44.

DEMONTIS, D., R. K. WALTERS, J. MARTIN, M. MATTHEISEN, T. D. ALS, E. AGERBO, G. BALDURSSON, R. BELLIVEAU, J. BYBJERG-GRAUHOLM, M. BÆKVAD-HANSEN, et al. (2019). "Discovery of the First Genome-wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder". *Nature Genetics* 51 (1), pp. 63–75.

DESCARTES, R. (1641). "Meditations on First Philosophy". *Descartes: Selected Philosophical Writings*. Ed. by J. COTTINGHAM. Ed. by D. MURDOCH. Ed. by R. STOOTHOFF. Ed. by A. KENNY. Cambridge: Cambridge University Press, pp. 73–122.

DOMINGUE, B. W., D. BELSKY, D. CONLEY, K. M. HARRIS, and J. BOARDMAN (2015). "Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health". *AERA Open* 1 (3), pp. 1–13.

DOMINGUE, B. W., S. TREJO, E. ARMSTRONG-CARTER, and E. M. TUCKER-DROB (2020). "Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges". *Sociological Science* 7 (19), pp. 465–486.

DUDBRIDGE, F. (2013). "Power and Predictive Accuracy of Polygenic Risk Scores". *PLOS Genetics* 9 (3), e1003348.

FALK, A., A. BECKER, T. J. DOHMEN, B. ENKE, D. B. HUFFMAN, and U. SUNDE (2018). "Global Evidence on Economic Preferences". *Quarterly Journal of Economics* 133 (4), pp. 1645–1692.

FLOOD, S., M. KING, R. RODGERS, S. RUGGLES, and J. R. WARREN (2020). "Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]". *Minneapolis, MN: IPUMS*.

FREDRIKSSON, P., B. ÖCKERT, and H. OOSTERBEEK (2013). "Long-term Effects of Class Size". *Quarterly Journal of Economics* 128 (1), pp. 249–285.

HANUSHEK, E. A. and S. G. RIVKIN (2010). "Generalizations about Using Value-added Measures of Teacher Quality". *American Economic Review* 100 (2), pp. 267–271.

HANUSHEK, E. A., S. G. RIVKIN, and J. C. SCHIMAN (2016). "Dynamic Effects of Teacher Turnover on the Quality of Instruction". *Economics of Education Review* 55, pp. 132–148.

HANUSHEK, E. A. and L. WOESSMANN (2008). "The Role of Cognitive Skills in Economic Development". *Journal of Economic Literature* 46 (3), pp. 607–668.

– (2012). "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation". *Journal of Economic Growth* 17 (4), pp. 267–321.

HECKMAN, J., S. H. MOON, R. PINTO, P. SAVELYEV, and A. YAVITZ (2010). "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program". *Quantitative Economics* 1 (1), pp. 1–46.

HECKMAN, J. and S. MOSSO (2014). "The Economics of Human Development and Social Mobility". *Annual Review of Economics* 6 (1), pp. 689–733.

HILL, A. J. and D. B. JONES (2018). "A Teacher Who Knows Me: The Academic Benefits of Repeat Student-teacher Matches". *Economics of Education Review* 64, pp. 1–12.

HOUMARK, M. A., V. RONDA, and M. ROSHOLM (2020). "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills". *IZA Discussion Paper Series* 13780.

HOWARD, D. M., M. J. ADAMS, T.-K. CLARKE, J. D. HAFFERTY, J. GIBSON, M. SHIRALI, J. R. COLEMAN, S. P. HAGENAARS, J. WARD, E. M. WIGMORE, et al. (2019). "Genome-wide Meta-analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions". *Nature Neuroscience* 22 (3), pp. 343–352.

HUMPHRIES, J. E. and F. KOSSE (2017). "On the Interpretation of Non-cognitive Skills—What is Being Measured and Why it Matters". *Journal of Economic Behavior & Organization* 136, pp. 174–185.

HWANG, N., B. KISIDA, and C. KOEDEL (2021). "A Familiar Face: Student-teacher Rematches and Student Achievement". *Economics of Education Review* 85, p. 102194.

ISPHORDING, I. E. and U. ZÖLITZ (2020). "The Value of a Peer". *University of Zurich, Department of Economics, Working Paper No. 342*.

JACKSON, C. K. (2019). "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes". *Journal of Political Economy* 126 (5), pp. 2072–2107.

JACKSON, C. K. and E. BRUEGMANN (Oct. 2009). "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers". *American Economic Journal: Applied Economics* 1 (4), pp. 85–108.

JACKSON, C. K., S. C. PORTER, J. Q. EASTON, A. BLANCHARD, and S. KIGUEL (2020). "School Effects on Socioemotional Development, School-based Arrests, and Educational Attainment". *American Economic Review: Insights* 2 (4), pp. 491–508.

JACKSON, M. I. (2009). "Understanding links between adolescent health and educational attainment". *Demography* 46 (4), pp. 671–694.

JACOB, B. A., J. E. ROCKOFF, E. S. TAYLOR, B. LINDY, and R. ROSEN (2018). "Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools". *Journal of Public Economics* 166, pp. 81–97.

JANSEN, P., K. WATANABE, S. STRINGER, N. SKENE, J. BRYOIS, A. HAMMERSCHLAG, C. DE LEEUW, J. BENJAMINS, A. MUÑOZ-MANCHADO, M. NAGEL, et al. (2019). "Genome-wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways". *Nature Genetics* 51 (3), pp. 394–403.

JOHNSON, R. C. and C. K. JACKSON (2019). "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending". *American Economic Journal: Economic Policy* 11 (4), pp. 310–49.

KELLER, M. (2014). "Gene × Environment Interaction Studies have not Properly Controlled for Potential Confounders: the Problem and the (Simple) Solution". *Biological Psychiatry* 75 (1), pp. 18–24.

KIRKEBØEN, L. (2021). "School Value-added and Long-term Student Outcomes". *Statistics Norway Discussion Papers* 970.

KLING, J. R., J. B. LIEBMAN, and L. F. KATZ (2007). "Experimental Analysis of Neighborhood Effects". *Econometrica* 75 (1), pp. 83–119.

KOELLINGER, P. D. and K. P. HARDEN (2018). "Using Nature to Understand Nurture". *Science* 359 (6374), pp. 386–387.

KONG, A. et al. (2018). "The Nature of Nurture: Effects of Parental Genotypes". *Science* 359 (6374), pp. 424–428.

KOSSE, F. (2016). "The Nutritional and Social Environment-Related Effects of Breastfeeding on Intelligence". *JAMA Pediatrics* 170 (2), pp. 173–174.

KRUEGER, A. B. (1999). "Experimental Estimates of Education Production Functions". *Quarterly Journal of Economics* 114 (2), pp. 497–532.

KRUEGER, A. B. and M. LINDAHL (2001). "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39 (4), pp. 1101–1136.

KWEON, H., C. A. BURIK, R. K. LINNÉR, R. de VLAMING, A. OKBAY, D. MARTSCHENKO, K. P. HARDEN, T. A. DIPRETE, and P. D. KOELLINGER (2020). "Genetic Fortune: Winning or Losing Education, Income, and Health". *Tinbergen Institute Discussion Papers* 20-053.

LALIBERTÉ, J.-W. (2021). "Long-term Contextual Effects in Education: Schools and Neighborhoods". *American Economic Journal: Economic Policy* 13 (2), pp. 336–377.

LAMARCK, J.-B. P. A. d. M. de (1838). *Histoire naturelle des animaux sans vertèbres*. Paris: Verdière.

LEE, J. J. et al. (2018). "Gene Discovery and Polygenic Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals". *Nature Genetics* 50 (8), pp. 1112–1121.

LEUVEN, E. and S. A. LØKKEN (2020). "Long-term Impacts of Class Size in Compulsory School". *Journal of Human Resources* 55 (1), pp. 309–348.

LIU, M., Y. JIANG, R. WEDOW, Y. LI, D. BRAZEL, F. CHEN, G. DATTA, J. DAVILA-VELDERRAIN, D. MCGUIRE, C. TIAN, et al. (2019). "Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use". *Nature Genetics* 51 (2), pp. 237–244.

LOCKE, J. (1690). *An Essay Concerning Human Understanding*. Hamilton: McMaster University Archive for the History of Economic Thought.

MARTIN, A. R., C. R. GIGNOUX, R. K. WALTERS, G. L. WOJCIK, B. M. NEALE, S. GRAVEL, M. J. DALY, C. D. BUSTAMANTE, and E. E. KENNY (2017). "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations". *American Journal of Human Genetics* 100 (4), pp. 635–649.

MARTSCHENKO, D., S. TREJO, and B. W. DOMINGUE (2019). "Genetics and Education: Recent Developments in the Context of an Ugly History and an Uncertain Future." *AERA Open* 5 (1), pp. 1–15.

MILLS, M. C., N. BARBAN, and F. C. TROPF (2020). *An Introduction to Statistical Genetic Data Analysis*. Cambridge: MIT Press.

MUSLIMOVA, D., H. van KIPPERSLUIS, C. A. RIETVELD, S. von HINKE, and S. F. W. MEDDENS (2020). "Dynamic Complementarity in Skill Production: Evidence from Genetic Endowments and Birth Order". *Tinbergen Institute Discussion Paper 2020-082/V*.

NICOLETTI, C. and B. RABE (2014). "School Inputs and Skills: Complementarity and Self-productivity". *IZA Working Paper Series* 8693.

OSTER, E. (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". *Journal of Business & Economic Statistics* 37 (2), pp. 187–204.

PAPAGEORGE, N. W. and K. THOM (2020). "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study". *Journal of the European Economic Association* 18 (3), pp. 1351–1399.

PAPAY, J. P. and M. A. KRAFT (2015). "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement". *Journal of Public Economics* 130, pp. 105–119.

RIVKIN, S. G., E. A. HANUSHEK, and J. F. KAIN (2005). "Teachers, Schools, and Academic Achievement". *Econometrica* 73 (2), pp. 417–458.

ROCKOFF, J. E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data". *American Economic Review* 94 (2), pp. 247–252.

RONDA, V., E. AGERBO, D. BLESES, P. B. MORTENSEN, A. BØRGLUM, D. M. HOUGAARD, O. MORS, M. NORDENTOFT, T. WERGE, and M. ROSHOLM (forthcoming). "Family Disadvantage, Gender and the Returns to Genetic Human Capital". *Scandinavian Journal of Economics*.

RONFELDT, M., S. LOEB, and J. WYCKOFF (2013). "How Teacher Turnover Harms Student Achievement". *American Educational Research Journal* 50 (1), pp. 4–36.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, and M. SOBEK (2020). "Integrated Public Use Microdata Series USA: Version 10.0 [dataset]". *Minneapolis, MN: IPUMS*.

SACERDOTE, B. (2014). "Experimental and Quasi-experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6 (1), pp. 253–272.

SAVAGE, J. E. et al. (2018). "Genome-wide Association Meta-analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence". *Nature Genetics* 50 (7), pp. 912–919.

SELZAM, S., S. J. RITCHIE, J.-B. PINGAULT, C. A. REYNOLDS, P. F. O'REILLY, and R. PLOMIN (2019). "Comparing Within- and Between-Family Polygenic Score Prediction". *American Journal of Human Genetics* 105 (2), pp. 351–363.

SHENHAV, N. (2021). "Lowering Standards to Wed? Spouse Quality, Marriage, and Labor Market Responses to the Gender Wage Gap". *Review of Economics and Statistics* 103 (2), pp. 265–279.

SOTOUDEH, R., K. M. HARRIS, and D. CONLEY (2019). "Effects of the Peer Metagenomic Environment on Smoking Behavior". *Proceedings of the National Academy of Sciences* 116 (33), pp. 16302–16307.

TREJO, S., D. BELSKY, J. BOARDMAN, J. FREESE, K. M. HARRIS, P. HERD, K. SICINSKI, and B. DOMINGUE (2018). "Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health". *Sociological Science* 5, pp. 513–540.

TURKHEIMER, E. (2000). "Three Laws of Behavior Genetics and What They Mean". *Current Directions in Psychological Science* 9 (5), pp. 160–164.

WARE, E. B., L. L. SCHMITZ, J. FAUL, A. GARD, C. MITCHELL, J. A. SMITH, W. ZHAO, D. WEIR, and S. L. KARDIA (2017). "Heterogeneity in Polygenic Scores for Common Human Traits". *BioRxiv 106062*.

WRAY, N. R., S. H. LEE, D. MEHTA, A. A. VINKHUYZEN, F. DUDBRIDGE, and C. M. MIDDELDORP (2014). "Research Review: Polygenic Methods and their Application to Psychiatric Traits". *Journal of Child Psychology and Psychiatry* 55 (10), pp. 1068–1087.

YENGO, L., J. SIDORENKO, K. E. KEMPER, Z. ZHENG, A. R. WOOD, M. N. WEEDON, T. M. FRAYLING, J. HIRSCHHORN, J. YANG, P. M. VISSCHER, et al. (2018). "Meta-analysis of Genome-wide Association Studies for Height and Body Mass Index in 700000 Individuals of European Ancestry". *Human Molecular Genetics* 27 (20), pp. 3641–3649.

# A SUPPLEMENTARY TABLES

**TABLE A.1 – Sample Representativeness**

| | Population (Cohorts 1974-1983) | | Analysis Sample | |
|---|---|---|---|---|
| | All | Non-Hispanic White | Unweighted | Re-Weighted |
| *Gender* | | | | |
| Male | 0.498 | 0.503 | 0.453 | 0.503 |
| Female | 0.502 | 0.497 | 0.547 | 0.497 |
| *Education Mother* | | | | |
| $\leq$ High School | 0.536 | 0.489 | 0.493 | 0.489 |
| $>$ High School; $<$ College Degree | 0.281 | 0.302 | 0.218 | 0.301 |
| $\geq$ College Degree | 0.183 | 0.209 | 0.289 | 0.210 |
| *Education Father* | | | | |
| $\leq$ High School | 0.472 | 0.425 | 0.491 | 0.425 |
| $>$ High School; $<$ College Degree | 0.255 | 0.271 | 0.196 | 0.271 |
| $\geq$ College Degree | 0.273 | 0.304 | 0.313 | 0.303 |
| *Age Mother at Birth* | | | | |
| $<$ 25 Years | 0.353 | 0.330 | 0.486 | 0.330 |
| $\geq$ 25 Years | 0.647 | 0.670 | 0.514 | 0.670 |
| *Parental Income* | | | | |
| $<$ \$50,000 | 0.557 | 0.491 | 0.530 | 0.515 |
| $\geq$ \$50,000; $<$ \$100,000 | 0.352 | 0.403 | 0.390 | 0.402 |
| $\geq$ \$100,000 | 0.091 | 0.106 | 0.080 | 0.083 |
| *Education Respondent* | | | | |
| $\leq$ High School | 0.301 | 0.225 | 0.181 | 0.173 |
| $>$ High School; $<$ College Degree | 0.327 | 0.344 | 0.399 | 0.403 |
| $\geq$ College Degree | 0.372 | 0.431 | 0.419 | 0.424 |

**Data:** National Longitudinal Study of Adolescent to Adult Health, American Community Survey (ACS), Current Population Survey (CPS).
**Note:** Own calculations. This table shows summary statistics of the core analysis sample in comparison to other population samples. It shows respondents' characteristics for the following samples: (i) the US population from birth cohorts 1974–1983, (ii) the Non-Hispanic White US population from birth cohorts 1974–1983, (iii) the core estimation sample, and (iv) the core estimation sample re-weighted to match (ii) with respect to *Gender*, *Education Mother*, *Education Father*, and *Age Mother at Birth*. Population data on *Gender* and *Education Respondent* from IPUMS ACS 2019 (Ruggles et al., 2020). Population data on *Education Mother*, *Education Father*, *Age Mother at Birth*, and *Parental Income* from IPUMS CPS 1994 (Flood et al., 2020).

**TABLE A.2 – Testing for Genetic Nurture and Selection into Schools**

| | Baseline Sample | Between Family (2) vs. Within-Family (3) | | Years of Educ. (4) vs. Predicted Years of Educ. (5) | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $PGS^{EA}$ | 0.368*** (0.037) | 0.384*** (0.109) | 0.394*** (0.148) | – | – |
| $I_{Quality}$ | 0.217*** (0.079) | – | – | 0.291*** (0.096) | 0.076 (0.052) |
| $I_{Quantity}$ | 0.078 (0.051) | – | – | 0.035 (0.058) | 0.067 (0.072) |
| Child Controls | ✓ | ✓ | × | ✓ | ✓ |
| Family Controls | ✓ | ✓ | × | ✓ | ✓ |
| Sibling Fixed Effect | × | × | ✓ | × | × |
| N | 3,075 | 525 | 525 | 2,109 | 2,109 |
| $R^2$ | 0.337 | 0.384 | 0.747 | 0.322 | 0.330 |
| Outcome Mean | 14.811 | 14.928 | 14.928 | 14.968 | 14.968 |
| Outcome SD | 2.249 | 2.262 | 2.262 | 2.247 | 0.955 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.

**Note:** Own calculations. This table shows the associations of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with years of education. The first panel shows estimates in our baseline sample. The second panel shows estimates in the sibling sample. Column (2) displays results from a between-family comparison. Column (3) displays results from a within-family comparison. The third panel shows estimates in a sample with information on breastfeeding, parental time investments, and parental income. Column (4) displays results for completed years of education. Column (5) displays results for predicted years of education. Predicted education is calculated from a fully interacted regression of completed years of education on breastfeeding, maternal time investments, paternal time investments, and log family income. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0$, $\sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

**TABLE A.3 – Tests for Equality of $PGS^{EA}$ Distributions**

| | Terciles of $I_{Quality}$ / $I_{Quantity}$ | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| *Panel (a): $I_{Quality}$* | | | |
| 1 | – | – | – |
| 2 | 0.70 | – | – |
| 3 | 0.08 | 0.45 | – |
| *Panel (b): $I_{Quantity}$* | | | |
| 1 | – | – | – |
| 2 | 0.63 | – | – |
| 3 | 0.99 | 0.84 | – |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows the results of pairwise Kolmogorov-Smirnov tests for the $PGS^{EA}$ distributions within different terciles of $I_{Quality}$ and $I_{Quantity}$, respectively. Results are summarized by the p-value for the null hypothesis that the two $PGS^{EA}$ distributions are equal within the corresponding terciles of $I_{Quality}$ and $I_{Quantity}$.

## TABLE A.4 – Alternative Polygenic Scores

| | Baseline | + Controls for Other Polygenic Scores | | | | | |
|---|---|---|---|---|---|---|---|
| Outcome: Years of Education | (1) | Body Mass Index (2) | ADHD (3) | Depressive Symptoms (4) | Intelligence (5) | Ever Smoker (6) | Sleep Duration (7) |
| $PGS^{EA}$ | 0.365*** | 0.351*** | 0.340*** | 0.366*** | 0.352*** | 0.341*** | 0.369*** |
| | (0.033) | (0.035) | (0.032) | (0.034) | (0.040) | (0.038) | (0.033) |
| $I_{Quality}$ | 0.214*** | 0.218*** | 0.216*** | 0.218*** | 0.219*** | 0.217*** | 0.219*** |
| | (0.079) | (0.079) | (0.078) | (0.078) | (0.077) | (0.076) | (0.078) |
| $PGS^{EA} \times I_{Quality}$ | -0.074** | -0.083** | -0.075** | -0.074** | -0.080** | -0.080** | -0.074** |
| | (0.033) | (0.035) | (0.033) | (0.034) | (0.037) | (0.037) | (0.033) |
| Other PGS | – | -0.072** | -0.132*** | -0.026 | 0.030 | -0.134*** | -0.005 |
| | | (0.032) | (0.034) | (0.032) | (0.040) | (0.041) | (0.032) |
| Other PGS $\times$ $I_{Quality}$ | – | -0.029 | 0.020 | 0.011 | 0.009 | -0.012 | 0.000 |
| | | (0.037) | (0.030) | (0.030) | (0.040) | (0.039) | (0.034) |
| Child Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 |
| $R^2$ | 0.339 | 0.339 | 0.341 | 0.338 | 0.338 | 0.341 | 0.338 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.

**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. We control for other PGS and their interaction with $I_{Quality}$ and $I_{Quantity}$ . The relevant PGS are indicated in the column header. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

**TABLE A.5 – Decomposition of $I_{Quality}$ and $I_{Quantity}$**

| | Baseline | Decomposition of $I_{Quality}$ and $I_{Quantity}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | Average Class Size (2) | STR (3) | Tenure $>$ 5 years (4) | Tenure $<$ 1 year (5) | MA Degree (6) | HI Race (7) | HI Hisp. (8) |
| $PGS^{EA}$ | 0.365*** (0.033) | 0.372*** (0.037) | 0.372*** (0.037) | 0.368*** (0.035) | 0.372*** (0.036) | 0.370*** (0.034) | 0.374*** (0.036) | 0.373*** (0.036) |
| $I_{Quality}$ | 0.214*** (0.079) | – | – | – | – | – | – | – |
| $PGS^{EA} \times I_{Quality}$ | -0.074** (0.033) | – | – | – | – | – | – | – |
| $I_{Quantity}$ | 0.078 (0.053) | – | – | – | – | – | – | – |
| $PGS^{EA} \times I_{Quantity}$ | 0.027 (0.033) | – | – | – | – | – | – | – |
| School Characteristic | – | -0.073 (0.058) | -0.000 (0.085) | 0.191*** (0.066) | 0.094 (0.066) | 0.182*** (0.067) | -0.017 (0.068) | -0.121*** (0.037) |
| $PGS^{EA} \times$ School Characteristic | – | -0.029 (0.036) | -0.047 (0.035) | -0.062** (0.030) | 0.041 (0.039) | -0.054 (0.033) | -0.048 (0.034) | -0.013 (0.026) |
| Child Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 | 3,075 |
| $R^2$ | 0.339 | 0.335 | 0.335 | 0.338 | 0.336 | 0.337 | 0.335 | 0.336 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.

**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$and the school characteristics used for the construction of $I_{Quality}$ and $I_{Quantity}$ with completed years of education. The relevant school characteristics are indicated in the column header. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

## TABLE A.6 – Robustness to Sample Composition

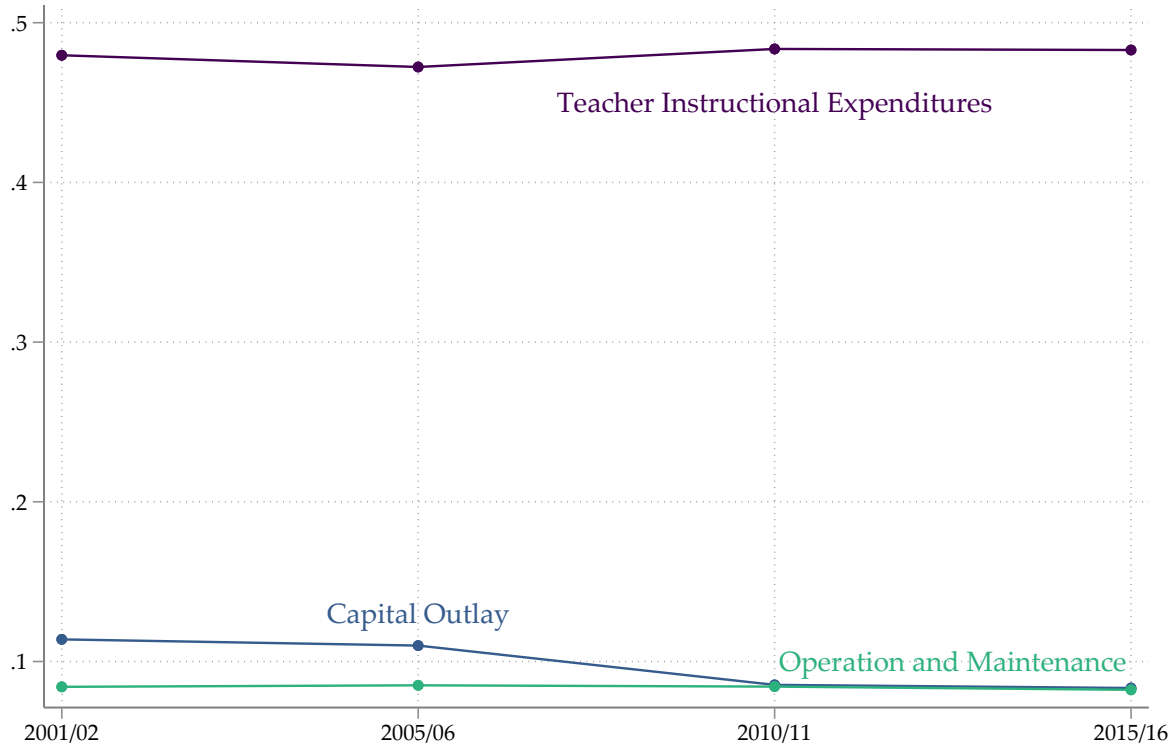
| | Baseline | Alternative Sample Composition | | |
|---|---|---|---|---|
| Outcome: Years of Education | (1) | Re-Weighted (2) | Excl. all (Potential) Movers (3) | Inc. all (Potential) Movers (4) |
| $PGS^{EA}$ | 0.365*** | 0.353*** | 0.355*** | 0.372*** |
| | (0.033) | (0.034) | (0.039) | (0.028) |
| $I_{Quality}$ | 0.214*** | 0.186** | 0.140 | 0.207*** |
| | (0.079) | (0.080) | (0.094) | (0.074) |
| $PGS^{EA} \times I_{Quality}$ | -0.074** | -0.076** | -0.070* | -0.047* |
| | (0.033) | (0.033) | (0.038) | (0.028) |
| $I_{Quantity}$ | 0.078 | 0.060 | 0.084 | 0.065 |
| | (0.053) | (0.059) | (0.065) | (0.070) |
| $PGS^{EA} \times I_{Quantity}$ | 0.027 | 0.039 | 0.012 | 0.038 |
| | (0.033) | (0.035) | (0.036) | (0.024) |
| Child Controls | ✓ | ✓ | ✓ | ✓ |
| Family Controls | ✓ | ✓ | ✓ | ✓ |
| N | 3,075 | 3,021 | 2,520 | 4,176 |
| $R^2$ | 0.339 | 0.317 | 0.330 | 0.324 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.

**Note:** Own calculations. This table shows the joint association of $PGS^{EA}$, $I_{Quality}$ and $I_{Quantity}$ with completed years of education. In column (2) we re-weight our analysis sample to match ACS and CPS with respect to gender composition, educational attainment of parents, and the age of mothers at birth—see also Appendix Table A.1. In column (3) we exclude respondents that visit feeder schools in wave 1 and for whom we do not have information on subsequent high schools. In column (4) we include respondents that are in Add Health high schools in wave 1 and for whom we do not have information on subsequent high schools. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. All right-hand side variables are standardized on the estimation sample so that $\mu = 0, \sigma = 1$. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (in parentheses) are clustered at the school level.

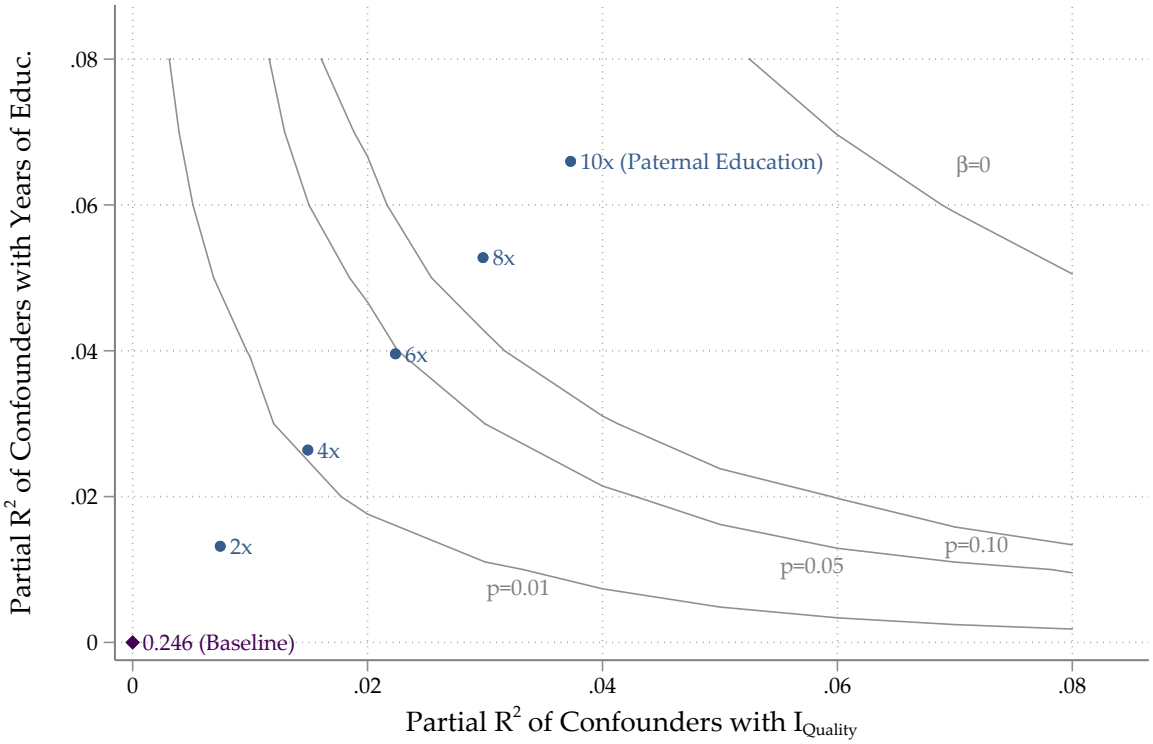
# B SUPPLEMENTARY FIGURES

**FIGURE B.1 – Top 3 School Expenditure Categories (in % of Total)**



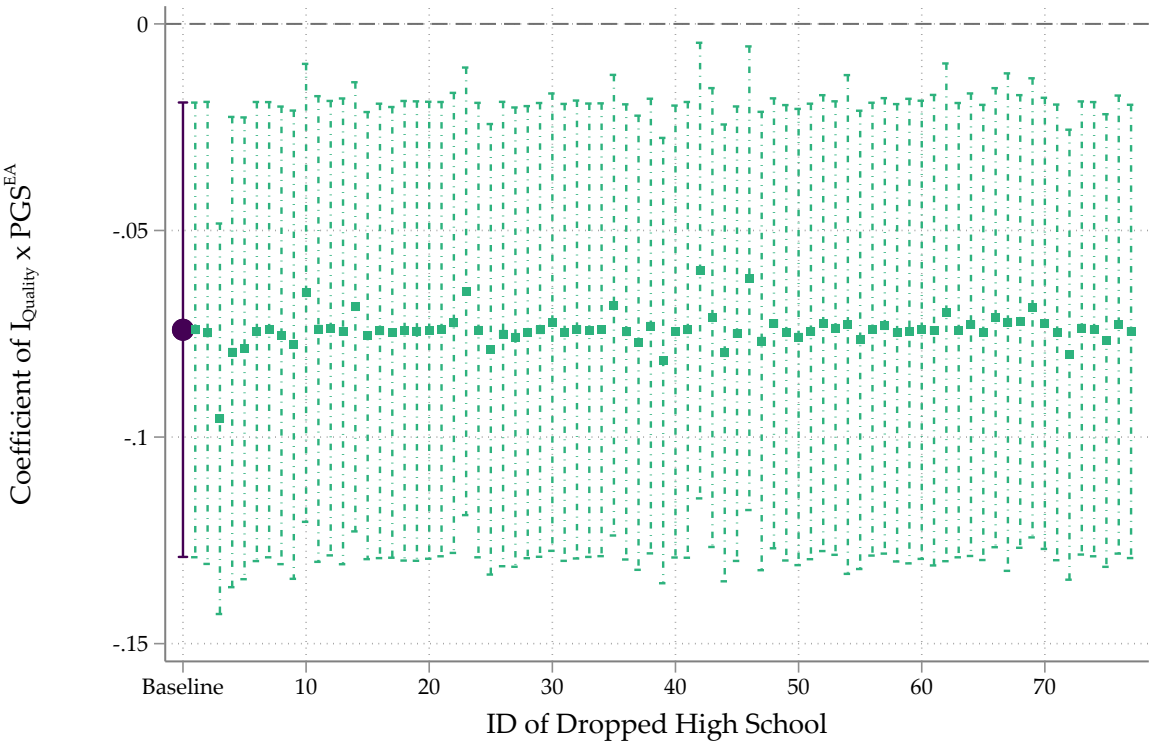

**Data:** Common Core of Data (CCD), National Public Education Financial Survey.
**Note:** Own calculations. This figure shows per-pupil expenditures shares in public elementary and secondary schools in the US. Teacher Instructional Expenditures includes teachers' salaries and employee benefits. Capital Outlay includes expenditures for property and for buildings and alterations completed by school district staff or contractors. Operation and Maintenance includes expenditures for the supervision of operations and maintenance, the operation of buildings, the care and upkeep of grounds and equipment, vehicle operations (other than student transportation) and maintenance, and security.

**FIGURE B.2 – Sensitivity of $I_{Quality}$ to Unobserved Confounders**



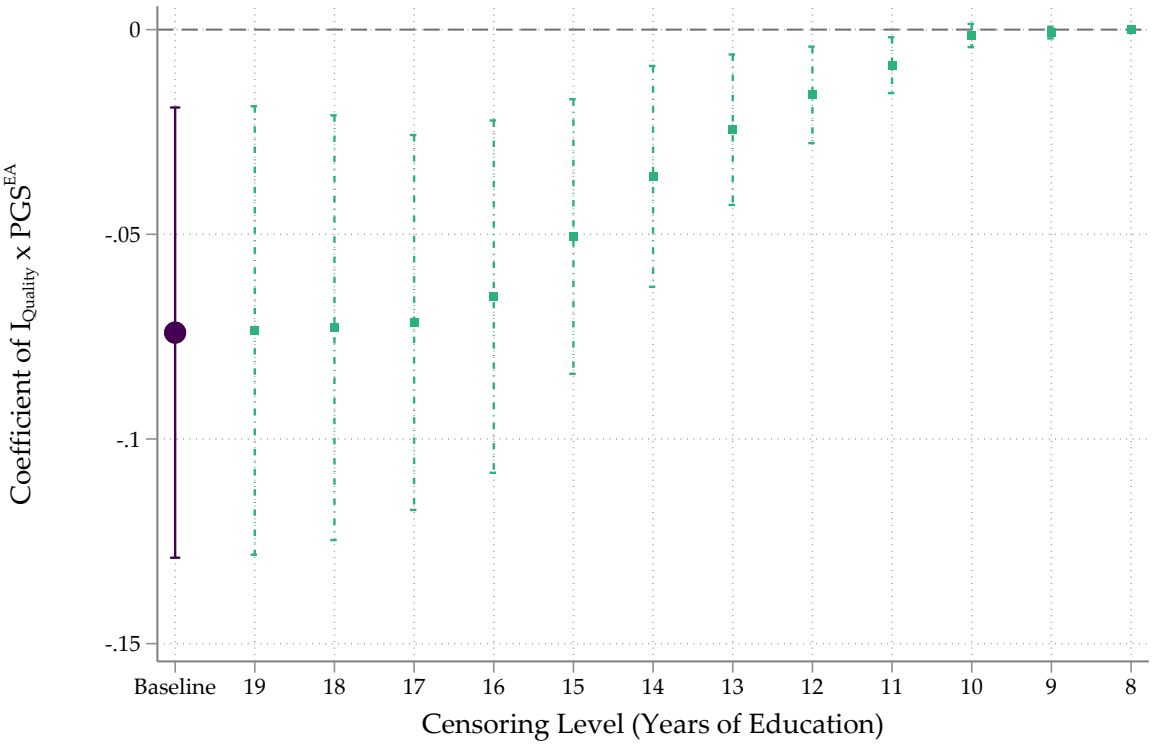

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows the sensitivity of the point estimate for $I_{Quality}$ to unobserved confounding variables. Following the procedure of Cinelli and Hazlett (2020), we calculate the bias-adjusted treatment effect of $I_{Quality}$ under different assumptions about the partial $R^2$ of confounding variables with the variables of interest and the partial $R^2$ of confounding variables with years of education. Each contour line shows p-values (or point estimates) for different combinations of the two partial $R^2$. Each circle shows resulting values for different multiples paternal education. The diamond shows baseline estimates from Figure 2. Standard errors are clustered at the school level.

**FIGURE B.3 – Sensitivity to Outlier Schools**



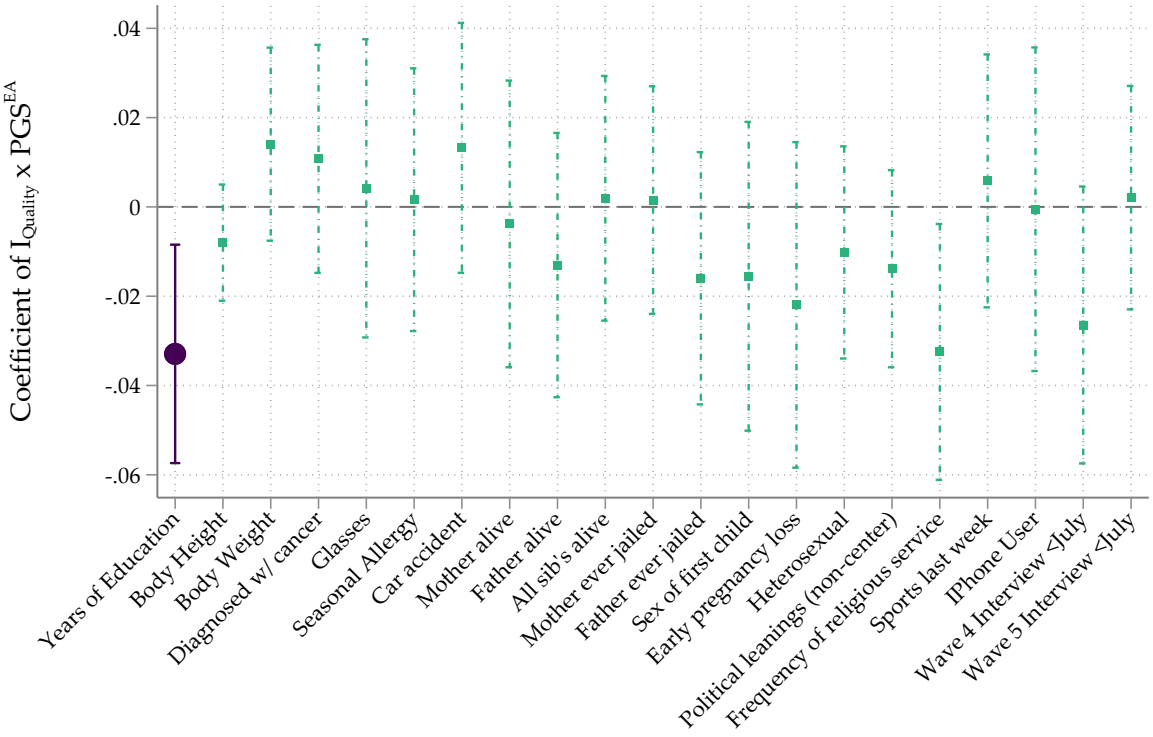

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows point estimates and 90% confidence bands of the interaction of $PGS^{EA}$ and $I_{Quality}$, and its association with years of education. Each estimate is derived from a subsample of the data in which we drop one High School, respectively. Estimates follow the specification of equation (7). *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. Standard errors are clustered at the school level.

**FIGURE B.4 – Sensitivity to Ceiling Effects**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows point estimates and 90% confidence bands of the interaction of $PGS^{EA}$ and $I_{Quality}$, and its association with years of education. Each estimate is derived from the full sample while censoring the outcome variable at different levels from above. Estimates follow the specification of equation (7). *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. Standard errors are clustered at the school level.

**FIGURE B.5 – Association of $PGS^{EA}$ and $I_{Quality}$ with Educational Attainment and Placebo Outcomes**




**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This figure shows point estimates and 90% confidence bands of the interaction of $PGS^{EA}$ and $I_{Quality}$, and its association with years of education as well as a series of placebo outcomes. Estimates follow the specification of equation (7). All outcomes are standardized on the estimation sample so that $\mu = 0$, $\sigma = 1$. *Child Controls*: Firstborn dummy, linear birth cohort trend (in months) by gender, 20 principal components of the full matrix of genetic data. *Family Controls*: Age of mother at birth, years of education of both mother and father, average potential wages of both mother and father, the standard deviation of potential wages of both mother and father, dummies for non-US born mothers and fathers, a dummy for Christian religion, state fixed effects. Standard errors are clustered at the school level.

# Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments

Benjamin Arold, Paul Hufe & Marc Stöckli

**Supplementary Material**

# A BIAS IN GENE-ENVIRONMENT INTERACTION

Assume the following population model:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \delta z_i + \epsilon_i,$$

where $x_1 \sim \mathcal{N}(0,1)$, $x_2 \sim \mathcal{N}(0,1)$, $x_3 = x_1 \times x_2$, $Cov(x_1, x_2) = 0$, and $Cov(\epsilon | x_1, x_2, z) = 0$. Suppose $z_i$ is unobserved and $\delta > 0$.

We estimate the following model via OLS:

$$y_i = \tilde{\alpha} + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \tilde{\beta}_3 x_{i3} + \tilde{\epsilon}_i,$$

Let's focus on the estimate of the interaction effect only:

$$plim\ \tilde{\beta}_3 = \beta_3 + \delta \frac{Cov(x_3, z)}{Var(x_3)} = \beta_3 + \underbrace{\delta \frac{Cov(x_1 \times x_2, z)}{Var(x_1) \times Var(x_2)}}_{=\text{Bias}}.$$

Following Bohrnstedt and Goldberger (1969), we can write $Cov(x_1 \times x_2, z)$ as follows:

$$Cov(x_1 \times x_2, z) = \mathbb{E}[x_1] \times Cov(x_2, z) + \mathbb{E}[x_2] \times Cov(x_1, z) + \mathbb{E}[\Delta x_1 \Delta x_2 \Delta z],$$

where $\Delta x_1 = (x_1 - \mathbb{E}[x_1])$, $\Delta x_2 = (x_2 - \mathbb{E}[x_2])$, $\Delta z = (z - \mathbb{E}[z])$.

Using $x_1, x_2 \sim \mathcal{N}(0,1)$ and $Cov(x_1, x_2) = 0$, this expression simplifies to:

$$Cov(x_1 \times x_2, z) = \mathbb{E}[\Delta x_1 \Delta x_2 \Delta z] = \mathbb{E}[\Delta x_1 \sqrt{\Delta z}] \mathbb{E}[\Delta x_2 \sqrt{\Delta z}].$$

Therefore, we can write the bias term as follows:

$$\delta \frac{\mathbb{E}[\Delta x_1 \sqrt{\Delta z}] \mathbb{E}[\Delta x_2 \sqrt{\Delta z}]}{Var(x_1) \times Var(x_2)}.$$

Note that $\mathbb{E}[\Delta x_1 \sqrt{\Delta z}] > 0 \leftrightarrow \tilde{\beta}_1 > \beta_1$ and $\mathbb{E}[\Delta x_2 \sqrt{\Delta z}] > 0 \leftrightarrow \tilde{\beta}_2 > \beta_2$. Therefore, the following conditions hold:

1. If $\tilde{\beta}_1 > \beta_1$ and $\tilde{\beta}_2 > \beta_2$ or $\tilde{\beta}_1 < \beta_1$ and $\tilde{\beta}_2 < \beta_2$: $\tilde{\beta}_3$ is upward biased.
2. If $\tilde{\beta}_1 < \beta_1$ and $\tilde{\beta}_2 > \beta_2$ or $\tilde{\beta}_1 > \beta_1$ and $\tilde{\beta}_2 < \beta_2$: $\tilde{\beta}_3$ is downward biased.

# B DATA APPENDIX

## *B.1 Outcome Variables*

**Educational attainment.** We measure educational attainment by total *years of education*. In each wave, respondents were asked about their highest level of education at the time of the interview. For each respondent, we use the most recent information and transform education levels into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: eighth grade or less (8), some high school (10), high school graduate (12), GED (12), some vocational/technical training (13), some community college (14), some college (14), completed vocational/technical training (14), associate or junior college degree (14), completed college (16), some graduate school (17), completed a master's degree (18), some postbaccalaureate professional education (18), some graduate training beyond a master's degree (19), completed post-baccalaureate professional education (19), completed a doctoral degree (20).

We use the most recent available information to construct the following measures for educational degrees: *High School* (including GED), *2-year College*, *4-year College*, and *Post-Graduate*. Two-year college degrees include associate and junior college degrees as well as vocational and technical training after high school. Four-year college degrees include bachelor's degrees. Post-graduate degrees include master's degrees, doctoral degrees, and post-baccalaureate professional degrees. If available, information is taken from wave 5; otherwise we take it from waves 4 or 3, respectively. We only include respondents for which we observe educational degrees when they are at least 27 years old at the time of observation. We assume an ordinal ranking of degrees (high school < 2-year college < 4-year college < post-graduate) and assign the possession of a lower-ranked degree if a respondent obtained a higher-ranked degree. For example, we assume that a respondent has finished high school if he or she has obtained a college degree, even if we don't have explicit information about high school graduation status.

**Health.** We proxy *subjective health* by quality-adjusted life years (QALY) that we derive from self-assessed health (SAH) measures. We use information from waves 3 and 4, where participants were asked "in general, how is your health?" We convert their (categorical) responses into a continuous measure using a mapping proposed by Van Doorslaer and Jones (2003). Using information about objective health—the Health Utility Index Mark III—Van Doorslaer and Jones (2003) scale the intervals of the SAH categories. This approach yields "quality weights" for health between 0 and 1. The values for each health status category are as follows (quality weights in parentheses): "excellent" (0.9833), "very good" (0.9311), "good" (0.841), "fair" (0.707), and "poor" (0.401).[1] We average resulting QALY measures across waves 3 and 4.

[1] See Table 4 in Van Doorslaer and Jones (2003).

We construct an index of *objective health* based on information from wave 4. Specifically, we sum the standardized values about whether a respondent (i) is obese, (ii) has stage one hypertension, and (iii) has high cholesterol (as indicated by the respondent). Each item was answered with either "yes" (= 1) or "no" (= 0). We reverse-code our measure of objective health such that higher values indicate better health.

**Cognitive skills.** The *Picture Vocabulary Test* (PVT) is a test for receptive hearing vocabulary and is a widely-used proxy for verbal ability and scholastic aptitude. To administer the PVT, an examiner presents a series of pictures to the respondent. There are four pictures per page, and the examiner speaks a word describing one of the pictures. The respondent then has to indicate the picture that the word describes. In our analysis we use age-adjusted PVT percentile ranks from wave 3 (Harris, 2020).

**preferences.** We construct two measures of preferences: *risk aversion* and *patience*. In waves 3 and 4, participants were asked (i) whether they like to take risks, and (ii) whether they live their life without much thought for the future. Questions were answered on a five-point Likert scale ranging from "strongly agree" to "strongly disagree." We reverse-code both measures and use averages from waves 3 and 4 in our analysis.

**Personality.** The Big Five personality traits are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Almlund et al., 2011). We use information from wave 4 to construct personality measures. Participants were asked a set of questions that each relate to one of the five personality traits. Questions were answered on a five-point Likert scale ranging from "strongly agree" to "strongly disagree." We use averages of the following questions in our analysis. *Openness*: (i) "I have a vivid imagination," (ii) "I have difficulty understanding abstract ideas" (reverse-coded), (iii) "I am not interested in abstract ideas" (reverse-coded), (iv) "I do not have a good imagination" (reverse-coded). *Conscientiousness*: (i) "I get chores done right away," (ii) "I like order," (iii) "I often forget to put things back in their proper place" (reverse-coded), (iv) "I make a mess of things" (reverse-coded). *Extraversion*: (i) "I am the life of the party," (ii) "I talk to a lot of different people at parties," (iii) "I don't talk a lot" (reverse-coded), (iv) "I keep in the background" (reverse-coded). Agreeableness: (i) "I sympathize with others' feelings," (ii) "I feel others' emotions," (iii) "I am not interested in other people's problems" (reverse-coded), (iv) "I am not really interested in others" (reverse-coded). *Neuroticism*: (i) "I have frequent mood swings," (ii) "I get upset easily," (iii) "I am relaxed most of the time" (reverse-coded), (iv) "I seldom feel blue" (reverse-coded).

**Parental investment (as-if unobserved).** To measure *parental time investments*, we use information on a series of activities that children have done with their mother or father in the last

four weeks. Specifically, the child is asked whether he or she has (i) gone shopping, (ii) played a sport, (iii) gone to a religious service or church-related event, (iv) talked about someone he or she is dating, or a party he or she went to, (v) gone to a movie, play, museum, concert, or sports event, (vi) had a talk about a personal problem he or she was having, (vii) had a serious argument about him or her behavior, (viii) talked about his or her school work or grades, (ix) worked on a project for school, (x) talked about other things he or she is doing in school. Questions were answered with "yes" (= 1) or "no" (= 0). We standardize answers to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health respondents and then sum by parent (Anderson, 2008; Kling et al., 2007).

Information about *breastfeeding* and *family income* is taken from wave 1. Parents were asked about whether the child was breastfed (yes or no) and about their income. We use the logarithm of the latter (replacing zero incomes with a 1 to prevent a loss of observations).

**Placebo outcomes** All placebo outcomes are binary variables (except body height and weight) and constructed using information from waves 4 or 5. Each respondent's *height* (in inches), *weight* (in pounds), and whether he or she (i) has even been diagnosed with *cancer/leukemia*, (ii) needs *vision correction*, (iii) has had *seasonal allergies* during the four weeks prior to the interview, and (iv) ever had a *motor vehicle accident* is taken from wave 5 if available, and wave 4 otherwise. Similarly, information about whether the respondent's biological mother or father is still *alive* and whether they have ever been in *jail*, whether any sibling has *died*, and the gender of the resondent's first-born child (male = 1, female = 0, no child = missing) is taken from wave 5 if available, and wave 4 otherwise. Whether the respondent has had a miscarriage (both males and females) is taken from wave 4 only (yes = 1, no = 0, respondent has never been pregnant/never had a pregnant partner = missing). Information about whether the respondent considers him or herself *heterosexual*, attends *religious service*, owns an *iPhone*, did any *sports* during the weeks prior to the interview, and his or her *political leanings* (not being "middle-of-the-road") is taken from wave 5 if available, and wave 4 otherwise. *Wave 4 Interview <July* and *Wave 4 Interview <July* equal 1 if the respective interview was conducted between January and June, and 0 otherwise.

### *B.2 Variables of Interest*

**Polygenic scores.** Add Health obtained saliva samples from consenting participants in wave 4. After quality control procedures, genotyped data is available for 9,974 individuals and 609,130 SNPs. Add Health uses this data and calculates a set of different PGS using summary statistics from existing GWAS. Our baseline measure $PGS^{EA}$ is based on statistics from Lee et al. (2018). In our analysis, we also use the PGS for body mass index (*BMI*) (Yengo et al., 2018), attention deficit hyperactivity disorder (*ADHD*) (Demontis et al., 2019), *depressive symptoms* (Howard et al., 2019), *intelligence* (Savage et al., 2018), *smoking* (Liu et al., 2019), and *sleep*

**TABLE S.1 – Summary Statistics (Outcomes)**

| | Obs. | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| *Educational Attainment* | | | | | |
| Years Education | 3,075 | 14.81 | 2.25 | 8.00 | 20.00 |
| High School Degree | 3,075 | 0.97 | 0.18 | 0.00 | 1.00 |
| 2-year College Degree | 3,075 | 0.53 | 0.50 | 0.00 | 1.00 |
| 4-year College Degree | 3,075 | 0.42 | 0.49 | 0.00 | 1.00 |
| Post-Graduate Degree | 3,075 | 0.15 | 0.36 | 0.00 | 1.00 |
| *Health* | | | | | |
| Subjective | 3,075 | 0.91 | 0.07 | 0.40 | 0.98 |
| Objective | 3,075 | 0.03 | 1.93 | -6.46 | 1.62 |
| *Cognitive Skills* | | | | | |
| Picture Vocabulary Test | 2,995 | 59.95 | 25.91 | 0.00 | 100.00 |
| *Preferences* | | | | | |
| Risk Aversion | 3,071 | 2.83 | 0.86 | 1.00 | 5.00 |
| Patience | 3,071 | 3.93 | 0.72 | 1.00 | 5.00 |
| *Personality* | | | | | |
| Openness | 3,053 | 3.63 | 0.63 | 1.00 | 5.00 |
| Conscientousness | 3,073 | 3.65 | 0.70 | 1.25 | 5.00 |
| Extraversion | 3,069 | 3.33 | 0.77 | 1.00 | 5.00 |
| Agreeableness | 3,071 | 3.87 | 0.58 | 1.00 | 5.00 |
| Neuroticism | 3,071 | 2.55 | 0.70 | 1.00 | 5.00 |
| *Parental Investments (As-If-Unobserved)* | | | | | |
| Breastfed | 2,844 | 0.48 | 0.50 | 0.00 | 1.00 |
| Time Investment Mother | 3,075 | 0.53 | 4.33 | -8.51 | 14.89 |
| Time Investment Father | 2,538 | 0.32 | 4.28 | -6.47 | 16.74 |
| Family income (log) | 2,588 | 3.76 | 0.73 | 0.00 | 6.91 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows summary statistics for outcome variables in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion.

*duration* (Jansen et al., 2019). All polygenic scores are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of genotyped Add Health respondents.

**School characteristics.** In wave 1 and 2, Add Health administered questionnaires to head-masters of Add Health schools. We use this information to construct indicators for high school investments using a principal components analysis that includes the following school-level

**TABLE S.2 – Summary Statistics (Placebo Outcomes)**

| | Obs. | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Body Height | 3,074 | 67.79 | 4.11 | 55.00 | 92.00 |
| Body Weight | 3,070 | 191.07 | 51.53 | 82.00 | 500.00 |
| Diagnosed w/ cancer | 3,075 | 0.02 | 0.15 | 0.00 | 1.00 |
| Glasses | 3,074 | 0.53 | 0.50 | 0.00 | 1.00 |
| Seasonal Allergy | 3,075 | 0.18 | 0.38 | 0.00 | 1.00 |
| Car accident | 3,075 | 0.10 | 0.30 | 0.00 | 1.00 |
| Mother alive | 3,073 | 0.94 | 0.24 | 0.00 | 1.00 |
| Father alive | 3,040 | 0.83 | 0.38 | 0.00 | 1.00 |
| All sib's alive | 2,960 | 0.09 | 0.28 | 0.00 | 1.00 |
| Mother ever jailed | 3,073 | 0.02 | 0.15 | 0.00 | 1.00 |
| Father ever jailed | 3,005 | 0.12 | 0.32 | 0.00 | 1.00 |
| Sex of first child | 2,147 | 0.51 | 0.50 | 0.00 | 1.00 |
| Early pregnancy loss | 1,730 | 0.22 | 0.41 | 0.00 | 1.00 |
| Heterosexual | 3,075 | 0.84 | 0.36 | 0.00 | 1.00 |
| Political leanings (non-center) | 3,011 | 0.53 | 0.50 | 0.00 | 1.00 |
| Frequency of religious service | 3,075 | 0.60 | 0.49 | 0.00 | 1.00 |
| Sports last week | 3,075 | 0.88 | 0.32 | 0.00 | 1.00 |
| IPhone User | 1,185 | 0.56 | 0.50 | 0.00 | 1.00 |
| Wave 4 Interview <July | 3,075 | 0.79 | 0.41 | 0.00 | 1.00 |
| Wave 5 Interview <July | 2,490 | 0.33 | 0.47 | 0.00 | 1.00 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows summary statistics for outcome variables in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion.

information: (i) average class size, (ii) share of teachers with a master degree, (iii) share of new teachers in the current school year, (iv) share of teachers with school-specific tenure of more than five years, and Herfindahl indices to measure teacher diversity with respect to (v) race and (vi) Hispanic background.[2] We also include school-level information about the average student-teacher ratio (number of full-time students per full-time equivalent teachers) in 1995/96 taken from the Common Core of Data (CCD) and the Private School Survey (PSS). We apply a factor rotation for interpretability reasons (oblique oblimin rotation of the Kaiser normalized matrix with $\gamma = 0$; see Gorsuch, 1983). The first component loads almost exclusively on average class size and average student-teacher ratio. Hence, we interpret this component, $I_{Quantity}$, as an indicator for the "quantity" of teachers. The second component primarily loads positively on the percentage of teachers with a master degree and the share of teachers with a tenure of more than five years; it loads negatively on the share of new teachers in the current

[2]Herfindahl indices are calculated by first squaring the share of each component and then summing up resulting values (i.e. $H = \sum_{i=1}^{N} a_i^2$, where $a_i$ is the share of component $i$, and $N$ is the total number of components). For the Herfindahl index for race, we include the schools' share of full-time classroom teachers that are (i) White, (ii) Black or African American, (iii) American Indian or Native American, (iv) and Asian or Pacific Islander. For the Herfindahl index for Hispanic background, we include the schools' share of full-time classroom teachers that are (i) Hispanic or of Spanish origin, and (ii) neither Hispanic nor of Spanish origin.

school year. We interpret this component, $I_{Quality}$, as an indicator for the "quality" of teachers. Both factors are coded such that higher values indicate higher school investments, i.e. higher teacher "quantity" investments (smaller classes) and higher teacher "quality" investments (better teachers), respectively. The calculated factors are orthogonal to each other by construction. They are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health high schools.[3]

**Family socio-economic status.** We use the *social origins factor score* constructed by Belsky et al. (2018). Their measure uses information about parental education, parental occupation, household income, and household receipt of public assistance in wave 1. The score is standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health respondents in wave 1.

**TABLE S.3 – Summary Statistics (Variables of Interest)**

| | Obs. | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| *Polygenic Scores* | | | | | |
| $PGS^{EA}$ | 3,075 | 0.05 | 1.00 | -4.13 | 3.39 |
| BMI | 3,075 | -0.02 | 1.00 | -3.42 | 3.56 |
| ADHD | 3,075 | -0.05 | 0.99 | -3.82 | 3.48 |
| Depressive Symptoms | 3,075 | -0.02 | 1.01 | -3.79 | 3.55 |
| Intelligence | 3,075 | 0.02 | 0.99 | -3.30 | 4.06 |
| Ever Smoker | 3,075 | -0.04 | 1.00 | -4.25 | 4.25 |
| Sleep Duration | 3,075 | 0.02 | 0.99 | -3.74 | 2.99 |
| *School Characteristics* | | | | | |
| $I_{Quality}$ | 3,075 | 0.06 | 0.91 | -3.04 | 1.79 |
| $I_{Quantity}$ | 3,075 | -0.02 | 0.82 | -2.65 | 2.59 |
| *Family SES* | | | | | |
| Social Origins Factor Score | 3,018 | 0.37 | 1.12 | -4.40 | 3.51 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows summary statistics for variables of interest in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion.

## *B.3 Control Variables*

**Child characteristics.** The child's *gender* (female or male, as indicated by the interviewer) is taken from the in-home questionnaire in wave 1.

[3] Note that in an oblique rotation, factors may be slightly correlated.

We calculate the child's *age* (in months) at each wave by subtracting the child's birth date from the date of interview. Because birth dates have minor inconsistencies across waves, we take averages across waves 1 to 4.

We use the first 20 *principal components* of full matrix of the genetic data. The components are obtained from a principal components analysis on the matrix of SNPs in Add Health (see Braudt and Harris, 2020, for a discussion). The principal components are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of genotyped Add Health respondents.

**Family socio-economic status.** We use information from wave 1 to construct measures of *parents' education*. We transform parents' highest degree into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: never went to school (0), eighth grade or less (8), some high school (10), completed vocational/technical training instead of high school (10), went to school but level unknown (12), respondent doesn't know (12), high school graduate (12), GED (12), completed vocational/technical training after high school (14), some college (14), completed college (16), professional training beyond a master's degree (19). Where available, mothers' and fathers' education refers to the resident parent. If this information is not available, we use the biological parents' education instead.

Information about *mother's age at birth* (in years) is obtained from wave 1 if available, and wave 2 otherwise. To calculate age at birth, we take information about mother's age (as indicated by the child) and subtract the age of the child at the respective wave.

Information about religion (*Christian* or not) is obtained from wave 1 (as indicated by the child).

We calculate *potential wages* for population group $g$ in time period $t$ according to the following formula (Shenhav, 2021):

$$\hat{w}_{gt} = \sum_{j} \frac{E_{jg,1970}}{E_{g,1970}} \times \sum_{o} \frac{E_{ojg,1970}}{E_{jg,1970}} \left(\pi_{ojt,-r}\right) \times w_{ojt,-r},$$

where $\frac{E_{jg,1970}}{E_{g,1970}}$ describes the group-specific employment share of industry $j$ in 1970, $\frac{E_{ojg,1970}}{E_{jg,1970}}$ describes the group- and industry-specific employment share of occupation $o$ in 1970, $\pi_{ojt,-r}$ describes the leave-region-out industry-specific employment growth in occupation $o$ for year $t$ relative to 1970 (scaled by the overall employment growth in occupation $o$ for year $t$ relative to 1970), and $w_{ojt,-s}$ describes the leave-region-out average hourly wage paid in year $t$ for each occupation/industry/region cell. We define groups $g$ by individuals that are homogeneous in gender (male, female), educational attainment ($<$ High School, High School, $>$ High School), and ethnicity (Non-Hispanic White, Hispanic, Non-Hispanic Black). We define regions $r$ by census regions (North-East, Midwest, South, West). Employment shares in 1970 are taken from the 1970 decennial census. Employment shares and wages in periods $t$ are taken from the March Supplements of the Current Population Survey (CPS) over the time period 1975-2000.

We match time series of $\hat{w}_{gt}$ to the parents of respondents in Add Health based on information about $g$. Then we calculate (i) mean potential wages across respondent ages 0–14, and (ii) the standard deviation in potential wages across respondent ages 0–14.

**School characteristics.** We use information about school *peer characteristics* from the in-school questionnaire in wave 1. Specifically, for each school we calculate average years of education of students' fathers, the share of single parents, and the average subjective likelihood of students to attend college. We transform the father's highest degree into years of education following the mapping suggested by Domingue et al. (2015). Numeric values in parentheses: never went to school (0), eighth grade or less (8), some high school (10), went to school but level unknown (12), respondent doesn't know (12), high school graduate (12), GED (12), completed vocational/technical training after high school (14), some college (14), completed college (16), professional training beyond a four-year college (19). For college aspiration, students indicate how likely it is that they will graduate from college. Responses range from "no chance" (= 0) to "it will happen" (= 8). We define a student to have college aspiration if his or her response is above "about 50-50" (= 4), and to have no college aspiration otherwise. To prevent mechanical correlation between school peer characteristics and respondent characteristics, we calculate averages and shares while excluding individual respondents (leave-one-out).

We use information from the school administrator questionnaire in wave 1 to construct measures of *sanction policies* by means of a principal components analysis. School administrators were asked what happens to a student who is caught in their school (i) cheating, (ii) fighting with another student, (iii) injuring another student, (iv) possessing alcohol, (v) possessing an illegal drug, (vi) possessing a weapon, (vii) drinking alcohol at school, (viii) using an illegal drug at school, (ix) smoking at school, (x) verbally abusing a teacher, (xi) physically injuring a teacher, and (xii) stealing school property. Responses are "minor action", "in-school suspension", "out-of-school suspension", and "expulsion." Administrators were asked about sanctions in response to both first and second occurrences. We apply a factor rotation for interpretability reasons (oblique oblimin rotation of the Kaiser normalized matrix with $\gamma = 0$; see Gorsuch, 1983). The first three components load on variables reflecting the school's strictness regarding (i) drug use, (ii) social misconduct, and (iii) academic misconduct. The calculated factors are orthogonal to each other by construction. They are standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of Add Health high schools.[4]

We calculate *value-added measures* with respect to GPAs in subject $s$ for cohort $c$ visiting high

[4]Note that in an oblique rotation, factors may be slightly correlated.

school $j$ following a two-step procedure (Chetty et al., 2014):

$$\text{GPA}^s_{igjc} = \beta^s Z_{igjc} + \text{VA}^s_{jc} + \epsilon^s_{igjc},$$
$$\widehat{\text{VA}}^s_{jc} = \frac{1}{N}\sum_{i \in jc}^{N}(\text{VA}^s_{jc} + \widehat{\epsilon}^s_{igjc}).$$

$Z_{igjc}$ contains grade fixed effects $\delta_g$, lagged GPAs from grade levels $g-1$ for English, Math and Science as well as current and lagged grade- and subject-specific indicators for academic tracks in English, Math and Science (3 levels per grade times subject cell). To avoid mechanical relationships, we predict $\widetilde{\text{VA}}^s_{jc}$ excluding data from cohort $c$ and choosing a weighting vector $\phi^s = [\phi^s_{c-5}, ..., \phi^s_{c+5}]$ that minimizes the out-of-sample mean-squared error. Hence, $\widetilde{\text{VA}}^s_{jc}$ is our best prediction based on other cohorts of how much school $j$ will increase GPAs in subject $s$ in one year of high school relative to the improvements of similar students at other schools. We calculate $\widetilde{\text{VA}}^s_{jc}$ for English, Math and Science. In turn, we run a principal component analysis and use the first principal component as the aggregate measure of school value-added. The principal component is standardized to $\mu = 0$ and $\sigma = 1$ on the full sample of high schools with available transcript data on Add Health respondents.

**TABLE S.4 – Summary Statistics (Controls)**

| | Obs. | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| *Child Characteristics* | | | | | |
| Female | 3,075 | 0.55 | 0.50 | 0.00 | 1.00 |
| Age in Months (Wave 1) | 3,075 | 193.63 | 19.77 | 144.00 | 256.00 |
| Principal Component 1 | 3,075 | 0.00 | 0.01 | -0.14 | 0.10 |
| Principal Component 2 | 3,075 | -0.00 | 0.01 | -0.37 | 0.07 |
| Principal Component 3 | 3,075 | 0.00 | 0.01 | -0.10 | 0.02 |
| Principal Component 4 | 3,075 | 0.00 | 0.01 | -0.09 | 0.65 |
| Principal Component 5 | 3,075 | -0.00 | 0.01 | -0.07 | 0.18 |
| Principal Component 6 | 3,075 | -0.00 | 0.01 | -0.14 | 0.19 |
| Principal Component 7 | 3,075 | -0.00 | 0.01 | -0.13 | 0.33 |
| Principal Component 8 | 3,075 | -0.00 | 0.01 | -0.37 | 0.08 |
| Principal Component 9 | 3,075 | 0.00 | 0.01 | -0.06 | 0.07 |
| Principal Component 10 | 3,075 | -0.00 | 0.01 | -0.58 | 0.26 |
| Principal Component 11 | 3,075 | 0.00 | 0.01 | -0.25 | 0.37 |
| Principal Component 12 | 3,075 | 0.00 | 0.01 | -0.39 | 0.18 |
| Principal Component 13 | 3,075 | -0.00 | 0.01 | -0.35 | 0.18 |
| Principal Component 14 | 3,075 | -0.00 | 0.01 | -0.12 | 0.23 |
| Principal Component 15 | 3,075 | 0.00 | 0.01 | -0.28 | 0.23 |
| Principal Component 16 | 3,075 | 0.00 | 0.02 | -0.15 | 0.66 |
| Principal Component 17 | 3,075 | -0.00 | 0.01 | -0.50 | 0.24 |
| Principal Component 18 | 3,075 | -0.00 | 0.01 | -0.29 | 0.20 |
| Principal Component 19 | 3,075 | 0.00 | 0.01 | -0.26 | 0.46 |
| Principal Component 20 | 3,075 | -0.00 | 0.01 | -0.18 | 0.27 |
| *Family SES* | | | | | |
| Education Mother (in Years) | 3,075 | 13.63 | 2.50 | 8.00 | 19.00 |
| Education Father (in Years) | 3,075 | 13.67 | 2.68 | 8.00 | 19.00 |
| Maternal Age at Birth | 3,075 | 25.49 | 4.83 | 16.00 | 44.33 |
| Christian | 3,075 | 0.82 | 0.38 | 0.00 | 1.00 |
| Potential Wage/Hour Mother (Mean) | 3,075 | 12.62 | 1.38 | 9.45 | 14.27 |
| Potential Wage/Hour Father (Mean) | 3,075 | 15.48 | 1.31 | 11.14 | 17.11 |
| Potential Wage/Hour Mother (SD) | 3,075 | 0.36 | 0.11 | 0.12 | 0.51 |
| Potential Wage/Hour Father (SD) | 3,075 | 0.40 | 0.08 | 0.20 | 0.65 |
| *School Characteristics* | | | | | |
| Peer Characteristics (Educ. Father) | 2,959 | 13.57 | 1.05 | 10.90 | 17.84 |
| Peer Characteristics (Single Parents) | 2,959 | 0.24 | 0.08 | 0.00 | 0.60 |
| Peer Characteristics (College Aspir.) | 2,959 | 0.76 | 0.08 | 0.44 | 1.00 |
| Sanction Policies (Drugs) | 2,993 | 0.06 | 0.85 | -2.65 | 4.16 |
| Sanction Policies (Social) | 2,993 | 0.12 | 0.76 | -2.98 | 2.36 |
| Sanction Policies (Acad.) | 2,993 | 0.05 | 0.83 | -2.42 | 1.69 |
| Value-Added (GPA) | 2,768 | 0.21 | 1.55 | -4.18 | 4.41 |

**Data:** National Longitudinal Study of Adolescent to Adult Health.
**Note:** Own calculations. This table shows summary statistics for control variables in our core analysis sample. The sample is restricted to genotyped individuals of (i) European descent, (ii) who visited an Add Health high school or an associated feeder school in wave 1, and (iii) who graduated from the same school. Observations with missing information in any of the displayed variables are dropped by list-wise deletion.

## References


ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, and T. KAUTZ (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education*. Ed. by E. A. HANUSHEK, S. MACHIN, and L. WOESSMANN. Vol. 4. Elsevier, pp. 1–181.

ANDERSON, M. L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American Statistical Association* 103 (484), pp. 1481–1495.

BELSKY, D. W. et al. (2018). "Genetic Analysis of Social-class Mobility in Five Longitudinal Studies". *Proceedings of the National Academy of Sciences* 115 (31), E7275–E7284.

BOHRNSTEDT, G. W. and A. S. GOLDBERGER (1969). "On the Exact Covariance of Products of Random Variables". *Journal of the American Statistical Association* 64 (328), pp. 1439–1442.

BRAUDT, D. and K. M. HARRIS (2020). "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health)—Release 2".

CHETTY, R., J. N. FRIEDMAN, and J. E. ROCKOFF (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates". *American Economic Review* 104 (9), pp. 2593–2632.

DEMONTIS, D., R. K. WALTERS, J. MARTIN, M. MATTHEISEN, T. D. ALS, E. AGERBO, G. BALDURSSON, R. BELLIVEAU, J. BYBJERG-GRAUHOLM, M. BÆKVAD-HANSEN, et al. (2019). "Discovery of the First Genome-wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder". *Nature Genetics* 51 (1), pp. 63–75.

DOMINGUE, B. W., D. BELSKY, D. CONLEY, K. M. HARRIS, and J. BOARDMAN (2015). "Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health". *AERA Open* 1 (3), pp. 1–13.

GORSUCH, R. (1983). *Factor Analysis*. 2nd ed. Hillsdale, NJ: Psychology Press.

HARRIS, K. M. (2020). "National Longitudinal Study of Adolescent to Adult Health, Public Use Peabody Picture Vocabulary Test (PVT) Score Data, Wave III [dataset]". *National Longitudinal Study of Adolescent to Adult Health*.

HOWARD, D. M., M. J. ADAMS, T.-K. CLARKE, J. D. HAFFERTY, J. GIBSON, M. SHIRALI, J. R. COLEMAN, S. P. HAGENAARS, J. WARD, E. M. WIGMORE, et al. (2019). "Genome-wide Meta-analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions". *Nature Neuroscience* 22 (3), pp. 343–352.

JANSEN, P., K. WATANABE, S. STRINGER, N. SKENE, J. BRYOIS, A. HAMMERSCHLAG, C. DE LEEUW, J. BENJAMINS, A. MUÑOZ-MANCHADO, M. NAGEL, et al. (2019). "Genome-wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways". *Nature Genetics* 51 (3), pp. 394–403.

KLING, J. R., J. B. LIEBMAN, and L. F. KATZ (2007). "Experimental Analysis of Neighborhood Effects". *Econometrica* 75 (1), pp. 83–119.

LEE, J. J. et al. (2018). "Gene Discovery and Polygenic Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals". *Nature Genetics* 50 (8), pp. 1112–1121.

LIU, M., Y. JIANG, R. WEDOW, Y. LI, D. BRAZEL, F. CHEN, G. DATTA, J. DAVILA-VELDERRAIN, D. MCGUIRE, C. TIAN, et al. (2019). "Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use". *Nature Genetics* 51 (2), pp. 237–244.

SAVAGE, J. E. et al. (2018). "Genome-wide Association Meta-analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence". *Nature Genetics* 50 (7), pp. 912–919.

SHENHAV, N. (2021). "Lowering Standards to Wed? Spouse Quality, Marriage, and Labor Market Responses to the Gender Wage Gap". *Review of Economics and Statistics* 103 (2), pp. 265–279.

VAN DOORSLAER, E. and A. JONES (2003). "Inequalities in Self-reported Health: Validation of a New Approach to Measurement". *Journal of Health Economics* 22 (1), pp. 61–87.

YENGO, L., J. SIDORENKO, K. E. KEMPER, Z. ZHENG, A. R. WOOD, M. N. WEEDON, T. M. FRAYLING, J. HIRSCHHORN, J. YANG, P. M. VISSCHER, et al. (2018). "Meta-analysis of Genome-wide Association Studies for Height and Body Mass Index in 700000 Individuals of European Ancestry". *Human Molecular Genetics* 27 (20), pp. 3641–3649.