Modeling and Measuring the Genetic Determinants of Child Development*

Francesco Agostinelli
University of Pennsylvania

Zach Weingarten
University of Pennsylvania

March 5, 2026

Abstract

The longstanding debate over whether human capabilities and skills are shaped more by "nature" or "nurture" has been revitalized by recent advances in genetics, particularly in the use of polygenic scores (PGSs) to proxy for genetic endowments. Yet, we argue that PGSs embed not only direct genetic effects but also indirect environmental influences, raising questions about their validity for causal analysis. We show that these conflated measures can mislead studies of gene–environment interactions, especially when parental behavior responds to children's genetic risk. To address this issue, we construct a new latent measure of genetic risk that integrates individual genotypes with diagnostic symptoms, using data from the National Longitudinal Study of Adolescent to Adult Health linked to restricted individual SNP-level genotypes from dbGaP. Exploiting multiple sources of variation—including the Mendelian within-family genetic randomization among siblings—we find consistent evidence that parents compensate by investing more in children with higher genetic risk for ADHD. Strikingly, these compensatory responses disappear when genetic risk is proxied by the conventional ADHD PGS, which also yields weaker—and in some cases reversed—predictions for long-run outcomes. Finally, we embed our latent measure of genetic endowments into a standard dynamic structural model of child development. The model shows that both parental investments and latent genetic risk jointly shape children's cognitive and mental health development, underscoring the importance of modeling the dynamic interplay between genes and environments in the formation of human capital.

Keywords: gene-environment interaction; human capital formation; parental investment; cognitive skills; mental health; ADHD.

JEL Codes: I24, I14, J24, J13, D15

*We thank James Heckman and Kevin Thom for helpful comments and suggestions. We also thank attendees at the ASSA 2026 annual meeting and seminar participants at the University of Pennsylvania for helpful suggestions that improved the paper. Contact information: Agostinelli: fagostin@upenn.edu; Weingarten: zachwein@upenn.edu.

1 Introduction

The question of whether differences in human capital arise primarily from genetic endowments or environmental influences is one of the oldest debates in the social sciences, with roots stretching from Enlightenment-era empiricism (Locke, 1689) through early discussions of heredity and behavior in the late nineteenth and early twentieth centuries (Galton, 1874; Watson, 1924), and later shaped by influential contributions across psychology, sociology, and economics. In recent years, this longstanding debate has been revisited in light of advances in genetics and the increasing availability of large-scale datasets linking genetic variation to health and behavioral outcomes. Following the completion of the Human Genome Project (Lander et al., 2001) and the emergence of genome-wide association studies (GWAS) in the mid-2000s (Manolio et al., 2009), researchers have developed polygenic scores—summary measures that aggregate the effects of many trait-associated genetic variants (single-nucleotide polymorphisms, SNPs). In standard GWAS analyses, these effects are estimated by regressing the trait of interest on each SNP while controlling for relevant covariates. Polygenic scores therefore combine the small contributions of many variants into a single predictive index for an individual. Since their introduction, they have been applied to a wide range of complex traits, including health, psychological, behavioral, economic, and educational outcomes (Benjamin et al., 2012; Belsky et al., 2016; Belsky and Harden, 2019).

These advances in quantitative genetics have sparked a growing literature exploring the interplay between genetic and environmental factors and their causal impact on educational attainment. The focus of quantitative inquiry has shifted from merely decomposing outcome variance explained by genetic variation to addressing a more fundamental causal question: to what extent do differences in genetic endowments cause variation in children's outcomes, and how do they moderate the heterogeneous effects of environmental influences across individuals?

In this paper, we argue that advancing toward a causal analysis of the interplay between genetic endowments and environments—often referred to as gene–environment ($G \times E$) interactions—requires a fundamental rethinking of how these endowments are measured. Standard polygenic scores (PGSs) are constructed as weighted sums of trait-associated genetic variants, with each variant assigned a "reduced-form" weight that captures its statistical association with the outcome of interest. We show that these estimated weights conflate different mechanisms: they reflect not only the direct influence of genetic variation on the trait but also indirect pathways through which genetic factors shape mediating environmental conditions.

Crucially, when the goal is to identify the causal $G \times E$ interaction—that is, how $\frac{\partial \mathbb{E}[Y|G,E]}{\partial G}$ varies

with E, where Y can represent any outcome such as test scores, educational attainment, or intermediate inputs in a child's skill production like parental investment—using polygenic scores (PGSs) to measure G can confound true genetic variation with endogenous variation in E. Because PGS weights themselves incorporate information about the environment, they cannot be treated as fixed in a hypothetical causal counterfactual. Instead, these weights are endogenously shaped by E, which may itself respond to changes in G. For example, family and parental influences, which play a pivotal role in children's skill formation (Heckman and Mosso, 2014), may adjust in response to shifts in genetic endowment.

To make progress on these challenges, we undertake two complementary tasks. First, we begin by constructing a novel measure of latent genetic predisposition that combines observed genotypes with early-life phenotypes that are informative of a given trait, avoiding the conventional polygenic score, as we are concerned that it attaches behaviorally confounded weights to the genetic variants. Second, we develop a structural model of child development that allows for direct genetic effects from the endogenous behavioral responses they induce in the family environment. In our dynamic model, parents make investment decisions in the presence of their child's latent genetic risk and in response to their child's evolving cognitive ability and mental health.

We apply this framework to study how genetic predisposition to Attention-Deficit/Hyperactivity Disorder (ADHD)—a neurodevelopmental condition affecting more than 11% of children in the United States (Visser et al., 2014)—influences children's cognitive and mental health development. ADHD provides an ideal setting to study the interplay between nature and nurture on child development. It is highly heritable, with twin studies attributing about 74% of its variation to genetic differences (Faraone and Larsson, 2019), yet its phenotypic expression—and the consequences of that expression for children's skill formation—is shaped by environmental factors, including family context, parenting practices, behavioral interventions, and broader social environments (Staff et al., 2023; Claussen et al., 2024). In terms of policy relevance, ADHD is also associated with many adverse long-run economic outcomes, including early school exit (Fleming et al., 2017), lower post-secondary attainment (Loe and Feldman, 2007), higher unemployment (Rietveld and Patel, 2019), and increased risk of criminal involvement (Fletcher and Wolfe, 2009).

Our empirical analyses draw on the integration of two complementary datasets. First, we use the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative panel that follows a cohort of adolescents from middle and high school into adulthood. Add Health provides rich longitudinal data on individual behavior, family environment, educational attainment, labor market outcomes, and survey-based measures of behaviors predictive of ADHD, as well as ADHD diagnoses. Most respondents consented to genetic data collection, and conventional polygenic scores are available within

the core dataset. However, SNP-level genotype data on children in Add Health are only accessible through the restricted-use Database of Genotypes and Phenotypes (dbGaP).[1] We merge this SNP-level genetic data from dbGaP with the Add Health survey to construct our final analysis dataset.

We construct measures of each child's cognitive skills, mental health, and parental investment from multiple survey-based indicators, applying the Bartlett scoring procedure to extract the underlying latent traits. Our main innovation is the development of a latent measure of genetic risk for ADHD. Instead of relying on the conventional ADHD PGS from Demontis et al. (2019)—which can embed environmental bias through its dependence on diagnosed cases and the associated non-random variation in parenting practices and home environments—we treat individual SNPs as noisy indicators of underlying genetic risk. We also incorporate the same 18 behavioral symptoms physicians use to diagnose ADHD in children as additional noisy proxies.

To guide the development of our structural model, we conduct a series of regression analyses that exploit within-family Mendelian randomization, leveraging genetic differences across siblings (Gray and Wheatley, 1991; Davey Smith and Ebrahim, 2003). The random assortment of alleles at conception generates variation in genetic endowments among siblings, providing a natural experiment that enables the identification of causal effects free from confounding due to shared familial genetic heritage. Although siblings inherit much of their genetic material from their parents, Mendelian randomization—rooted in the principles established in the mid-19th century by Gregor Mendel, often regarded as the founder of modern genetics—captures the random variation in genetic traits that arises within families.

We examine how parents adjust their investments in response to random variation in their child's genetic risk for ADHD. We find suggestive evidence of compensatory behavior: parents systematically invest more in siblings with higher genetic risk. We then extend this analysis to examine long-run outcomes in adulthood and find that our measure of latent genetic risk for ADHD predicts ADHD diagnosis, arrests, and illicit drug use. By contrast, when we replicate these analyses using the ADHD polygenic score (ADHD PGS), the results diverge: parental investments no longer respond to children's genetic risk, the predictive power for actual ADHD diagnosis falls by two-thirds, and the effects on adult outcomes reverse in sign. These discrepancies are consistent with our earlier discussion that PGSs may embed environmental influences in their construction, thereby generating bias, and they underscore the need for caution when relying on the PGS as a proxy for genetic risk in economic and behavioral research.

Motivated by these findings, we develop and estimate a structural model in which parents choose

[1] Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation across individuals. Each SNP represents a difference in a single DNA building block and serves as a basic unit in genome-wide analyses. See Section 2 for more details.

investments in response to their child's cognitive ability, mental health, and latent genetic risk for ADHD. In turn, both cognitive and mental health skills evolve dynamically as functions of parental investments, genetic endowments, and current skill levels. The evolution of skills follows a technology with dynamic complementarities across inputs. Siblings differ in genetic risk and initial endowments but share a family-level unobservable that may be correlated with both, influencing their skill development over time.

We estimate the model via indirect inference, matching simulated and empirical moments from both pooled and within-family regressions. This approach ensures that we leverage the same identifying variation as in our reduced-form analysis, including the quasi-random genetic variation generated by Mendelian randomization across siblings. The framework thus integrates reduced-form evidence with structural estimation, enabling us to quantify the causal role of genetic endowments in shaping parental behavioral responses and to conduct counterfactual quantitative exercises.

Our model shows that genetic endowment is a critical input in the production of both cognitive ability and mental health during adolescence. Genetic risk for ADHD directly hinders the development of each skill and indirectly shapes outcomes through parental responses, as parents compensate with greater investment. The estimated production functions reveal positive complementarities between genetic endowment and existing skills, meaning that current skill levels amplify the effect of genetics. As a result, when we take a step back in the dynamic process, today's parental investments not only raise current skills but also shape how genetic endowments will influence child development in the future. We also find negative complementarities between genetic endowment and parental investments, implying that, all else equal, returns to parental investments are higher for children with greater genetic risk. Taken together, these patterns highlight the dynamic nature of early disadvantage: genetic risk affects both the level of skills and their marginal productivity, with consequences shaped by how families adapt their behavior in response.

To benchmark our findings against the conventional approach of using the PGS as the measure for genetic endowment, we re-estimate the model using the ADHD PGS in place of our latent genetic ADHD risk measure. This alternative specification yields substantially different results. The estimated direct effect of genetic endowment on skill formation reverses sign, and the behavioral response to genetic risk attenuates. The complementarity between genetics and cognitive skills is attenuated by 74%. As a result of these differences, the model produces divergent predictions about the long-run effects of genetic endowment. For instance, increasing genetic endowment by 10% in the model estimated with our latent measure raises adult cognitive skills by approximately 0.07 standard deviations and raises adult mental health by 0.3 standard deviations. Conversely, this same increase in genetic endowment in a model estimated with

the standard PGS decreases cognitive skills in adulthood by 0.14 standard deviations and increases adult mental health by 0.03 standard deviations. Consistent with our reduced-form evidence, these differences highlight how conventional genetic proxies, which conflate genetic and environmental influences, are susceptible to bias and can distort conclusions about the role of nature in skill formation.

Related literature. Our paper relates to several strands of the literature on child development and human capital formation. First, we contribute to work that studies the interplay between nature and nurture in early skill formation (Cunha and Heckman, 2007, 2008; Cunha, Heckman and Schennach, 2010). Within this framework, we continue to build on the seminal work of Becker and Tomes (1979), in which parental investment in time and resources is conceptualized as a key component of the nurture environment (Todd and Wolpin, 2007; Cunha and Heckman, 2008; Del Boca, Flinn and Wiswall, 2014; Caucutt et al., 2020; Attanasio et al., 2020; Giannola, 2024; Attanasio, De Paula and Toppeta, 2025; Agostinelli and Wiswall, 2025; Agostinelli et al., 2025). These studies show that parental behavior can amplify or attenuate children's initial skill endowments, and that these responses account for a significant share of the disparities in human capital observed in adulthood.

We extend this strand by developing a structural model of child development that embeds both genetic endowment and parental investment within a multidimensional technology of skill formation. The model allows for dynamic complementarities and substitution between nature and nurture, and explicitly incorporates how parental investment decisions respond to a child's endowments. Using ADHD as an empirical application, we quantify how variation in genetic risk interacts with parental behavior to shape the joint evolution of cognitive skills and mental health across adolescence.

The second strand of literature our work relates to examines how genetic endowments interact with environments in the development of human capital. Early work relied on twin and adoption designs to separately identify the role of heredity and the role of the environment (Sacerdote, 2002; Krueger and Ashenfelter, 1992; Plug and Vijverberg, 2003; Sacerdote, 2007). More recently, the advent of polygenic scores has enabled economists to use GWAS results to construct individual-level measures of genetic risk and genetic endowments, which have been used to study the role of nature in labor and health economics (Domingue et al., 2015; Barth, Papageorge and Thom, 2020; Amin et al., 2021; Bertoni, Marin-Lopez and Sanz-de Galdeano, 2023; Jeong et al., 2024). This work has also combined polygenic scores with twin and sibling designs and leveraged parental genotype data to separate direct genetic effects from environmentally mediated ones (Sanz-de Galdeano and Terskaya, 2022; Cawley et al., 2023; Houmark, Ronda and Rosholm, 2024; Muslimova et al., 2024). Relative to this strand, we leverage rich data on adolescents' geno-

types and phenotypes to construct a latent genetic measure that separates direct genetic influence from environmental mediation. We then embed this measure in a structural model with endogenous parental investment to quantify how gene–environment interactions shape the joint evolution of cognitive skills and mental health in adolescence.

Finally, our work relates to a growing literature that incorporates genetic endowments into structural models of human capital formation. These studies build on the skill formation framework of Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010) and extend it to allow for variation in endowments and in the productivity of parental and environmental inputs (Rustichini et al., 2023; Houmark, Ronda and Rosholm, 2024; Bolt et al., 2025). To the best of our knowledge, Barth, Papageorge and Thom (2020) is the only remaining paper that embeds genetics into a dynamic structural model, although their focus is on the role of genetics in wealth accumulation during adulthood. Each of these studies relies on the educational attainment polygenic score (Lee et al., 2018). We instead focus on ADHD and use the Demontis et al. (2019) polygenic score as the conventional benchmark in our setting. Relative to this prior work, we develop a new measure of genetic endowment that does not rely directly on the polygenic score and show that the latter is susceptible to omitted variable biases. We then compare two versions of the same structural model, one incorporating our measure and the other the conventional polygenic score, and find that the latter produces different implications for parental investment responses and the long-run evolution of cognitive skills and mental health.

The rest of the paper is organized as follows. Section 2 provides an overview of genetic mechanisms and discusses potential biases that arise when using polygenic scores to measure pure genetic endowment. Section 3 describes our data and presents results from the causal inference analysis. Section 4 develops a structural model of child development with endogenous parental investment decisions in the presence of genetic endowments. Section 5 outlines our estimation strategy and presents the results. Section 6 examines the discrepancies that emerge when comparing our latent measure of genetic risk to the conventional polygenic score. Finally, Section 7 concludes.

2 Genetics

2.1 A Brief Primer on Genetics

This section offers a concise overview of key concepts and recent advances in genetics. Benjamin et al. (2024) provides a comprehensive discussion of these developments.

Human DNA is distributed across 23 pairs of chromosomes and consists of sequences of the nucleotide

Figure 1: The Process of Mendelian Inheritance for Single-Nucleotide Polymorphisms

Maternal Base-Pair Sequence:	TT	AA	TT	AA	AT
Paternal Base-Pair Sequence:	TT	AA	AT	AT	AT
	↓	↓	↓	↓	↓
Potential Child Sequence:	TT	AA	TT AT	AT AA	TT AT AA
Set of Potential SNP Values:	$\{0\}$	$\{2\}$	$\{0, 1\}$	$\{1, 2\}$	$\{0, 1, 2\}$

NOTES: The above figure displays a single-nucleotide polymorphism occurring at some base-pair locus for theoretical pairs of parents. We consider a SNP with adenine and thymine possible realizations. Children's potential sequence realizations are displayed in the third row, with the corresponding SNP values beneath.

bases adenine (A), cytosine (C), guanine (G), and thymine (T). Each human cell contains approximately 3 billion base pairs, where the bases form complementary pairs: adenine pairs with thymine (A-T), and cytosine pairs with guanine (C-G). These base-pair locations collectively constitute the human genome. Nearly all of the base-pair locations across the human genome are identical in the population; however, the variation that occurs at less than 1% of locations induces the observed and unobserved differences in humans. Whenever a particular base-pair location experiences sizable variation in the realized nucleotide, it is referred to as a single-nucleotide polymorphism (SNP).[2] SNPs are often measured based on the number of reference minor alleles present at a given base-pair location, meaning they can take values of 0, 1, or 2, depending on whether an individual has zero, one, or two copies of the minor allele.

Figure 1 displays an example of a hypothetical SNP at some base-pair location for five different pairs of parents. Each column shows the possible set of genetic endowments a child could receive based on their parents' SNPs. Importantly, this inheritance process is independent between parents and across siblings. The first two columns demonstrate instances in which siblings would share SNPs with certainty. The last three columns provide examples of within-family randomization. For example, a child born to a mother with TT and a father with AT will have, in expectation, AT with probability 0.5 and TT with probability 0.5.[3]

Over the last two decades, geneticists have conducted genome-wide association studies (GWAS) to determine whether any SNPs predict a particular trait of interest, such as heart disease or schizophrenia.

[2] Here, "sizable" refers to at least 1% of the human population.

[3] The child receives T from their mother with certainty and receives adenine from their father 50% of the time. Therefore, the possible SNP realizations for the child are AT, AT, TT, and TT.

In a standard GWAS, the sample includes a large number of treated (those with the trait) and control (those without) participants. The researchers then estimate J equations of the form

$$y_i = \beta_{0j} + \beta_{1j}\text{SNP}_{ij} + \nu_{ij}, \tag{1}$$

where y_i is the trait of interest, $\text{SNP}_{ij} \in \{0, 1, 2\}$ is SNP $j \in \{1, ..., J\}$ of participant i, and ν_{ij} is the associated error in the estimating equation.[4] Since the coefficients of interest $\{\beta_{1j}\}_{j=1}^{J}$ are recovered independently of one another, the second step of the GWAS estimation process involves an adjustment procedure (e.g., LDpred3) that accounts for the conditional effect of SNP_j given the effect of all other SNPs, $\{\text{SNP}_{j'}\}_{j' \neq j}$. Due to the fact that hundreds of thousands of SNPs may be tested in a single GWAS, a SNP is considered statistically significant only if its associated p-value is less than or equal to 5×10^{-8}.

The collection of adjusted coefficients are used to construct polygenic risk scores—indexes for the trait of interest that allow the full set of SNP realizations to be expressed as a single summary statistic. The polygenic score of participant i takes the linear form

$$\text{PGS}_i = \sum_{j=1}^{J} \omega_j \text{SNP}_{ij}, \tag{2}$$

where ω_j is the adjusted weight attached to SNP j in determining the trait. Recent studies have detected significant SNPs—and subsequently constructed polygenic risk scores—for educational attainment (Lee et al., 2018), the Big Five personality types (Lo et al., 2017), ADHD (Demontis et al., 2019, 2023), dyslexia (Doust et al., 2022), and risky behavior (Karlsson Linnér et al., 2019), among many others.

The recent availability of polygenic scores has allowed social scientists to study important questions that incorporate genetic information without having direct access to complete DNA data. In economics, the majority of work that incorporates genetic information focuses on the effects of a polygenic risk score for educational attainment (Barth, Papageorge and Thom, 2020; Houmark, Ronda and Rosholm, 2024; Sanz-de Galdeano and Terskaya, 2022; Cawley et al., 2023). However, recent work has started to adopt polygenic scores of heritable traits and diseases directly (Bertoni, Marin-Lopez and Sanz-de Galdeano, 2023; Jeong et al., 2024).

[4] A refined version of this model includes a small set of control variables, usually age and sex only.

2.2 The Problem with Polygenic Scores

Ultimately, we would like to determine the relative impact of nature, nurture, and their possible joint influence on children's outcomes. Following seminal work in Cunha and Heckman (2007), measures of parental investment are used as proxies for "nurture." The recent advent of large-scale GWAS has provided economists with improved measures of "nature" via the polygenic score. A key assumption in genoeconomics work is that these polygenic scores are unbiased proxy indexes for some latent genetic measures. In this section, we detail why this is unlikely to be the case and how this misspecification may yield biased estimates of genetic treatment effects.

To illustrate our point, and consistent with much of the recent economics literature, we first consider a child i's educational attainment as the outcome of interest (y_i). For simplicity, suppose that educational attainment is produced via an additively separable production function that depends, in part, on nature (genetics, g_i) and nurture (parental investment, I_i). Here, genetics are not proxied by some uni-dimensional score but instead refer to the full vector of SNPs that partially influences y_i. In other words, $g_i \coloneqq \mathbf{SNP}_i = [\text{SNP}_{i,1}, \ldots, \text{SNP}_{i,J}]'$. We write the underlying educational attainment production function as

$$y_i = \sum_{j=1}^{J} \alpha_{1,j} \text{SNP}_{ij} + \alpha_2 I_i + \sum_{j=1}^{J} \alpha_{3,j} \text{SNP}_{ij} I_i + \eta_i, \tag{3}$$

where $\alpha_{1,j}$ captures the direct effect of a particular SNP j on y_i, α_2 captures the direct return to parental investment, and $\alpha_{3,j}$ captures the possible complementarity between SNP j and investment. The final term η_i is the error term. Finally, in the model, parents make endogenous investment decisions in part due to their child's genetic endowments according to some underlying structural equation,

$$I_i = f(\mathbf{SNP}_i), \tag{4}$$

where $f(\cdot)$ is a general function of the child's genetics. Our argument does not rely on any particular parametric form of $f(\cdot)$. As written, we focus on the simplest case without family-specific unobserved heterogeneity. All of our results still hold in this setting, as long as parental investments respond to genetic variation in endowments across siblings.

Recall from Equation 2 that the polygenic score includes the weights ω_j, which are measured as the estimated marginal effect of SNP_j on educational attainment y_i ($\frac{\partial \mathbb{E}[y_i|\text{SNP}_j]}{\partial \text{SNP}_j}$). However, in the above model,

combining Equation 3 and Equation 4, this marginal effect can be derived as

$$\omega_j \equiv \frac{\partial \mathbb{E}[y_i \mid \text{SNP}_j]}{\partial \text{SNP}_j} = \alpha_{1,j} + \alpha_2 \frac{\partial \mathbb{E}[f(\cdot \mid \text{SNP}_j)]}{\partial \text{SNP}_j} + \alpha_{3,j} \mathbb{E}[f(\cdot \mid \text{SNP}_j)] + \sum_{j=1}^{J} \alpha_{3,j} \text{SNP}_{ij} \frac{\partial \mathbb{E}[f(\cdot \mid \text{SNP}_j)]}{\partial \text{SNP}_j}. \tag{5}$$

The right-hand side of Equation 5 shows that the marginal effect of one particular SNP may be determined by more than its direct effect ($\alpha_{1,j}$) alone. If parents endogenously make investment decisions on the basis of their child's genetic endowment, then the effect of SNP_j on the outcome of interest is a function of this behavioral response. Furthermore, if the interplay between nature and nurture is such that increases to parental investment amplify or mitigate the effect of SNP_j ($\alpha_{3,j} \neq 0$), then estimates of ω_j will be further biased from the omission of environmental factors. In particular, $\alpha_{3,j} \neq 0$ suggests that children's environments fundamentally shape the *expression* of their genes, thereby complicating the presumed separation between the effects of nature and nurture (see also Heckman, 2008, on this issue).

Equation 5 also has broader implications for the pitfalls of relying on polygenic scores. Suppose a researcher seeks to estimate the effect of genes on parental investment using the PGS:[5]

$$\begin{aligned} I_i &= \gamma_0 + \gamma_1 \text{PGS}_i + \varepsilon_i && (6) \\ &= \gamma_0 + \gamma_1 \sum_{j=1}^{J} \omega_j \text{SNP}_{ij} + \varepsilon_i && \text{(from Equation 2)} \\ &= \gamma_0 + \gamma_1 \sum_{j=1}^{J} \Psi_j(I_i, \mathbf{SNP}_i; \alpha) \text{SNP}_{ij} + \varepsilon_i\,, && \text{(from Equation 5)} \end{aligned}$$

where, in the last step, we abbreviate the full definition of ω_j from Equation 5 as $\Psi_j(\cdot)$. As shown in Equation 6, estimating this regression model with OLS—even when relying solely on within-family Mendelian randomization across siblings—yields biased estimates of the nurture response to nature. The bias arises from endogeneity due to simultaneity: parental investments influence the PGS measure itself. Put differently, regressing proxies of nurture on the PGS score produces biased estimates because the PGS partly reflects nurture, with parental investment appearing on both sides of the estimating equation. The direction of this bias depends on how investments shape the weights of the polygenic score. This source of bias helps explain why the recent literature on parental responses to children's genetic endowment has

[5] The same reasoning applies to the case of the genetic determinants of ADHD considered later in this paper.

reached mixed conclusions.[6,7,8]

An open question, therefore, is how—if at all—parents adjust their investments in response to their children's genetic endowments. The magnitude and direction of this behavioral response are key to assessing the extent to which polygenic scores provide inadequate measures of children's latent genetic endowments. To address this, the next section describes the data and empirical strategy we employ to estimate the causal effect of children's genetics on parental investment decisions.

3 Data and Empirical Results

3.1 Data

We rely on data from The National Longitudinal Study of Adolescent to Adult Health (Add Health) to conduct our analyses. Add Health ran school-based surveys on a nationally representative sample of 90,118 middle and high school students between 1994 and 1995. A subset of this sample (20,745 adolescents) participated in an additional at-home survey, which formed the longitudinal sample. Approximately 85% of this sample had a parent or guardian participate in a separate parent interview in 1994-1995. These two at-home surveys provide us with detailed information on children's and parents' age, race, gender, education level, and socioeconomic status. Add Health subsequently conducted four additional follow-up surveys on the longitudinal sample in 1996 (Wave II), 2001-2002 (Wave III), 2008-2009 (Wave IV), and 2016-2018 (Wave V). For this study, we focus on the first four waves of Add Health.

Add Health is suitable to address our research question for several reasons. First, the data set includes rich information that allows us to construct longitudinal measures of both latent cognitive skills and latent mental health in children. We use information on students' academic performance in the first two waves to construct measures of latent cognitive skill levels. In order to construct measures of mental health, we leverage the fact that Add Health asks adolescents a battery of questions related to depression and

[6] For instance, Breinholt and Conley (2023) find that parents invest more in children with higher polygenic scores for educational attainment, whereas Sanz-de Galdeano and Terskaya (2022) report the opposite.

[7] In bridging this discussion with our empirical setting of child development in the presence of ADHD, we note that polygenic scores for ADHD have recently become available to researchers. Economists have begun implementing these score into their analyses to determine the effects of ADHD on outcomes like income in adulthood (Rietveld and Patel, 2019) and medical diagnoses (Bertoni, Marin-Lopez and Sanz-de Galdeano, 2023). However, these scores are based on the Demontis et al. (2019) GWAS, which has since been updated by Demontis et al. (2023). Their larger meta-analysis, which pools data from more individuals and has greater power to detect small effect sizes, finds that only half of the previously identified SNPs remain significant. This suggests that the original GWAS is prone to Type I error, potentially biasing the polygenic scores derived from it. Polygenic scores derived from Demontis et al. (2023) are not currently available to researchers.

[8] This potential bias also arises in within-family sibling regression designs, to the extent that parental investments respond to genetic variation in endowments across siblings. In the empirical section below, we demonstrate the empirical relevance of this bias across different model specifications.

depressive tendencies. We use 19 survey questions in both Waves I and II related to depression.[9]

Add Health also offers unique survey data on how often parents participate in various activities with their children, which we use to construct a latent measure of parental investment. We further group these activities into two categories of investment. The first (*activity investment*) includes activities like playing sports, shopping, and going to the movies, while the second (*emotional investment*) includes activities like talking about a romantic partner or a personal problem. We measure all of the aforementioned latent measures (skills, mental health, and parental investment) via the Bartlett factor score, following Heckman, Pinto and Savelyev (2013).[10]

Another benefit of using Add Health is its explicit focus on ADHD. In Wave III, respondents retrospectively report on 18 behavioral indicators of hyperactivity and inattentiveness that they experienced between ages 5 and 12. These questions correspond directly to the 18 diagnostic criteria for ADHD outlined in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5), which physicians use to assess ADHD in children. Appendix C lists these diagnostic criteria in detail.

Perhaps the most attractive feature of the survey for our purposes is the fact that Add Health asked respondents for a saliva sample in Wave IV. Approximately 80% of adult respondents consented to having their genetic information stored and made available to researchers. The standard Add Health dataset includes three genetic measures pertinent to our study: an ADHD polygenic score constructed from the Demontis et al. (2019) genome-wide association meta-analysis, an indicator for whether the respondent has the 7R allele of the *DRD4* gene, and an indicator for the presence of the 10R allele of the *DAT1* gene. These two genes have been found to be associated with ADHD in the genetics literature (Turic, Swanson and Sonuga-Barke, 2010).

In addition, we utilize restricted-access DNA-level data on Wave IV respondents provided by the database of Genotypes and Phenotypes (dbGaP). These data include nearly one million SNP realizations for nearly 10,000 Add Health respondents. From this, we obtain 21 SNPs identified as predictive of ADHD, 19 of which were identified by Demontis et al. (2023), with the remaining two coming from Demontis et al. (2019). We combine the SNP-level genetic data with the retrospective ADHD behavioral data to construct our measure of latent ADHD risk. In practice, we create a Bartlett factor score that incorporates as measures the 18 hyperactive/inattentive tendencies, the 21 SNPs identified in Demontis et al. (2019, 2023), and

[9] Some examples of these questions include: "How often was the following true during the past week or month? You felt depressed." and "How often was the following true during the past week? You felt that you could not shake off the blues, even with help from your family and your friends."

[10] Each latent measure is constructed via exploratory factor analysis, retaining all factors with positive eigenvalues. The first factor—which by construction explains the greatest share of common variance among the constituent indicators—is then used to compute a Bartlett factor score.

the *DRD4* and *DAT1* alleles.

Finally, Add Health also includes two questions related to ADHD diagnosis. The first, asked in Wave III, is: "For which of the following conditions have you taken prescription medication in the past 12 months? Attention problems or ADD or ADHD." The second, asked in Wave IV, is: "Has a doctor, nurse, or other health care provider ever told you that you have or had attention problems or ADD or ADHD?" Although these measures are informative, we do not use them directly to identify ADHD status in children, as they are subject to screening and diagnostic biases.[11] By contrast, our latent ADHD measure, which integrates genetic variation with retrospective reports of inattentive and hyperactive tendencies, captures predisposition independently of access to diagnosis.[12] For validation, we classify respondents as diagnosed if they answer "Yes" to either question, and demonstrate that our latent measure strongly predicts this outcome.[13]

Looking ahead to our model of child development, we seek to include cognitive skills, mental health, and genetics in the multidimensional technology of skill formation. In order to ensure that the stock of each of these inputs has a positive monotonicity interpretation, we negate our mental health and genetic measures. "Mental health" is constructed from survey questions related to depression, so the negated score will indicate lower levels of depression and therefore higher levels of mental health. In a similar fashion, the negated ADHD score will carry a positive interpretation, meaning that children with a high propensity for ADHD will belong to the left of the distribution. For consistency, we also negate the ADHD polygenic score.

Finally, Add Health over-samples siblings. As will be discussed in Section 3.4, the large number of full sibling pairs is imperative for our causal inference strategy. Add Health also provides detailed information about whether sibling pairs are dizygotic (fraternal) or monozygotic (identical) twins, which is important to control for in a model of genetic and environmental interplay.

3.2 Inclusion of Behavioral Phenotypes in the Latent Genetic Risk Measure

A natural concern with our latent measure of genetic risk is that it combines SNP-level genotypes with behavioral phenotypes—the 18 retrospective ADHD symptom reports. One might worry that symptoms incorporate environmental influences and therefore reintroduce the confounding we seek to avoid. We

[11] For example, ADHD diagnoses were much less common in the 1990s than today, making the diagnosed subsample in Add Health an unrepresentative reflection of the underlying population at risk for the disorder. Moreover, Table 1 shows that in the Add Health sample, diagnosed children differ substantially from their peers across several demographic and socio-economic dimensions, including race and family income.

[12] The retrospective design may still introduce recall bias. Adults who were diagnosed with ADHD as children may be more likely to overstate their childhood hyperactivity, while those without a diagnosis may understate such behaviors.

[13] In Section 3.4, we show that our measure has greater predictive power in our Add Health sample for ADHD diagnoses than the available ADHD PGS.

argue that this concern does not undermine our approach. Given current limits of GWAS-based prediction for ADHD, excluding behavioral phenotypes would yield a less informative measure of the underlying genetic liability. We develop this argument along three dimensions.

First, SNPs alone provide a noisy proxy for genetic liability given missing heritability. Twin studies estimate ADHD heritability at roughly 74% (Faraone and Larsson, 2019), yet the set of genome-wide significant SNPs explains only a small share of this heritable variance.[14] In genetics, this gap is known as the missing heritability problem (Manolio et al., 2009). Accordingly, a score constructed solely from the 21 ADHD-associated SNPs available in our data would capture only a narrow component of genetic liability. Symptom reports, by contrast, summarize the realized behavioral expression of ADHD risk and therefore capture polygenic variation that a small set of significant loci cannot. Consistent with this view, our latent measure predicts ADHD diagnosis at roughly three times the rate of the PGS alone (Panel C of Table 4), implying that symptoms add heritable signal beyond the observed SNPs.

Second, our within-family design limits the scope for environmental confounding in symptom-based measures. A natural concern is that self-reported symptoms reflect shared family environment. The sibling fixed-effects design differences out all factors shared by siblings, including family background, household resources, and parental characteristics that are common within the family. Conditional on family fixed effects, residual differences in symptom counts across siblings are less likely to reflect shared environment and more likely to reflect differences in genetic endowment and idiosyncratic influences.[15] The symptom items are retrospective and refer to behaviors between ages 5 and 12, which predates many adolescent shocks and narrows the scope for contemporaneous confounding, though recall bias may remain.[16]

Third, the measurement model provides a disciplined way to combine heterogeneous indicators. In a factor model, an indicator is valid if it is partially determined by the latent trait and measured with error, provided the link between the indicator and the factor is correctly specified. Both SNPs and symptom reports satisfy this condition because each is shaped in part by the same underlying genetic liability for ADHD. The model does not require indicators to be purely genetic. It requires them to be informative about the latent trait. Given the high heritability of ADHD, symptom reports are substantially genetically influenced. Environmental noise in symptoms and the small effects of individual SNPs therefore enter as

[14] Demontis et al. (2023) report a SNP heritability of approximately 0.14.

[15] Because our analytic sample is observed in adolescence, unobserved non-shared environments may still differentially shape behavior and gene expression by the time outcomes are measured. However, this limitation is not specific to the genetic risk measure. It reflects the timing and structure of the data and applies more broadly to measured inputs and outcomes in our setting.

[16] Adults diagnosed with ADHD as children may be more likely to report childhood symptoms, while those without a diagnosis may underreport them. We view this as a limitation of symptom measures in isolation, but one that is mitigated by anchoring the latent factor to objective SNP-level variation.

measurement error, which is exactly what the factor structure is designed to accommodate. The symptom items also map directly to DSM-5 diagnostic criteria (American Psychiatric Association, 2013), which were developed through decades of clinical research to capture the behavioral manifestations most reliably associated with ADHD.

With each of these arguments in mind, we proceed with a latent genetic risk measure that combines SNP-level genotypes and retrospective symptom reports as complementary indicators of ADHD liability.

3.3 Analysis Sample

We restrict our analysis sample to the 6,987 respondents for whom we have valid measures of ADHD risk, parental investment, cognitive skills, and mental health. Table 1 presents descriptive statistics of the analysis sample in the first two columns. In the next two columns, we present the same statistics for the subsample of respondents that have a full sibling in Add Health, including twins and non-twins. The siblings sample comprises 23% of the overall sample. The final two columns restrict to those with an ADHD diagnosis on record (5% of the analysis sample). In Table A2 through Table A5, we additionally report summary statistics about the measures used for each latent factor across these three groups.

Panel A reports demographic characteristics of the respondents. In comparing the different segments of the sample, several patterns emerge. Despite comprising approximately half of the total sample, female respondents only make up 38% of the diagnosed sample. This is consistent with national trends that show boys are two times as likely to receive an ADHD diagnosis compared to girls. The diagnosed sample also contains a greater proportion of white respondents (84%) compared to the overall sample (62%). The diagnosed sample also reports slightly lower GPA's compared to respondents not diagnosed with ADHD.

Panel B provides summary statistics for respondents' mothers. We find little discrepancy between the analysis sample and the sibling subsample, but again note differences in the ADHD group. Mothers with diagnosed children report household incomes that are 9% higher than the analysis sample average. These same parents are also more likely to report that they invest in their child, as measured by both time and emotional investment factors. We find little evidence of differences in the propensity for these mothers to be married or highly educated.

Lastly, Panel C reports average characteristics related to respondents' health, including measures of mental health, genetics, and behavioral manifestations of ADHD. A little more than 5% of the overall sample has an ADHD diagnosis, which is close to national trends between 1997 and 2008 (Boyle et al., 2011). Those with a diagnosis report significantly higher counts of hyperactive and inattentive tendencies. These respondents have ADHD polygenic scores that are 14% of a standard deviation (0.14σ) above average.

Table 1: Descriptive Statistics

	Analysis Sample		Siblings Sample		Diagnosed Sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Panel A: Child Demographics						
Female	0.541	0.498	0.525	0.500	0.376	0.485
White	0.696	0.460	0.708	0.455	0.903	0.296
Black	0.213	0.410	0.189	0.392	0.079	0.264
Age	15.50	1.708	15.58	1.661	15.43	1.648
Twin	0.092	0.289	0.401	0.490	0.091	0.289
GPA	2.799	0.745	2.831	0.753	2.522	0.738
Panel B: Mother Demographics						
Household Income ($1,000's)	47.42	50.03	47.23	60.30	51.84	46.91
Investment	−0.000	1.295	0.008	1.269	0.023	1.360
Investment (Activity)	−0.000	1.675	0.027	1.618	0.025	1.783
Investment (Emotional)	0.000	1.367	−0.011	1.359	0.029	1.426
Married	0.745	0.436	0.754	0.431	0.753	0.432
College Educated	0.242	0.428	0.218	0.413	0.241	0.428
Panel C: Health Characteristics						
Mental Health	0.000	1.059	−0.043	1.074	−0.148	1.169
ADHD Diagnosis	0.053	0.225	0.052	0.221	1.000	0.000
Hyperactivity Count	1.734	1.972	1.692	1.903	3.780	2.641
Inattentiveness Count	1.200	1.887	1.244	1.924	3.626	2.797
ADHD PGS	0.022	1.000	0.052	1.016	−0.143	1.033
Latent ADHD Factor	0.000	1.047	−0.010	1.047	−1.271	1.337
Observations	6,987		1,608		372	

Notes: The above table presents summary statistics for different sub-samples of the Add Health survey. The analysis sample refers to those who agreed to take part in the Wave IV DNA sample. We maintain this restriction in the next two groupings as well. The sibling sample refers to those with at least one full sibling linkage or one twin linkage in Add Health. The diagnosed sample includes adolescents that are coded as having ADHD through at least one of the following ways: taken medication for ADHD in the past 12 months (Wave III) or have had a doctor say they have ADHD (Wave IV). Hyperactivity and impulsivity counts come from retrospective survey responses about respondents' own behavior during childhood. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Listed observations refer to the greatest possible sample size. Some measures, like those in Panel B, have fewer observations due to non-response in the Add Health.

Similarly, those with a diagnosis have a much higher propensity for ADHD and depression, as measured by our latent factor scores for each.

3.4 Empirical Results

ADHD is linked to a number of negative outcomes for children in both the short- and long-run. Recent work inside and outside of economics have found that children diagnosed with ADHD are more likely to exit school before age 16 (Fleming et al., 2017), less likely to obtain post-secondary education (Loe and Feldman, 2007), more likely to incur unemployment (Rietveld and Patel, 2019), and more likely to commit crime in adulthood (Fletcher and Wolfe, 2009). However, establishing a causal link between ADHD and any of these outcomes is difficult for several reasons, which we detail below.

First, ADHD is a latent characteristic that exists on a spectrum in individuals. It is "identified" by physician diagnoses, but not all children are screened for ADHD and not all physicians diagnose identically (Marquardt, 2022). These biases make ADHD diagnosis unsuitable as a pure measure of latent ADHD risk. In fact, analyses that rely on ADHD diagnosis status will recover the treatment effects induced by a diagnosis, rather than the direct, causal effect of ADHD. For this reason, we construct our own measure of ADHD risk using children's genotypes (relevant SNPs and genes) and phenotypes (ADHD behavioral tendencies). We benchmark analyses that use this latent factor score against analyses that instead use the Demontis et al. (2019) polygenic score to test for discrepancies in results stemming from differences in these two measures.

Finally, given our focus on the gene–environment interplay—where children's environments may shape the expression of their genes, and genetic predispositions may in turn influence parents' investment decisions—we estimate how parental investments respond to genetic factors and run regressions of children's adult outcomes to assess the long-run effects of genetic endowments during adolescence.

For our first set of motivating empirical results, we begin our analysis by first estimating the following linear model using ordinary least squares (OLS),

$$y_{if} = \alpha g_i + \boldsymbol{\beta}' \boldsymbol{x}_i + \varepsilon_i, \tag{7}$$

where y_{if} denotes the outcome for adolescent i who belongs to family f, g_i is the adolescent's latent (negated) risk for ADHD, $\boldsymbol{x}_i$ is a vector of controls—including age, gender, race, twin status, maternal education, and, most importantly, our measures of latent cognitive skill and latent mental health—and ε_i is the error term.

Our parameter of interest is α, the conditional association between children's genetic risk for ADHD and a given outcome. To estimate this relationship, we control for latent cognitive and non-cognitive (in our framework, mental health) skills, following the literature that underscores the importance of accounting for these latent factors in shaping both child development and parental investment decisions (Cunha and Heckman, 2007; Cunha, Heckman and Schennach, 2010; Agostinelli and Wiswall, 2025). When parental investment (I_i) is used as an outcome variable (y_i), the specification reduces to a reduced-form representation consistent with the investment equation commonly employed in the child development literature, which relates parental investment to latent measures of children's skills. Our contribution is to extend this framework by explicitly incorporating genetic risk. In addition, the pooled specification leverages variation across both siblings and unrelated individuals, thereby improving estimation precision relative to

Table 2: Empirical Results for Parental Investment using Latent Genetic Risk for ADHD

	(1)	(2)	(3)	(4)	(5)
Panel A: Pooled OLS					
Latent Genetic Factor	0.001	−0.002	−0.035	−0.035	−0.036
	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
Observations	6,987	6,987	6,987	6,987	6,987
Panel B: Within-Family					
Latent Genetic Factor	−0.049	−0.047	−0.081	−0.081	—
	(0.044)	(0.044)	(0.047)	(0.047)	
Observations	1,608	1,608	1,608	1,608	
Cognitive Control	✓	✓	✓	✓	✓
Mental Health Control	✗	✓	✓	✓	✓
Demographic Controls	✗	✗	✓	✓	✓
Twin Control	✗	✗	✗	✓	✓
Maternal Education Control	✗	✗	✗	✗	✓

Notes: This table presents within-family fixed effects regression results with latent parental investment as the dependent variable and our latent measure of (negated) ADHD risk as the independent variable of interest. "Cognitive Control" refers to the latent factor for cognitive skills, measured in Wave I. "Mental Health Control" refers to the latent factor for depression (mental health), measured in Wave I. "Demographic Controls" include gender, age, and a quadratic term for age. "Twin Control" includes an indicator for whether the respondent is a twin. "Maternal Education Control" is an indicator for whether the child's mother is college-educated. Within-family specifications omit race and maternal education due to multicollinearity. Robust standard errors are reported in Panel A. Standard errors are clustered at the family level in Panel B.

the more restrictive within-family specification introduced later. This precision gain is particularly valuable given the small effect sizes typically observed in genome-wide association studies (GWAS). However, although these controls help mitigate potential confounding, we remain cautious in assigning a causal interpretation to the parameter α.

Panel A of Table 2 reports OLS estimates of the pooled specification in Equation 7, using parental investment as the dependent variable. We iteratively introduce controls across columns (1)–(5) to examine how the coefficient on genetic risk evolves with the sequential inclusion of controls. Focusing on column (5), which includes the full vector of controls, we find that parents increase investment by roughly 0.03 standard deviations in a child with a one-standard deviation higher risk of ADHD (p-value=0.023).[17] This association is consistent with parents using endogenous investments as a compensatory response to higher genetic risk for ADHD.

Although these pooled OLS estimates are informative about the association between parental investment and genetic risk, our ultimate goal is to obtain a causal relationship. As discussed in Section 3.2, a

[17] We report coefficients using the latent investment factor, which is not normalized to have unit variance. Rescaling to a one-standard-deviation change in investment yields an effect of −0.027. Furthermore, because we negate our measure of genetic risk, higher values of g_i correspond to lower ADHD risk, so a positive coefficient indicates greater investment in children at lower genetic risk.

meta-analysis of twin studies shows that ADHD has a heritability rate of 74% on average (Faraone and Larsson, 2019), meaning that nearly three-fourths of the variation in ADHD is attributable to differences in genetic realizations across individuals. It is therefore important to control for parental genetic influence when determining a causal link between ADHD and some outcome of interest. There are two main strategies employed in the literature to control for parental genetic influence, both of which require genetic information on full-sibling pairs. In the first strategy, researchers use available information on the population-wide distribution of genes at a certain base-pair location and the realization of these genes between sibling pairs to impute the parental gene pool in a process called "Mendelian imputation of parental genotypes" (Young et al., 2022). In the second, researchers utilize a within-family design that differences out all environmental-specific fixed effects shared between siblings, including parental genotypes and ADHD heritability. Carvalho (2025) shows that these two strategies deliver very similar estimates, suggesting that our choice between them is not consequential. Given this result, we rely on the within-family approach to conduct our analysis.

We estimate within-family regressions that leverage the random allocation of genetic endowments across siblings. As shown in Figure 1, conditioning on parental genotypes ensures that the residual variation in children's genetic risk for ADHD is driven solely by Mendelian randomization. The trade-off, however, is a substantial reduction in sample size, which lowers the statistical precision of some estimates. Restricting the analysis to linked full siblings in Add Health, we estimate the following within-family specification:

$$y_{if} = \gamma g_i + \boldsymbol{\beta}' \boldsymbol{x}_i + \phi_f + \varepsilon_{if}, \tag{8}$$

where y_{if} is some outcome for adolescent i who belongs to family f, ϕ_f is a family-specific fixed effect, ε_{if} is the error term, and the controls are as before.[18] Our parameter of interest is γ, which we interpret as the causal effect of a one-standard deviation reduction in ADHD risk on the outcome, under the assumption that unobserved confounders linked to genetic variation are constant within families (across siblings).

Panel B of Table 2 reports the within-family estimates of Equation 8, again using parental investment as the dependent variable. In the most saturated specification, we find that parents allocate investments differentially across siblings based on their relative genetic risk for ADHD. We find that parents invest more in children with greater latent risk for ADHD, suggesting a compensatory response to this risk. Although only marginally significant, our point estimates indicate that parents reduce investment by roughly 0.062 standard deviations in a child whose sibling has a one-standard deviation higher risk of ADHD (p-

[18] Because we focus on full siblings, the within-family model absorbs shared demographic attributes. We therefore omit race and maternal education but otherwise retain all previously listed controls.

Table 3: Empirical Results for Parental Investment using ADHD PGS

	(1)	(2)	(3)	(4)	(5)
Panel A: Pooled OLS					
PGS	0.040	0.040	0.042	0.042	0.035
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Observations	6,987	6,987	6,987	6,987	6,987
Panel B: Within-Family					
PGS	−0.007	−0.008	−0.003	−0.003	—
	(0.070)	(0.070)	(0.071)	(0.071)	
Observations	1,608	1,608	1,608	1,608	
Cognitive Control	✓	✓	✓	✓	✓
Mental Health Control	✗	✓	✓	✓	✓
Demographic Controls	✗	✗	✓	✓	✓
Twin Control	✗	✗	✗	✓	✓
Maternal Education Control	✗	✗	✗	✗	✓

Notes: This table presents within-family fixed effects regression results with latent parental investment as the dependent variable and the (negated) ADHD PGS from Demontis et al. (2019) as the independent variable of interest. "Cognitive Control" refers to the latent factor for cognitive skills, measured in Wave I. "Mental Health Control" refers to the latent factor for depression (mental health), measured in Wave I. "Demographic Controls" include gender, age, and a quadratic term for age. "Twin Control" includes an indicator for whether the respondent is a twin. "Maternal Education Control" is an indicator for whether the child's mother is college-educated. Within-family specifications omit race and maternal education due to multicollinearity. Robust standard errors are presented in Panel A. Standard errors are clustered at the family level in Panel B.

value=0.080).[19] Importantly, Table A6 demonstrates that this compensatory pattern is robust to alternative choices in constructing the latent factor.

In comparing the two approaches, we find that the within-family coefficient is nearly double the pooled estimate. One possible explanation is that the pooled model combines both within- and between-family variation, and the latter can attenuate the estimated relationship if unobserved family characteristics are correlated with both genetic risk and parental investment. By conditioning on family fixed effects, the within-family design removes these unobserved family-level confounders and identifies the underlying compensatory response. This difference highlights the standard bias–variance trade-off: pooled OLS yields more precise estimates but is likely biased by unobserved heterogeneity, whereas within-family models reduce bias at the cost of higher variance, driven by smaller sample sizes (77% smaller in our case).

Regardless of specification, the evidence points to compensatory behavior by parents in response to increases in children's genetic risk for ADHD. As discussed above, such behavioral responses of nurture to nature complicate the interpretation of PGS as a "pure" measure of genetic endowments. To address this concern, we next examine whether our findings replicate when we substitute our latent genetic factor with

[19] As in Panel A, the reported coefficients use the latent investment factor, which is not normalized to have unit variance. Rescaling to a one-standard-deviation change in investment yields the corresponding standardized effect.

the conventional ADHD polygenic score (PGS) from Demontis et al. (2019). Specifically, we re-estimate Equation 7 and Equation 8, replacing g_i with the (negated) PGS. The results are reported in Table 3.

We reverse our findings under the PGS-based model. In the pooled specification, a one-standard deviation decrease in the PGS is associated with a 0.03 standard deviation *increase* in parental investment (p-value=0.024).[20] This sign reversal implies the opposite interpretation: rather than compensating for higher genetic risk, parents appear to reinforce existing differences in genetic endowment. By contrast, the within-family specification shows no systematic relationship between genetic risk and investment, suggesting that parents do not adjust investments differentially across siblings on the basis of genetic risk. Following our discussion of simultaneity bias induced by estimating investment equations with the PGS (Section 2.2), this result is in line with the hypothesis that information about parental investment is included in the PGS and attenuates the corresponding point estimates of genetic effects. This is of concern because researchers may reach diametrically opposed conclusions about the role of parental investment depending solely on which measure of genetic risk they adopt.[21]

In our final reduced-form empirical exercise, we consider the longer-run effects of ADHD. Following the literature on the impacts of ADHD in adulthood, we consider three classes of outcomes: labor, social, and health. Our labor market outcomes include the future stock of human capital (as measured by our latent factor for cognitive skills), number of years of education, employment, and log income. We include as social outcomes an indicator for being married in adulthood and an indicator for ever being arrested. For health outcomes, we consider an indicator for realized ADHD diagnosis status, the future stock of mental health (again as measured by our negated latent factor for depression), and illicit drug use. We measure drug use with a Bartlett factor score that uses adolescents' self-reported recent usage of illicit drugs as measures. We show estimates of these long-run, within-family equations in Table 4.

As before, we consider both our own latent ADHD score and the PGS as separate proxies for underlying risk. We do not find any statistically significant impact of ADHD on labor outcomes using either measure. However, siblings with greater latent risk for ADHD are more likely to ever be arrested—a result that does not replicate with the PGS. Both measures generate sensible point estimates for the probability of being diagnosed with ADHD, though the PGS effect size is attenuated to roughly one-third of the estimate obtained using our latent measure. Finally, siblings with higher latent genetic risk are more likely to use illicit drugs, a pattern again not replicated with the PGS.

[20] As with the latent factor measure, we convert the reported coefficients into standard deviation units by rescaling by the standard deviation of the latent investment measure.

[21] Table A10 presents similar results for overall investment, activity-based investment, and emotional investment by genetic measure. See Section 3.1 for details on the construction of these measures.

Table 4: Within-Family Model Estimates of ADHD on Long-Run Outcomes

Panel A: Education/Labor Measures								
	Future Skills		Education (Years)		Employed		(Log) Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latent Factor Score	0.018		0.079		0.006		0.096	
	(0.034)		(0.073)		(0.012)		(0.120)	
Polygenic Score		−0.024		0.112		0.008		0.033
		(0.044)		(0.106)		(0.017)		(0.178)
Observations	1,315		1,608		1,602		1,556	
Panel B: Social Measures								
	Ever Married		Ever Arrested					
	(1)	(2)	(3)	(4)				
Latent Factor Score	0.019		−0.043					
	(0.021)		(0.018)					
Polygenic Score		−0.005		0.015				
		(0.026)		(0.024)				
Observations	1,607		1,604					
Panel C: Health Measures								
	ADHD Diagnosis		Future Mental Health		Illicit Drug Use			
	(1)	(2)	(3)	(4)	(5)	(6)		
Latent Factor Score	−0.068		0.038		−0.105			
	(0.012)		(0.037)		(0.049)			
Polygenic Score		−0.023		0.077		0.048		
		(0.012)		(0.052)		(0.058)		
Observations	1,608		1,508		1,608			

Notes: Each regression includes controls for child age, age squared, gender, initial skills, initial mental health, and twin status. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Robust standard errors clustered at the family level are presented in parentheses. Future skills and future depression are measured in the Wave II, illicit drug use is a latent factor score that uses multiple measures of illicit drug use in Wave III, and all other outcomes are measured by Wave IV.

Although our estimates provide suggestive evidence on the long-run impacts of ADHD genetic risk and outcomes, they should be interpreted with caution. As our previous results indicate, genetic influences that are apparent in adolescence may diminish over time, particularly if families actively respond to early differences. Parents may compensate for or even fully offset these disparities through targeted investments, making long-run outcomes less directly attributable to genetic variation. In this sense, reduced-form estimates risk overstating persistent genetic effects, as they conflate initial endowments with the successful efforts of parents to narrow developmental gaps. To more carefully disentangle these forces, in Section 4 we develop a structural framework that embeds genetic risk within the broader process of skill formation and explicitly models the endogenous parental response.

4 A Dynamic Model of Child Development

4.1 Environment and Initial Conditions

Our goal is to develop a model that explains how children's skills evolve under the joint influence of genetic endowments (nature) and parental investments (nurture). We extend the prior literature on structural models of parental investments and skill formation by explicitly incorporating genetics (Del Boca, Flinn and Wiswall, 2014; Caucutt and Lochner, 2020; Caucutt et al., 2020; Agostinelli et al., 2024; Mullins, 2025). The model features families consisting of one parent and two siblings. Although parental investment decisions are assumed to be independent across siblings, the siblings are connected through a shared unobserved family-level factor and through correlated skill and genetic endowments. The model tracks skill development throughout adolescence, from $t = 1$ to $t = T = 6$, corresponding to grades 7 through 12.[22]

In a similar fashion to Cunha, Heckman and Schennach (2010), we consider skills as multidimensional. Denote by $\theta_{i,t} = (\theta^C_{i,t}, \theta^H_{i,t})$ the stock of cognitive skills and mental health held by child i in period t. Each adolescent is also endowed with an immutable genetic stock, g_i. Cognitive skills, mental health, and genetics are all defined as "goods," meaning that an increase in the stock of any attribute is assumed to be beneficial to the child. Thus, adolescents with high levels of depression have low levels of $\theta^H_{i,t}$ and those with a high risk for ADHD have a low endowment of g_i. Finally, parents in the model are able to influence the development of skills via period-specific endogenous investment decisions, $I_{i,t}$.

Consistent with our empirical analysis, we allow for heterogeneity by race (r) and gender (s). We also incorporate unobserved family-level heterogeneity, ψ, which is shared by siblings i and j. The initial conditions are specified as a four-dimensional joint distribution that may vary across (r, s) groups:

$$\begin{bmatrix} \ln \theta^C_0 \\ \ln \theta^H_0 \\ \ln g \\ \ln \psi \end{bmatrix} \Big| \, (r, s) \sim \mathcal{N}(\mu_{r,s}, \Sigma_{r,s}). \tag{9}$$

This formulation captures systematic differences across race-gender groups while allowing for unobserved correlation structures in accumulated skills at the outset of the model.

[22] The choice of this age range is not essential to the theoretical framework but reflects the availability of data, which is collected during these years of the child's schooling.

4.2 The Technology of Skill Formation

With the initial conditions established, we are ready to develop a theory for how skills develop across adolescence. We follow Agostinelli and Wiswall (2025) in adopting a trans-log specification of the skill technology, which provides flexibility in the sign of static complementarity between each of the included inputs. For a skill type $k \in \{C, H\}$, we write

$$\begin{aligned} \ln \theta^k_{i,t+1} = \ln A^k_{f(i)} + \alpha^k_1 \ln \theta^C_{i,t} + \alpha^k_2 \ln \theta^H_{i,t} + \alpha^k_3 \ln g_i + \alpha^k_4 \ln I_{i,t} \\ + \alpha^k_5 (\ln \theta^k_{i,t} \cdot \ln g_i) + \alpha^k_6 (\ln \theta^k_{i,t} \cdot \ln I_{i,t}) + \alpha^k_7 (\ln g_i \cdot \ln I_{i,t}). \end{aligned} \tag{10}$$

The first line in Equation 10 captures the direct effect of each input on the future production of skill k. For example, $\alpha^k_1 > 0$ suggests that $\ln \theta^C_{i,t}$ is a productive input and leads to increases in the future stock of skill k. The second line in the technology equation captures possible input complementarity. For instance, $\alpha^k_5 > 0$ indicates that the marginal return to skill k increases with g. Similarly, a positive value of α^k_7 reflects gene $\times$ environment interplay, indicating that the effectiveness of parental investment depends on the child's genetic endowment.

The term $\ln A^k_{f(i)}$ denotes the total factor productivity (TFP) of the skill-formation technology, which is skill-type specific (k) and depends on the family environment $f(i)$. We parameterize TFP as a function of unobserved family-level heterogeneity:

$$\ln A^k_{f(i)} = \alpha^k_{A,0} + \alpha^k_{A,1} \ln \psi_{f(i)}. \tag{11}$$

Since this heterogeneity is shared by siblings ($\psi_{f(i)} = \psi_{f(j)}$), TFP is constant within families; that is, $A^k_{f(i)} = A^k_{f(j)}$.

4.3 The Parent's Problem

4.3.1 Utility

We assume that parents in our model are paternalistic, meaning they derive utility directly from their children's skills. Parents face a trade-off between the benefits of having more skilled children, which requires higher levels of investment through the skill formation technology, and the costs associated with that investment. We specify the parent's period-specific utility as

$$u_{i,t} = \gamma_1 \ln \theta^C_{i,t} + \gamma_2 \ln \theta^H_{i,t} + \gamma_3 I_{i,t} + \gamma_4 I^2_{i,t} \ln \psi_{f(i)}. \tag{12}$$

The parameters γ_1 and γ_2 reflect the relative importance that parents place on cognitive skills and mental health. The term $\gamma_3 < 0$ captures the disutility that parents experience from increasing investment in their child's development. Lastly, γ_4 accounts for potentially convex investment costs and allows for heterogeneity in the disutility of costly investment through the interaction with $\ln \psi_{f(i)}$. In other words, this term captures family-level unobserved heterogeneity in the determinants of parental investments, which also enters the technology of skill formation.

We highlight a key feature of our utility specification that greatly simplifies both solving and estimating the model. As specified, our framework is equivalent to one in which parents evaluate each child independently. Specifically, for parents with multiple children i and j, utility is additively separable in the skill levels of each child ($u_{i,t}$ and $u_{j,t}$). This implies that parents choose child-specific investments independently across siblings. Introducing within-family complementarities in parental investments across siblings would substantially increase computational complexity. Although understanding how genetics interact with family-wide investment decisions and family composition is a promising direction for future work, we believe our approach is the right first step for studying the measurement of genetic endowments and their role in shaping conclusions about gene–environment interactions. In our model, siblings are linked only through shared family-level heterogeneity (ψ), which induces correlation in their skill endowments, but are otherwise treated as independent.

4.3.2 Recursive Formulation of the Parent's Problem

Parents choose an optimal investment level for child i in each period t. The technology of skill formation links skills in period $t+1$ to the stock of skills in period t, parental investments in period t, and a child's genetic endowment. In other words, the technology of skill formation defines how a child's skills evolve over time due to current endowments and parental choices. We can therefore write the parent's investment problem recursively,

$$V_t(\theta^C_{i,t}, \theta^H_{i,t} \mid \psi_{f(i)}) = \max_{I_{i,t} \geq 0} \left\{ u\left(I_{i,t}, \theta^C_{i,t}, \theta^H_{i,t} \mid \psi_{f(i)}\right) + \beta V_t(\theta^C_{i,t+1}, \theta^H_{i,t+1} \mid \psi_{f(i)}) \right\}, \tag{13}$$

where $V(\cdot)$ is the value function with cognitive skills and mental health as endogenous state variables, while $u(\cdot)$ is the period utility (Equation 12).

We follow Del Boca, Flinn and Wiswall (2014) in writing the parent's continuation utility after their child finishes high school as

$$V_T(\theta^C_{i,T}, \theta^H_{i,T}) = \gamma_1 \ln \theta^C_{i,T} + \gamma_2 \ln \theta^H_{i,T}, \tag{14}$$

which captures the continuation value beyond the model's time horizon, reflecting parents' perspective on their child's development.

Adult outcomes for child i are thus taken as outside the scope of the model. Taken together, this framework provides a parsimonious yet flexible way to study how parents respond to differences in children's skills and genetic endowments, and how those responses shape the development of human capital during adolescence.

5 Estimation

We estimate our model using indirect inference, specifically through the Simulated Method of Moments (SMM). In practice, this involves replicating and matching a targeted set of moments from the Add Health data within the model's simulations. In what follows, we present the additional parametric restrictions (Section 5.1) and the identifying assumptions (Section 5.2), describe the moments targeted in the SMM procedure (Section 5.3), report the parameter estimates (Section 5.4), and assess the model's fit to the data (Section 5.5).

5.1 Additional Parameterization

We now detail the normalizations and parameter restrictions imposed on the parametric distribution of the initial conditions during the estimation of the model. As discussed in Equation 9, skills, genetics, and family-level heterogeneity are jointly distributed with a 4-by-1 mean vector $\mu_{r,s}$ and a 4-by-4 covariance matrix $\Sigma_{r,s}$, both varying by race and gender. In total, this specification entails 56 free parameters. To maintain tractability in estimation, we impose additional restrictions that reduce the dimensionality of the parameter space.

The first three components of the mean vector in Equation 9, $\mu_{r,s} = [\mu_{C,r,s}, \mu_{H,r,s}, \mu_{g,r,s}, \mu_{\psi,r,s}]$, are obtained directly from the data as the group-specific conditional means of the measured initial conditions for (log) cognitive skills, mental health, and genetics (reported in Panel A of Table A15). The fourth component, corresponding to the mean of (log) family-level heterogeneity ($\ln \psi$), is estimated freely by (r, s) group within the SMM procedure. For identification, we normalize this mean to zero for Black female children, who serve as the reference group. Consequently, the estimated means of family-level heterogeneity for all other (r, s) groups are expressed relative to this benchmark. As an additional normalization, we set the variance of family-level heterogeneity to one ($\sigma_{\psi}^{2,(r,s)} = 1$). We further restrict this unit variance to be identical across all (r, s) groups, which saves three parameters relative to the alternative approach

of normalizing variance for only one reference group. Given our normalizations and assumption on the variance of $\ln \psi$, the covariance matrix takes the following form,

$$\Sigma_{r,s} = \begin{bmatrix} \sigma_C^{2,(r,s)} & \sigma_{C,H}^{(r,s)} & \sigma_{C,g}^{(r,s)} & \sigma_{C,\psi}^{(r,s)} \\ \sigma_{C,H}^{(r,s)} & \sigma_H^{2,(r,s)} & \sigma_{H,g}^{(r,s)} & \sigma_{H,\psi}^{(r,s)} \\ \sigma_{C,g}^{(r,s)} & \sigma_{H,g}^{(r,s)} & \sigma_g^{2,(r,s)} & \sigma_{g,\psi}^{(r,s)} \\ \sigma_{C,\psi}^{(r,s)} & \sigma_{H,\psi}^{(r,s)} & \sigma_{g,\psi}^{(r,s)} & 1 \end{bmatrix}. \tag{15}$$

The leading 3×3 block of $\Sigma_{r,s}$ corresponding to the variance covariance matrix of $(\ln C_0,\ \ln H_0,\ \ln g)$ is estimated directly from the data for each (r, s) type.

The covariances involving the family-level unobserved heterogeneity $\ln \psi$ are estimated within the SMM procedure. We assume that the correlations between $\ln \psi$ and each of the other variables are invariant across groups, $\rho_{C,\psi},\ \rho_{H,\psi},\ \rho_{g,\psi}$ are constant over (r, s). This restriction implies that, rather than estimating three separate covariances for each of the four (r, s) groups, we estimate only three common correlation parameters. As a result, the number of free parameters falls by nine, reducing the dimensionality of the estimation problem and improving tractability. Importantly, this does not mean that the covariances between family-level unobserved heterogeneity and a child's initial endowments of skills and genes are constant across groups; instead, it constrains their structure as follows:

$$\sigma_{C,\psi}^{(r,s)} = \rho_{C,\psi} \cdot \sigma_C^{(r,s)} \cdot \sigma_\psi^{(r,s)}, \qquad \sigma_{H,\psi}^{(r,s)} = \rho_{H,\psi} \cdot \sigma_H^{(r,s)} \cdot \sigma_\psi^{(r,s)}, \qquad \sigma_{g,\psi}^{(r,s)} = \rho_{g,\psi} \cdot \sigma_g^{(r,s)} \cdot \sigma_\psi^{(r,s)},$$

where $\sigma_X^{(r,s)} \equiv \sqrt{\sigma_X^{2\,(r,s)}}$ denotes the standard deviation of $X \in \{C, H, g, \psi\}$ in group (r, s), and $\sigma_{X,\psi}^{(r,s)}$ the corresponding covariance. Finally, we fix the discount factor in Equation 13 at $\beta = 0.96$. All other parameters introduced in Section 4 are left free for estimation. We now turn to the details of our identification strategy.

5.2 Identification

We model parental investment as an endogenous response to children's cognitive ability, mental health, and genetic endowment. To identify the structural parameters, we exploit multiple sources of variation, including the random allocation of genetic endowments across siblings and the longitudinal structure of the Add Health data, which provides repeated measures of skills and parental investment during adolescence. In the discussion that follows, we provide a heuristic account of identification, which relies on the joint restrictions imposed by the full set of moments, rather than a one-to-one mapping between moments

and parameters.

One source of identification comes from within-family differences in genetic endowment. Full siblings inherit genetic material randomly from their parents, conditional on the parental genotype. This biological process generates quasi-random variation in genetic endowment across siblings, which we treat as exogenous conditional on family background and initial skill levels.[23] By exploiting these sibling-level differences, we hold constant all shared family and environmental unobserved heterogeneity, allowing us to isolate the impact of genetic endowments on parental investment and adolescent outcomes.

Bringing this logic into our empirical framework, we account for unobserved family-level heterogeneity by estimating within-family fixed-effects regressions, which exploit within-family Mendelian randomization of genetic endowments, and using the resulting auxiliary coefficients as targets in our SMM procedure. This variation is particularly informative for identifying parental preferences, behavioral responses to genetic variation, and differences in skill endowments.

We complement the within-family variation with pooled variation, captured through auxiliary regression coefficients that incorporate differences across families. These moments help match broader population patterns. Although they do not permit a causal interpretation, they are particularly valuable for uncovering the role of family-level unobserved heterogeneity in shaping parental investments, as well as for highlighting components of the skill production function that warrant stronger empirical support, such as interaction effects and nonlinearities. By targeting both within-family and pooled moments, we combine credible identification of causal responses with an accurate replication of population-level relationships between investments, skills, and endowments.

We allow for heterogeneity in initial conditions by modeling the joint distribution of cognitive skills, mental health, genetic endowments, and family-level unobserved heterogeneity. This distribution is permitted to vary by children's race and gender. The initial conditions for cognitive skills, mental health, and genetic endowments are directly identified in the data through the mean and variance–covariance structure of the observed measures. By contrast, heterogeneity in parental investments across families, together with its cross-sectional correlation with children's endowments, identifies the parameters governing the distribution of family-level heterogeneity and its structural correlation with children's endowments, both across and within sibling pairs.

Finally, we identify the parameters of the skill production function by leveraging variation in the dynamic evolution of cognitive and mental health skills over time, as a function of the current stock of cognitive skills, mental health, genetic endowments, and parental investments. By jointly exploiting these

[23] See Section 2.1 for a detailed description of this inheritance mechanism.

dynamic relationships—both across families and within families—we are able to recover the parameters that govern the process of human capital accumulation.

5.3 Targeted Moments

First, we estimate outside the model a subset of the initial conditions—the joint distribution of cognitive skills, mental health, and genetic endowment—outside the model and separately for each of the four race-by-gender subgroups. This step simplifies the estimation process for the remaining structural parameters.[24] Each latent skill is constructed using Bartlett factor scores, following Heckman, Pinto and Savelyev (2013), which combine multiple noisy indicators into an unbiased predictor. We apply the same procedure to build a latent measure of parental investment. Initial skills are measured at 7th grade, which we set as $t = 0$ in the model.

For the SMM estimation, we target a set of estimated coefficients from several auxiliary regression models. Table 5 reports the full set of regression-based moments used in the procedure, organized into four conceptual blocks that mirror the structure of the model. The first block captures the age profile of skill accumulation in childhood by pooling regressions of cognitive and mental health outcomes on age, thereby identifying the parameters that govern skill development during adolescence. The second and third blocks consist of pooled and within-family regressions of parental investment on child skill and genetic endowments. These moments identify parental preferences for each child, the cost of investment as specified in the utility function (Equation 12), and the structural correlation between family-level unobserved heterogeneity and children's skills and genetic endowments. In particular, the comparison between pooled and within-family regression results on parental investments highlights the extent of bias induced by endogeneity—arising from unobserved family-level heterogeneity and its correlation with child skills and genetic endowments—which in turn informs the model's identification of this structural correlation. The fourth block includes reduced-form regressions for skill formation—estimated in both pooled and within-family specifications—which provide information on the skill production technology, including complementarities and interaction effects between inputs.

We also incorporate moments that capture heterogeneity in parental responses across demographic groups. Specifically, we estimate regressions of investment on race, gender, and their interaction to identify mean differences in family-level unobserved heterogeneity, which translate into heterogeneity across subgroups in the preference cost of parental investments.

In simulations, each child i is paired with a unique sibling j, and both share family-level heterogeneity,

[24] The computation of the standard errors accounts for the parameter uncertainty introduced by this first-step procedure.

Table 5: Main Sources of Identification for Model Parameters

Model Characteristic	Regression Specification
I. Evolution of Skills Across Adolescence	$\theta_i^k = \beta_0^k + \beta_1^k t + \varepsilon_i^k$
II. Endogenous Investment Decision	$I_{i,t} = \beta_0 + \beta_1 \ln \theta_{i,t}^C + \beta_2 \ln \theta_{i,t}^H + \beta_3 \ln g_i + \varepsilon_{i,t}$ $I_{i,f,t} = \beta_0^F + \beta_1^F \ln \theta_{i,t}^C + \beta_2^F \ln \theta_{i,t}^H + \beta_3^F \ln g_i + \phi_f + \varepsilon_{i,f,t}$
III. Technology of Skill Formation	$\ln \theta_{i,t+1}^k = \beta_0^{\theta,k} + \beta_1^{\theta,k} \ln \theta_{i,t}^C + \beta_2^{\theta,k} \ln \theta_{i,t}^H + \beta_3^{\theta,k} \ln g_i + \beta_4^{\theta,k} \ln I_{i,t}$ $+\beta_5^{\theta,k}(\ln \theta_{i,t}^C \cdot \ln I_{i,t}) + \beta_6^{\theta,k}(\ln \theta_{i,t}^H \cdot \ln I_{i,t}) + \beta_7^{\theta,k}(\ln g_i \cdot \ln I_{i,t})$ $+\beta_8^{\theta,k}(\ln \theta_{i,t}^C \cdot \ln \theta_{i,t}^H) + \beta_9^{\theta,k}(\ln \theta_{i,t}^C \cdot \ln g_i) + \beta_{10}^{\theta,k}(\ln \theta_{i,t}^H \cdot \ln g_i) + \varepsilon_{i,t}^{\theta,k}$ $\ln \theta_{i,f,t+1}^k = \beta_0^{F,k} + \beta_1^{F,k} \ln \theta_{i,t}^C + \beta_2^{F,k} \ln \theta_{i,t}^H$ $+\beta_3^{F,k} \ln g_i + \beta_4^{F,k} \ln I_{i,t} + \beta_5^{F,k}(\ln g_i \cdot \ln \theta_{i,t}^k) + \phi_f + \varepsilon_{i,f,t}^{F,k}$
IV. Heterogeneity in Investment	$I_{i,t} = \beta_0^x + \beta_1^x x_i^f + \beta_2^x x_i^b + \beta_3^x (x_i^f \cdot x_i^b) + \varepsilon_i^x$

Notes: This table provides an overview of the targeted moments used in the Simulated Method of Moments (SMM) procedure. Consistent with Section 4, we denote by $\theta_{i,t}^k$ the stock of skill $k \in \{C, H\}$ for child i in period t. The term ϕ_f is a family fixed effect. The indicator x_i^f (x_i^b) takes on the value 1 if child i is female (Black). In pooled regressions, we control for gender, race, age, and age squared. We omit race controls in within-family specifications. When simulating the empirical moments, we include these controls but do not include their corresponding coefficients as moments to target. Similarly, we do not target regression intercepts in characteristic groups II, III, or IV.

$\psi_{f(i)} = \psi_{f(j)}$, which enters both the utility function and the skill production technology. The simulated sample replicates the empirical panel design, and we estimate pooled and fixed effects regressions on the simulated data using the same specifications as in the Add Health data. Denote the vector of model parameters by Ω. Let M denote the vector of empirical moments and $M(\Omega)$ the corresponding model-simulated moments. The SMM estimator is obtained by minimizing the quadratic distance between empirical and simulated moments, as follows:

$$\widehat{\Omega} = \underset{\Omega}{\operatorname{argmin}} \Big(M - M(\Omega)\Big)' \Big(M - M(\Omega)\Big). \tag{16}$$

We compute standard errors using a nonparametric bootstrap. Specifically, we draw 100 bootstrap samples of the Add Health data (resampling with replacement) to generate 100 alternative realizations of the targeted moments and the initial-condition parameters estimated outside the model. For each bootstrap sample, we re-estimate the model to recover a new parameter vector. Standard errors are then calculated as the standard deviation of the bootstrap empirical distribution of these parameter estimates. This procedure captures the sampling uncertainty associated with both the initial conditions—estimated outside the model—and the structural parameters estimated within the model.

Table 6: Parameter Estimates

	Estimate	Std. Error
Panel A: Utility		
Cognitive Skills (γ_1)	0.501	0.033
Mental Health (γ_2)	0.455	0.044
Investment (γ_3)	−0.496	0.054
Heterogeneous Taste for Investment (γ_4)	−0.213	0.017
Panel B: Technology of Skill Formation (Cognitive)		
Cognitive Skills (α_1^C)	0.546	0.031
Mental Health (α_2^C)	0.038	0.013
Genetic Endowment (α_3^C)	0.052	0.009
Investment (α_4^C)	0.024	0.011
Cognitive Skills × Genetic Endowment (α_5^C)	0.072	0.015
Cognitive Skills × Investment (α_6^C)	0.021	0.012
Genetic Endowment × Investment (α_7^C)	−0.009	0.012
Total Factor Productivity Constant ($\alpha_{A,0}^C$)	−0.100	0.019
Unobserved Family-Level Heterogeneity ($\alpha_{A,1}^C$)	−0.074	0.007
Panel C: Technology of Skill Formation (Mental Health)		
Cognitive Skills (α_1^H)	0.076	0.018
Mental Health (α_2^H)	0.479	0.024
Genetic Endowment (α_3^H)	0.078	0.010
Investment (α_4^H)	0.013	0.012
Mental Health × Genetic Endowment (α_5^H)	0.017	0.018
Mental Health × Investment (α_6^H)	−0.010	0.013
Genetic Endowment × Investment (α_7^H)	−0.012	0.010
Total Factor Productivity Constant ($\alpha_{A,0}^H$)	0.101	0.011
Unobserved Family-Level Heterogeneity ($\alpha_{A,1}^H$)	−0.049	0.007

NOTES: The table reports point estimates and standard errors (in a separate column) for the parent utility specification (Panel A) and the technology of skill formation for cognitive skills (Panel B) and mental health (Panel C). Standard errors are obtained from 100 family-clustered nonparametric bootstraps of the estimation procedure.

5.4 Parameter Estimates

Table 6 displays the results of our main parameter estimates.[25] Panel A reports the estimated utility parameters, while Panel B (Panel C) presents the parameters governing the technology of skill formation for cognitive skills (mental health).

Turning to Panel A, our estimates suggest that parents value their children's cognitive skills and mental health almost equally. We find that investment is costly and that this disutility increases with the latent family-level heterogeneity. We interpret this heterogeneity as unobserved family disadvantage, consistent with our negative estimates in the skill formation technology showing that this variable adversely affects

[25] Table A16 presents additional estimates, including the race-by-gender-specific means ($\mu_{r,s}$) and scalar correlations ($\rho_{r,s}$) between initial conditions and the family-level heterogeneity.

skill development—a point we return to later in this section.

Consistent with the child development literature, Panel B shows that lagged cognitive skills have a strong and persistent effect on future cognitive skills. Lagged mental health also contributes meaningfully, at roughly 7% the size of the direct cognitive skill effect. The direct effect of a child's genetic endowment is sizable and statistically significant, amounting to approximately 10% of the cognitive skill effect. Parental investment is a significant and productive input in the technology. We also find evidence of positive and significant complementarities: the interaction between cognitive skills and genetic endowment, as well as that between cognitive skills and investment. These estimates suggest that the effectiveness of investment and genetic factors are amplified in children with higher levels of cognitive skills. The negative interaction between genetic endowment and investment is consistent with our reduced-form evidence in Section 3.4 and helps explain why parents exert less effort when their children have higher genetic endowments (i.e., lower ADHD risk).

The results for the mental health production function (Panel C) closely mirror those for cognitive skill formation. Lagged mental health is the most productive input in the production of future mental health, followed by genetic endowment and cognitive skills. The direct effects of genetic endowment and lagged cognitive skills are each approximately 16% the size of the direct effect of lagged mental health. In terms of input complementarities, we find that mental health is more productive among children with higher genetic endowments, as indicated by the positive coefficient on the interaction term (α_5^H). In contrast, investment becomes slightly less productive at higher levels of mental health and genetic endowment, as reflected in the negative coefficients on α_6^H and α_7^H. These results imply that investment is most effective in improving mental health among children with lower baseline mental health and greater genetic risk for ADHD.

5.5 Model's Fit

Table A17 through Table A21 report direct comparisons between the targeted empirical moments and the simulated moments generated at the model optimum. Each table corresponds to a distinct set of auxiliary regression coefficients used in the SMM procedure. Overall, the model successfully replicates the targeted moments. We briefly summarize the results below.

Table A17 focuses on moments related to the evolution of skills during adolescence (Model Characteristic I in Table 5). The intercepts are closely reproduced and we match the sign of the slope coefficient for mental health closely. The model overestimates the slope coefficient in the cognitive skill regression, although this targeted moment is itself imprecisely estimated.

Table A18 presents moments that inform the identification of parents' endogenous investment decisions. We consider both a pooled regression and a within-family specification of the reduced-form investment equations. In both cases, the model closely matches both the sign and magnitude of each coefficient. The model is also able to replicate the pattern of parents demonstrating a higher propensity to compensate against genetic risk in the within-family specification. Finally, the model captures the shift in the regression coefficient between pooled and within-family specifications, indicating that it recovers the correct correlation between family-level unobserved heterogeneity and children's skills and genetic endowments—the mechanism generating the discrepancy between the two empirical estimates.

Table A19 and Table A20 evaluate the model's ability to replicate reduced-form skill production functions for cognitive skills and mental health. Each specification includes a pooled regression with rich interaction terms and a simpler within-family version. Across all regressions, the model replicates the direct returns to inputs with high accuracy. It also matches most of the complementary returns, with small deviations in sign when the empirical coefficients are close to zero.

Finally, Table A21 compares regression coefficients related to heterogeneity in parental investment. The model aligns well with the data, capturing the pattern that parents invest more in daughters than they do sons, as well as the fact that Black parents tend to invest more on average in our sample. Our model is further able to replicate the negative coefficient attached to the interaction between female and Black child indicators.

6 Genetic Measures and Parents' Behavior

In Section 3.4, we showed that the reduced-form evidence on parental responses to children's genetic endowments either disappears or reverses when we replace our latent measure of genetic risk with the conventional polygenic score (PGS). In this section, we examine the extent to which relying on the PGS to measure g_i affects the estimation of our model and, by extension, whether it alters the model's predictions about parental investment decisions and their implications for subsequent skill formation. To do so, we re-estimate the structural model via SMM using the exact same empirical moments described in Table 5, replacing the latent genetic factor with the PGS for ADHD. We present the results of this exercise in Section 6.1. In Section 6.2, we then conduct counterfactual analyses to assess the influence of genetic endowments and the role of parental investment under the two models.

6.1 Model Estimates with PGS

To examine whether the discrepancies identified in Section 3.4 persist in our structural framework, we re-estimate the full model described in Section 4, using the PGS for ADHD from Demontis et al. (2019)—available in the Add Health data—as our measure of g_i. Let M_{PGS} denote the set of empirical moments constructed from Add Health using the PGS, and let $M_{\text{PGS}}(\Omega)$ represent the corresponding simulated moments generated by the model given parameter vector Ω.[26] As before (Equation 16), parameter estimates are obtained by solving the following SMM problem:

$$\widehat{\Omega}_{\text{PGS}} = \underset{\Omega}{\text{argmin}} \Big(M_{\text{PGS}} - M_{\text{PGS}}(\Omega)\Big)' W \Big(M_{\text{PGS}} - M_{\text{PGS}}(\Omega)\Big). \tag{17}$$

Table 7 presents the results of the estimation procedure.[27] For convenience, the first column reproduces the original estimates derived from our latent factor for the genetic risk of ADHD. Column (2) presents the accompanying estimates from a model that uses the PGS. The two approaches yield notably different results. In the discussion that follows, we focus on three key mechanisms to guide our comparison of the results: (i) the returns to genetics; (2) the returns to investment; and (3) parents' preferences.

The returns to genetics. Panel B shows that, in our preferred model, a one-standard deviation increase in genetic endowment (i.e., a one-standard deviation reduction in the genetic risk for ADHD) increases the production of cognitive skills at approximately 10% of the rate associated with lagged cognitive skills. By contrast, the PGS-based model produces a counterintuitive result: the same increase in genetic endowment decreases future cognitive skills at a rate equivalent to 13% of the effect of lagged cognitive skills. Consistent with the reduced-form results presented in Section 3.4, we therefore reverse our findings when relying on the PGS. Although Panel C does not present the same reversal of effects, our preferred specification suggests that genes are a productive input in the production of mental health, whereas the PGS-based model suggests that genes have no direct effect on mental health.[28]

In addition to these direct effects, the interaction between lagged skills and genetic endowment (captured by α_5^k) provides insight into how the persistence of skills varies with one's genetic endowment. In the latent factor specification, we estimate a large, positive complementarity between cognitive skills and

[26] The use of the PGS affects only the subset of moments derived from auxiliary regressions that include genetic endowment. All other moments remain identical across the two estimation procedures.Table A22 compares the set of moments involving genetic variables used in Equation 16 with those used in Equation 17.

[27] Table A23 presents additional estimates related to the unobserved family heterogeneity by race and gender. We do not find substantial differences between models.

[28] These findings are consistent with Table A22 wherein the reduced-form technology of skill formation shows a dampened effect on the direct returns to genetics. In the within-family specification of the auxiliary regression, the sign reverses between measures of genetic endowment.

Table 7: Parameter Estimates by Genetic Endowment Measure

	Latent Factor (1)	PGS (2)
Panel A: Utility		
Cognitive Skills (γ_1)	0.501	0.489
	(0.033)	(0.105)
Mental Health (γ_2)	0.455	0.104
	(0.044)	(0.040)
Investment (γ_3)	−0.496	−0.468
	(0.054)	(0.065)
Heterogeneous Taste for Investment (γ_4)	−0.213	−0.220
	(0.017)	(0.036)
Panel B: Technology of Skill Formation (Cognitive)		
Cognitive Skills (α_1^C)	0.546	0.568
	(0.031)	(0.056)
Mental Health (α_2^C)	0.038	0.049
	(0.013)	(0.031)
Genetic Endowment (α_3^C)	0.052	−0.076
	(0.009)	(0.026)
Investment (α_4^C)	0.024	0.020
	(0.011)	(0.028)
Cognitive Skills × Genetic Endowment (α_5^C)	0.072	0.019
	(0.015)	(0.025)
Cognitive Skills × Investment (α_6^C)	0.021	0.021
	(0.012)	(0.029)
Genetic Endowment × Investment (α_7^C)	−0.009	−0.009
	(0.012)	(0.022)
Total Factor Productivity Constant ($\alpha_{A,0}^C$)	−0.100	−0.100
	(0.019)	(0.028)
Unobserved Family-Level Heterogeneity ($\alpha_{A,1}^C$)	−0.074	−0.072
	(0.007)	(0.030)
Panel C: Technology of Skill Formation (Mental Health)		
Cognitive Skills (α_1^H)	0.076	0.092
	(0.018)	(0.026)
Mental Health (α_2^H)	0.479	0.470
	(0.024)	(0.051)
Genetic Endowment (α_3^H)	0.078	−0.006
	(0.010)	(0.022)
Investment (α_4^H)	0.013	0.007
	(0.012)	(0.022)
Mental Health × Genetic Endowment (α_5^H)	0.017	0.028
	(0.018)	(0.024)
Mental Health × Investment (α_6^H)	−0.010	−0.010
	(0.013)	(0.031)
Genetic Endowment × Investment (α_7^H)	−0.012	−0.012
	(0.010)	(0.025)
Total Factor Productivity Constant ($\alpha_{A,0}^H$)	0.101	0.099
	(0.011)	(0.033)
Unobserved Family-Level Heterogeneity ($\alpha_{A,1}^H$)	−0.049	0.001
	(0.007)	(0.029)

NOTES: The above table shows the estimates of the parent utility specification (Panel A) and the technology of skill formation for cognitive skills (Panel B) and mental health (Panel C). Column (1) corresponds to a model that uses our latent factor measure of the genetic risk for ADHD, whereas column (2) corresponds to the same model that instead uses the Demontis et al. (2019) PGS for ADHD. Standard errors in parentheses are obtained from 100 family-clustered non-parametric bootstraps of the estimation procedure.

genetic endowment, which is reduced by 74% in the PGS model and no longer statistically significant. This reduction implies that, relative to the latent factor specification, the PGS model predicts a smaller marginal return to investing in low-risk children with higher cognitive skills, weakening the incentives for parents to direct resources toward them. These differences are less pronounced in the mental health technology.

The returns to investment. Panel C reveals that a model that uses the PGS to measure genetic endowment predicts a return to investment on mental health that is 54% of the return predicted by a model that uses our latent factor. The same coefficient for cognitive skills shows a similar reduction in returns (17%). It is also important to consider the differences in the dynamic returns to lagged skills (Cunha and Heckman, 2008). In the model, the dynamic return to investment depends not only on the direct effect of investment, but also on how investment interacts with existing skills (α_6^k) and, indirectly, on how those skills evolve through their interaction with genetic endowment (α_5^k). The latent factor specification produces larger estimates of α_5^C and identical estimates of α_6^C, implying that investment in children with high cognitive skills and low genetic risk generates greater future skill accumulation by increasing the productivity of current skills. In contrast, the smaller values in the PGS model dampen this compounding effect, reducing the model's predicted long-run gains from investment in these children. Again, the differences across specifications are less pronounced for the mental health technology.[29]

Parents' preferences. In our preferred specification, parents value their children's cognitive skills and mental health approximately equally. In contrast, the PGS-based model implies that parents value cognitive skills roughly five times as much as mental health. This reduction in preferences for mental health is explained by the reduction in how parental investments respond to changes in mental health in the PGS-based, within-family investment regression (Panel A of Table A22). Because parents make a single investment decision that affects both skills simultaneously, this shift in relative preferences changes the overall marginal benefit of investment in ways that interact with the differences in the estimated technology of skill formation. These combined differences in preferences and technology have direct implications for how parental investment responds to changes in genetic endowment, which we explore in the following section.

[29] This attenuation in dynamic returns is consistent with Table A22, where the coefficients on the direct effect of investment, the interaction of investment with cognitive skills, and the interaction of investment with mental health in the pooled specifications are all smaller when using the PGS. In the within-family specifications, the coefficient on cognitive skills interacted with genetic endowment also reverses sign.

6.2 The Implications of Genetic Measures on Model Predictions

To evaluate the implications of the model for the role of genetic endowments and parental investment in skill formation, we conduct a quantitative exercise that exogenously increases each child's genetic endowment by 10%. For each skill type $k \in \{C, H\}$, we estimate the average treatment effect of this increase under two distinct scenarios. In a slight abuse of notation, let $\theta_T^k(g_i, \boldsymbol{I}_i)$ denote the final stock of skill k for child i as a function of their genetic endowment g_i and the full path of parental investments across adolescence, $\boldsymbol{I}_i \coloneqq \{I_{i,t}\}_{t=0}^T$. Let g_i denote the baseline genetic endowment and $g_i' = 1.1 \times g_i$ the counterfactual one. Let $\boldsymbol{I}_i$ and $\boldsymbol{I}_i'$ analogously denote the baseline and counterfactual sequences of endogenous parental investment.

The first object of interest is the *direct effect* of increasing g_i, which holds parental investment fixed at its baseline level. Formally, we define

$$\tau_g^k \coloneqq \mathbb{E}\left[\ln \theta_T^k(g_i', \boldsymbol{I}_i) - \ln \theta_T^k(g_i, \boldsymbol{I}_i)\right]. \tag{18}$$

This treatment effect τ_g^k captures the mechanical contribution of genetic endowment to the production of skill k, abstracting entirely from behavioral responses. The second object, which we refer to as the *total effect*, allows parents to respond optimally to the updated value of g_i in each period. We define this effect as

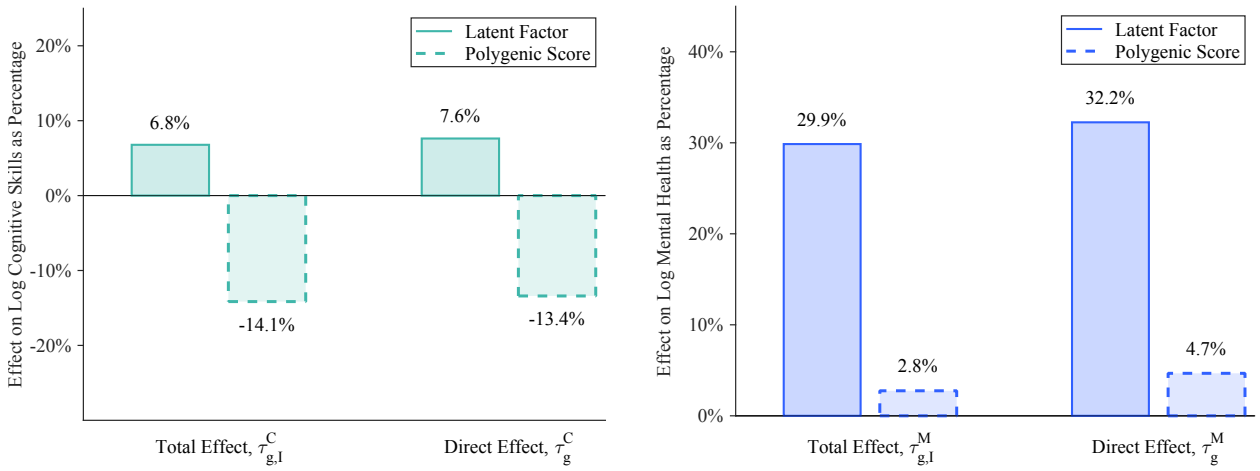
$$\tau_{g,I}^k \coloneqq \mathbb{E}\left[\ln \theta_T^k(g_i', \boldsymbol{I}_i') - \ln \theta_T^k(g_i, \boldsymbol{I}_i)\right]. \tag{19}$$

This expression incorporates both the direct influence of genetic endowments and the indirect effect operating through endogenous parental investment. We report both effects in standard deviation units, normalizing the average treatment effect on log skill by the standard deviation of baseline log skill in period T. This provides us with a scale-free measure that can be compared across skill types and model specifications.

Importantly, we can use these two effects to construct a measure of the relative importance of parental investment in response to changes in genetic endowment. In particular, the ratio $\tau_{g,I}^k / \tau_g^k$ captures the degree to which the total impact of increased genetic endowment is amplified (or attenuated) by endogenous parental behavior. A value greater than one indicates that parents reinforce the effect of improved endowment through increased investment, while a value less than one implies that parental responses partially offset the direct genetic advantage. This ratio therefore provides insight into the complementarity (or substitutability) between parental investment and genetic endowment in the production of child skills.

Figure 2 reports the total and direct effects of increasing g_i on each skill type under both measures of

Figure 2: Effects of Increasing Genetic Endowment by 10%, by Genetic Measure



(a) Cognitive Skills **(b)** Mental Health

NOTES: Each panel reports the effect of increasing genetic endowment g_i by 10% on (a) cognitive skills (green) and (b) mental health (blue). Effects are computed as the log-difference in outcomes between counterfactual and baseline levels, normalized in terms of the population-level standard deviation of final (log) skills at baseline. Within each panel, the first two bars show the *total* effect $(\tau^{C}_{g,I}, \tau^{M}_{g,I})$, which allows for a endogenous parental investment response. The next two bars show the *direct* effect $(\tau^{C}_{g}, \tau^{M}_{g})$, which fixes parental investment to the baseline level. We consider two different measure for genetic endowment: our latent factor for genetic risk of ADHD (solid outline) and the ADHD polygenic score from Demontis et al. (2019) (dotted outline).

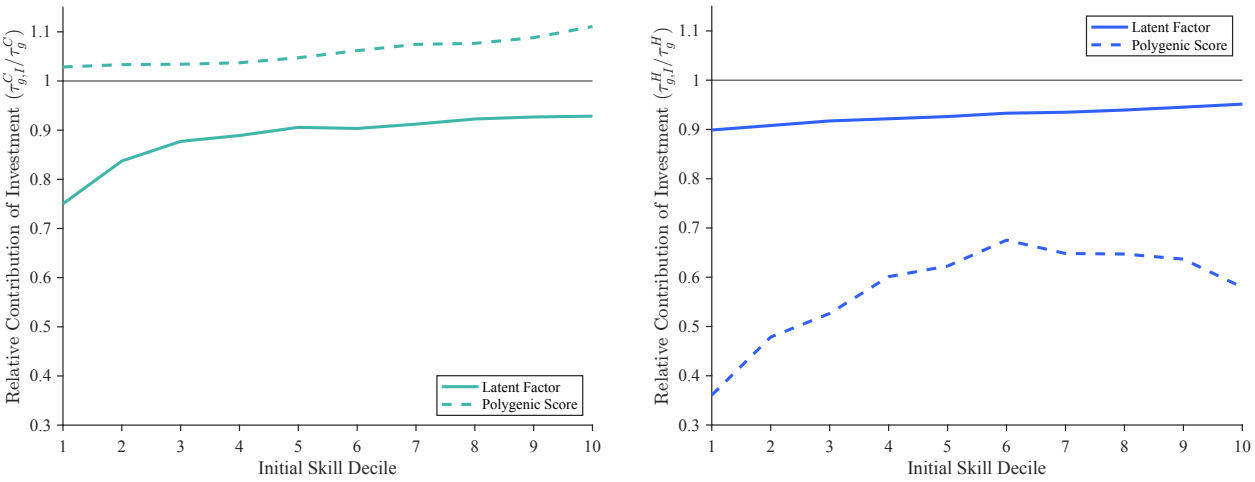
genetic endowment. The results differ markedly between the latent-factor and PGS-based specifications. Using our preferred measure, a 10% increase in genetic endowment raises terminal cognitive skills by about 0.07 standard deviations and terminal mental health by about 0.3 standard deviations.[30] In contrast, under the PGS-based model, the same change reduces cognitive skills by roughly 0.14 standard deviations and increases mental health by only 0.03 standard deviations.

Comparing total and direct effects reveals how parental behavior mediates these impacts. For the latent factor, the direct effect on cognitive skills is about 12% larger than the total effect, implying parents reduce investments when genetic endowments improve. For the PGS, the direct effect is about 5% smaller than the total, suggesting parents increase investments instead. In other words, a model based on the latent factor predicts that parents compensate against genetic risk, whereas a model based on the PGS predicts that parents reinforce genetic advantages through greater investment. These opposing behavioral inferences—along with the PGS model's counterintuitive negative cognitive effect—underscore the potential limitations of polygenic scores as proxies for genetic endowment.

We next compare the heterogeneity in the contribution of parental investment to mediating genetic effects, $\tau^{k}_{g,I}/\tau^{k}_{g}$, across the two measures of genetic endowment. Figure 3 summarizes these results by

[30] While the estimated effect is large, this finding aligns with evidence that depression is more prevalent in adults with ADHD, with reported prevalence as high as 53.3% (Fu et al., 2025).

Figure 3: Relative Contribution of Investment in Mediating Genetic Effects, by Skill Type and Measure



(a) Cognitive Skills **(b)** Mental Health

NOTES: This figure plots the ratio $\tau^k_{g,I}/\tau^k_g$, where τ^k_g captures the direct effect of increasing g_i by 10% while holding investment fixed, and $\tau^k_{g,I}$ allows for parental reoptimization. Each line displays this ratio by baseline skill decile, separately for our latent factor measure of genetic risk and for the polygenic score (PGS) from Demontis et al. (2019). Panel (a) displays this ratio for cognitive skills and panel (b) displays this value for mental health.

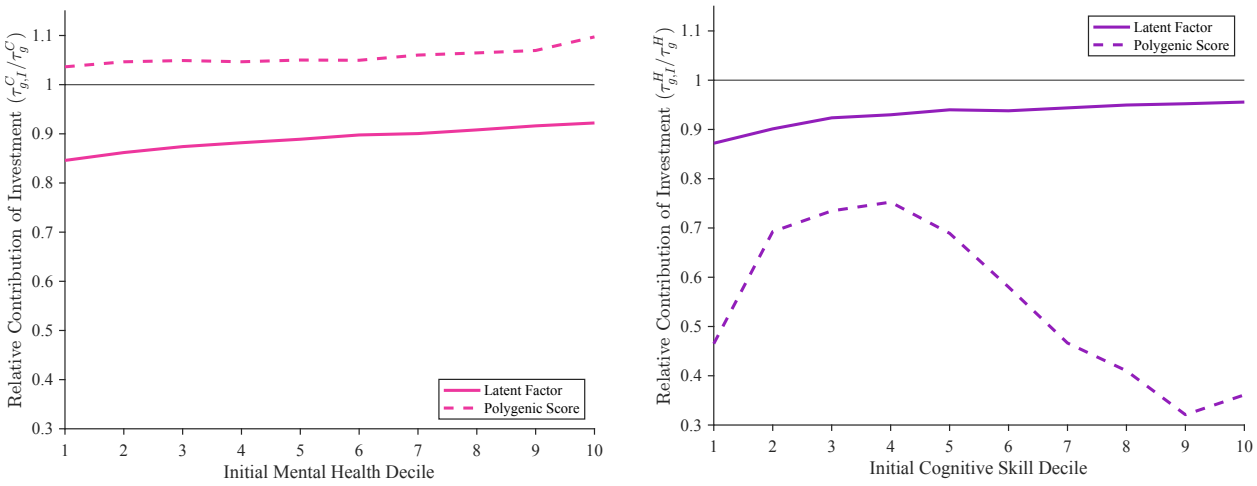
deciles of initial skill endowments. Panel (a) presents the results for cognitive skills, and Panel (b) shows the corresponding patterns for mental health.

Consistent with our reduced-form findings (Table 2), we find that, under our preferred specification, this ratio ranges from 0.75 in the lowest to 0.91 in the highest decile of initial cognitive skills. This pattern indicates that parents reduce their investment in response to higher genetic endowments, with the contribution of parental investments in mediating genetic effects being considerably stronger among children at the lower end of the cognitive skill distribution. In contrast, when relying on the PGS, we obtain ratios exclusively greater than one—reversing the pattern observed in the latent-factor model. In other words, the way genetic endowment is measured leads to conflicting conclusions about both the direction of parental responses and the role of parental investments in mediating genetic influences on skill formation.

The heterogeneity analysis in our preferred model reveals that the extent to which parental investments mediate genetic effects on skill formation diminishes at higher initial skill levels. By contrast, the PGS yields a more uniformly increasing profile, with mediation effects that strengthen modestly across deciles of initial skills.

For mental health, both measures indicate compensatory behavior. Under our preferred specification, the profile is relatively flat around 0.9, suggesting that the extent to which parental investments mediate genetic effects on mental health is largely homogeneous across the distribution of initial mental health skills.

Figure 4: Relative Contribution of Investment by Cross-Skill Type and Measure



(a) Cognitive Skills by Initial Mental Health Decile

(b) Mental Health by Initial Cognitive Skill Decile

NOTES: This figure plots the ratio $\tau^k_{g,I}/\tau^k_g$, where τ^k_g captures the direct effect of increasing g_i by 10% while holding investment fixed, and $\tau^k_{g,I}$ allows for parental reoptimization. Each line displays this ratio by *cross* baseline skill decile, separately for our latent factor measure of genetic risk and for the polygenic score (PGS) from Demontis et al. (2019). Panel (a) displays this ratio for cognitive skills as a function of the initial mental health decile and panel (b) displays this value for mental health as a function of the initial cognitive skill decile.

In contrast, the PGS exhibits a concave-down pattern, implying that mediation effects are most pronounced near the tails of the distribution. Overall, these results underscore that the choice of genetic measure can substantially affect both the magnitude and direction of inferred parental investment responses, as well as the implied consequences for genetic influences on skill formation.

In Figure 4, we extend the analysis to cross-skill dynamics by plotting the ratio $\tau^k_{g,I}/\tau^k_g$ for skill k as a function of the initial decile of the other skill, $k' \neq k$. Panel (a) shows how the total effect on cognitive skills varies with baseline mental health, while Panel (b) shows the converse pattern for mental health as a function of initial cognitive skills.

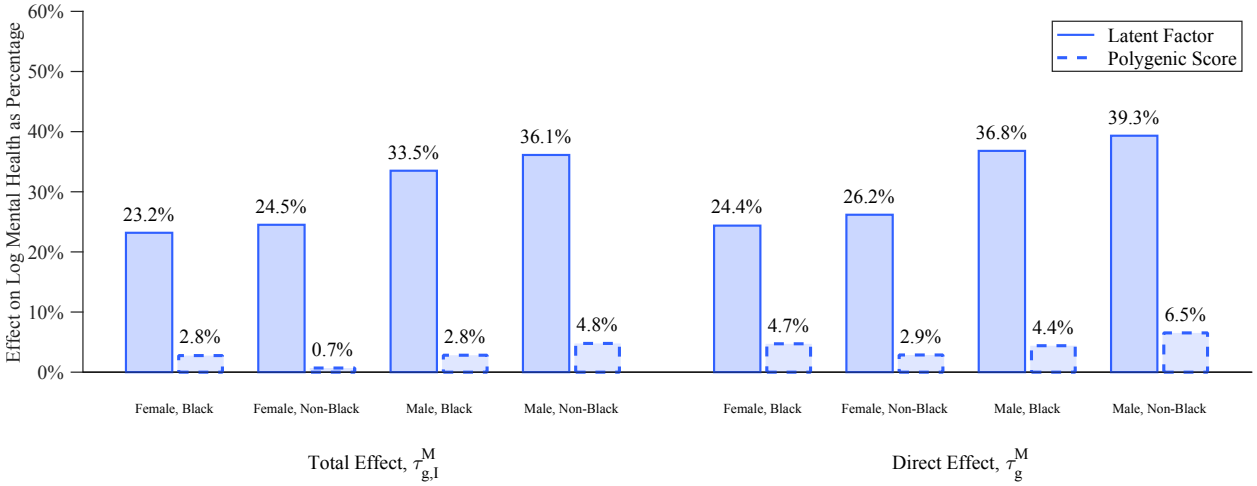
For cognitive skills, the latent-factor measure starts below one and rises modestly with baseline mental health, consistent with mild compensatory investment that weakens at higher mental health levels. In contrast, the PGS profile lies slightly above one and remains largely flat, indicating limited variation in investment responses across deciles.

We conclude this section by examining effect heterogeneity on the basis of race and gender. Figure 5 repeats the analyses conducted in Figure 3 separately by the four race-gender groups. In panel (a), we focus on the effect of a 10% increase in g_i on terminal (log) cognitive skills. Across both genetic measures, we find little variation across race and gender. However, in turning our attention to mental health in panel (b), we find stark differences. The latent factor exhibits a clear gender-based pattern—namely, an increase

Figure 5: Heterogeneity in Effects by Race and Gender



(a) Cognitive Skills



(b) Mental Health

NOTES: Each panel reports the effect of increasing genetic endowment g_i by 10% on (a) cognitive skills (green) and (b) mental health (blue). Effects are computed as the log-difference in outcomes between counterfactual and baseline levels, normalized in terms of the population-level standard deviation of final (log) skills at baseline. We show these effects by each (r, s) group for race r and gender s. Within each panel, the first eight bars show the *total* effect $(\tau^{C}_{g,I}, \tau^{M}_{g,I})$, which allows for a endogenous parental investment response. The next eight bars show the *direct* effect $(\tau^{C}_{g}, \tau^{M}_{g})$, which fixes parental investment to the baseline level. We consider two different measure for genetic risk: our latent factor for genetic risk of ADHD (solid outline) and the ADHD polygenic score from Demontis et al. (2019) (dotted outline).

in genetic endowment leads to larger mental health gains for male adolescents, with little evidence of additional race-based heterogeneity. The PGS-based model does not generate this pattern; instead, Black adolescents share similar effect sizes regardless of gender.

Taken together, these findings provide empirical support for the identification concerns outlined in Section 2.2. As shown in Equation 5 and Equation 6, PGS can yield biased estimates of the causal effects of genetic endowments when parental investments are endogenous and gene–environment interactions

are present. This arises because PGS combine direct genetic effects with indirect effects operating through parental responses and environmental complementarities, introducing simultaneity and omitted-variable bias. Consistent with this prediction, we find that estimates from the PGS-based model differ substantially from those obtained using our latent genetic factor. In particular, the PGS model generates contrasting counterfactual predictions about skill formation and the contribution of genetics to inequality in skills among children, relative to our preferred specification based on the latent-factor measure of genetic endowment. Overall, these results underscore the risks of treating polygenic scores as exogenous proxies for genetic endowments and highlight the importance of explicitly modeling behavioral responses and gene–environment interplay.

7 Conclusion

This paper develops and estimates a structural model of child development that separately identifies direct genetic effects from endogenous behavioral responses by parents. We construct a novel latent measure of genetic endowment by treating individual SNPs as noisy indicators of underlying risk and anchoring this measure to the 18 behavioral criteria used to diagnose ADHD. This approach mitigates the environmental biases inherent in polygenic scores. Embedding this measure within a dynamic model of skill formation and parental investment, we show that parents respond to greater genetic risk with increased investment, particularly when children begin with lower cognitive skills or poorer mental health.

Our findings demonstrate that conventional polygenic scores, which conflate genetic and environmental factors, yield misleading inferences about the role of nature in skill formation. Using the PGS in place of our latent measure reverses the estimated direction of genetic effects and substantially attenuates complementarities between endowment and other inputs. These discrepancies affect predictions of long-run outcomes and have implications for how researchers interpret gene–environment interactions. More broadly, our work contributes to the literature on human capital formation by showing how genetic endowments and parental behavior jointly shape skill development in adolescence. By separating nature from nurture in both measurement and modeling, this framework provides new tools to inform future research and policy aimed at improving adolescent development and designing interventions that help close achievement gaps linked to early-life disadvantage.

References

Agostinelli, Francesco, and Matthew Wiswall. 2025. “Estimating the Technology of Children’s Skill Formation.” *Jour-*

nal of Political Economy, 133(3): 846–887.

Agostinelli, Francesco, Domenico Ferraro, Xincheng Qiu, and Giuseppe Sorrenti. 2024. "Intra-Household Insurance and the Intergenerational Transmission of Income Risk." National Bureau of Economic Research Working Paper 32096.

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2025. "It Takes a Village: The Economics of Parenting with Neighborhood and Peer Effects." Forthcoming, *Journal of Political Economy*.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*. Vol. 5, American psychiatric association Washington, DC.

Amin, Vikesh, Jere R Behrman, Jason M Fletcher, Carlos A Flores, Alfonso Flores-Lagunes, and Hans-Peter Kohler. 2021. "Genetic risks, adolescent health, and schooling attainment." *Health economics*, 30(11): 2905–2920.

Attanasio, Orazio P, Áureo De Paula, and Alessandro Toppeta. 2025. "Intergenerational mobility in socio-emotional skills." *Journal of Public Economics*, 248: 105423.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. "Estimating the production function for human capital: results from a randomized controlled trial in Colombia." *American Economic Review*, 110(1): 48–85.

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

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of political Economy*, 87(6): 1153–1189.

Belsky, Daniel W., and Kathryn Paige Harden. 2019. "Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down." *Current Directions in Psychological Science*, 28(1): 82–90.

Belsky, Daniel 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): 957–972.

Benjamin, Daniel J., David Cesarini, Christopher F. Chabris, Edward L. Glaeser, David I. Laibson, et al. 2012. "The promises and pitfalls of genoeconomics." *Annual Review of Economics*, 4: 627–662.

Benjamin, Daniel J, David Cesarini, Patrick Turley, and Alexander Strudwick Young. 2024. "Social-science genomics: Progress, challenges, and future directions."

Bertoni, Marco, Blas A Marin-Lopez, and Anna Sanz-de Galdeano. 2023. "Subjective Gender-Based Patterns in ADHD Diagnosis." IZA Discussion Papers.

Bolt, Uta, Eric French, Varun Warrier, Qianyu Yang, and Weilong Zhang. 2025. "Genetic Endowments and Lifetime Earnings: Understanding the Mechanisms." Faculty of Economics, University of Cambridge.

Boyle, Coleen A, Sheree Boulet, Laura A Schieve, Robin A Cohen, Stephen J Blumberg, Marshalyn Yeargin-Allsopp, Susanna Visser, and Michael D Kogan. 2011. "Trends in the prevalence of developmental disabilities in US children, 1997–2008." *Pediatrics*, 127(6): 1034–1042.

Breinholt, Asta, and Dalton Conley. 2023. "Child-driven parenting: Differential early childhood investment by off-

spring genotype." *Social Forces*, 102(1): 310–329.

Carvalho, Leandro Siqueira. 2025. "Genetics and Socioeconomic Status: Some Preliminary Evidence on Mechanisms." *Journal of Political Economy: Microeconomics*.

Caucutt, Elizabeth M., and Lance Lochner. 2020. "Early and Late Human Capital Investments, Borrowing Constraints, and the Family." *Journal of Political Economy*, 128(3): 1065–1147.

Caucutt, Elizabeth M, Lance Lochner, Joseph Mullins, and Youngmin Park. 2020. "Child skill production: Accounting for parental and market-based time and goods investments." National Bureau of Economic Research.

Cawley, John, Euna Han, Jiyoon Kim, and Edward C Norton. 2023. "Genetic nurture in educational attainment." *Economics & Human Biology*, 49: 101239.

Claussen, Angelika H, Joseph R Holbrook, Helena J Hutchins, Lara R Robinson, Jeanette Bloomfield, Lu Meng, Rebecca H Bitsko, Brenna O'Masta, Audrey Cerles, Brion Maher, et al. 2024. "All in the family? A systematic review and meta-analysis of parenting and family environment as risk factors for attention-deficit/hyperactivity disorder (ADHD) in children." *Prevention Science*, 25(Suppl 2): 249–271.

Cunha, Flavio, and James Heckman. 2007. "The technology of skill formation." *American economic review*, 97(2): 31–47.

Cunha, Flavio, and James J Heckman. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of human resources*, 43(4): 738–782.

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

Davey Smith, George, and Shah Ebrahim. 2003. "Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease?" *International Journal of Epidemiology*, 32(1): 1–22.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household choices and child development." *Review of Economic Studies*, 81(1): 137–185.

Demontis, Ditte, G Bragi Walters, Georgios Athanasiadis, Raymond Walters, Karen Therrien, Trine Tollerup Nielsen, Leila Farajzadeh, Georgios Voloudakis, Jaroslav Bendl, Biau Zeng, et al. 2023. "Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains." *Nature genetics*, 55(2): 198–208.

Demontis, Ditte, Raymond K Walters, Joanna Martin, Manuel Mattheisen, Thomas D Als, Esben Agerbo, Gísli Baldursson, Rich Belliveau, Jonas Bybjerg-Grauholm, Marie Bækvad-Hansen, et al. 2019. "Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder." *Nature genetics*, 51(1): 63–75.

Domingue, Benjamin W, Daniel W Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D Boardman. 2015. "Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health." *AERA open*, 1(3): 2332858415599972.

Doust, Catherine, Pierre Fontanillas, Else Eising, Scott D Gordon, Zhengjun Wang, Gökberk Alagöz, Barbara Molz, Beate St Pourcain, et al. 2022. "Discovery of 42 genome-wide significant loci associated with dyslexia." *Nature genetics*, 54(11): 1621–1629.

Faraone, Stephen V, and Henrik Larsson. 2019. "Genetics of attention deficit hyperactivity disorder." *Molecular psychiatry*, 24(4): 562–575.

Fleming, Michael, Catherine A Fitton, Markus FC Steiner, James S McLay, David Clark, Albert King, Daniel F Mackay, and Jill P Pell. 2017. "Educational and health outcomes of children treated for attention-deficit/hyperactivity disorder." *JAMA pediatrics*, 171(7): e170691–e170691.

Fletcher, Jason, and Barbara Wolfe. 2009. "Long-term consequences of childhood ADHD on criminal activities." *The journal of mental health policy and economics*, 12(3): 119.

Fu, Xinyu, Weige Wu, Yuru Wu, Xiaofu Liu, Wanting Liang, Rongchuan Wu, and Yun Li. 2025. "Adult ADHD and comorbid anxiety and depressive disorders: a review of etiology and treatment." *Frontiers in Psychiatry*, 16: 1597559.

Galton, Francis. 1874. *English Men of Science: Their Nature and Nurture.* London:Macmillan and Co.

Giannola, Michele. 2024. "Parental investments and intra-household inequality in child human capital: Evidence from a survey experiment." *The Economic Journal*, 134(658): 671–727.

Gray, Richard, and Keith Wheatley. 1991. "How to avoid bias when comparing bone marrow transplantation with chemotherapy." *Bone marrow transplantation*, 7: 9–12.

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

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

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review*, 103(6): 2052–2086.

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm. 2024. "The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills." *American Economic Review*, 114(2): 385–425.

Jeong, Yeongmi, Nicholas W Papageorge, Meghan Skira, and Kevin Thom. 2024. "Genetic Risk for Alzheimer's Disease and Related Dementias: Cognition, Economic Behavior, and Clinically Actionable Information." National Bureau of Economic Research.

Karlsson Linnér, Richard, Pietro Biroli, Edward Kong, S Fleur W Meddens, Robbee Wedow, Mark Alan Fontana, Maël Lebreton, Stephen P Tino, Abdel Abdellaoui, Anke R Hammerschlag, et al. 2019. "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature genetics*, 51(2): 245–257.

Krueger, Alan B, and Orley C Ashenfelter. 1992. "Estimates of the economic return to schooling from a new sample of twins."

Lander, Eric S, L M Linton, Bruce Birren, and et al. 2001. "Initial sequencing and analysis of the human genome." *Nature*, 409(6822): 860–921.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, 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): 1112–

1121.

Lo, Min-Tzu, David A Hinds, Joyce Y Tung, Carol Franz, Chun-Chieh Fan, Yunpeng Wang, Olav B Smeland, Andrew Schork, Dominic Holland, Karolina Kauppi, et al. 2017. "Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders." *Nature genetics*, 49(1): 152–156.

Locke, John. 1689. *An Essay Concerning Human Understanding*. London:Thomas Bassett.

Loe, Irene M, and Heidi M Feldman. 2007. "Academic and educational outcomes of children with ADHD." *Journal of pediatric psychology*, 32(6): 643–654.

Manolio, Teri A, Francis S Collins, Nancy J Cox, David B Goldstein, Lucia A Hindorff, David J Hunter, Mark I McCarthy, Erin M Ramos, Lon R Cardon, Aravinda Chakravarti, et al. 2009. "Finding the missing heritability of complex diseases." *Nature*, 461(7265): 747–753.

Marquardt, Kelli. 2022. "Mis (sed) Diagnosis: Physician Decision-Making and ADHD."

Mullins, Joseph. 2025. "Designing Cash Transfers in the Presence of Children's Human Capital Formation." University of Minnesota Working Paper.

Muslimova, Dilnoza, Hans Van Kippersluis, Cornelius A Rietveld, Stephanie Von Hinke, and S Fleur W Meddens. 2024. "Gene-environment complementarity in educational attainment."

Plug, Erik, and Wim Vijverberg. 2003. "Schooling, family background, and adoption: Is it nature or is it nurture?" *Journal of political Economy*, 111(3): 611–641.

Rietveld, Cornelius A, and Pankaj C Patel. 2019. "ADHD and later-life labor market outcomes in the United States." *The European Journal of Health Economics*, 20(7): 949–967.

Rustichini, Aldo, William G Iacono, James J Lee, and Matt McGue. 2023. "Educational attainment and intergenerational mobility: A polygenic score analysis." *Journal of Political Economy*, 131(10): 2724–2779.

Sacerdote, Bruce. 2002. "The nature and nurture of economic outcomes." *American Economic Review*, 92(2): 344–348.

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

Sanz-de Galdeano, Anna, and Anastasia Terskaya. 2022. "Sibling Differences in Genetic Propensity for Education: How do Parents React?" *The Review of Economics and Statistics*, 1–44.

Staff, Anouck I, Jaap Oosterlaan, Saskia Van der Oord, Barbara J van den Hoofdakker, and Marjolein Luman. 2023. "The relation between classroom setting and ADHD behavior in children with ADHD compared to typically developing peers." *Journal of Attention Disorders*, 27(9): 939–950.

Todd, Petra E, and Kenneth I Wolpin. 2007. "The production of cognitive achievement in children: Home, school, and racial test score gaps." *Journal of Human capital*, 1(1): 91–136.

Turic, Darko, James Swanson, and Edmund Sonuga-Barke. 2010. "DRD4 and DAT1 in ADHD: Functional neurobiology to pharmacogenetics." *Pharmacogenomics and personalized medicine*, 61–78.

Visser, SN, ML Danielson, RH Bitsko, JR Holbrook, MD Kogan, RM Ghandour, R Perou, SJ Blumberg, and S Visser. 2014. "Trends in the Parent-report of Health Care Provider Diagnosed and Medicated ADHD: United States, 2003–2011 HHS Public Access." *J Am Acad Child Adolesc Psychiatry*, 53(1): 34–46.

Watson, John B. 1924. *Behaviorism.* New York:People's Institute Publishing Company.

Young, Alexander I, Seyed Moeen Nehzati, Stefania Benonisdottir, Aysu Okbay, Hariharan Jayashankar, Chanwook Lee, David Cesarini, Daniel J Benjamin, Patrick Turley, and Augustine Kong. 2022. "Mendelian imputation of parental genotypes improves estimates of direct genetic effects." *Nature genetics*, 54(6): 897–905.

Online Appendix

A Additional Tables

Table A1: Reproduction of Table 1 in Demontis et al. (2023)

Genomic locus	Chr.	Position (bp)	rs ID	A1	A2	Odds Ratio	Freq. Cases	Freq. Controls
1	1	44076469	rs549845	G	A	1.082	0.321	0.326
2	2	145714354	rs1438898	A	C	1.065	0.762	0.769
3	3	20724204	rs2886697	G	A	1.061	0.634	0.643
4	3	43691501	rs9877066	G	A	0.888	0.944	0.951
5	3	49916710	rs7613360	C	T	0.948	0.598	0.614
6	3	51884072	rs2311059	G	A	0.944	0.314	0.308
7	3	71499401	rs17718444	C	T	1.063	0.695	0.660
8	3	87015142	rs114142727	C	G	1.285	0.988	0.988
9	4	112217523	rs17576773	C	T	1.101	0.888	0.880
10	4	147099654	rs6537401	G	A	0.945	0.660	0.655
11	5	87854395	rs4916723	A	C	0.918	0.553	0.573
12	5	103964585	rs77960	G	A	0.929	0.665	0.682
13	5	144474779	rs10875612	C	T	0.947	0.483	0.470
14	6	70858701	rs2025286	A	C	0.947	0.553	0.550
15	7	67685754	rs73145587	A	T	1.107	0.910	0.901
16	7	114158954	rs9969232	G	A	0.934	0.344	0.382
17	8	93277087	rs7844069	T	G	1.057	0.428	0.399
18	8	145802447	rs4925811	T	G	0.944	0.515	0.531
19	10	8784773	rs11255890	C	A	1.054	0.389	0.401
20	10	106453832	rs11596214	G	A	1.054	0.597	0.569
21	11	28602173	rs2582895	C	A	1.075	0.634	0.618
22	12	89771903	rs704061	T	C	0.946	0.554	0.560
23	14	98690923	rs76284431	T	A	0.922	0.847	0.842
24	16	61966703	rs1162202	C	T	1.063	0.630	0.606
25	18	5871800	rs76857496	C	A	1.083	0.870	0.859
26	18	50625779	rs7506904	G	A	0.946	0.343	0.372
27	20	21250843	rs6082363	T	C	1.073	0.296	0.291

Notes: This table displays the material presented in Table 1 in Demontis et al. (2023) for the SNPs used in our analysis. We omit several columns from the original table to preserve space, including information on the p-value, the recovered standard error, and indicators for whether the gene is new relative to those found in Demontis et al. (2019). Two additional SNPs (rs11420276 and rs5886709) not present in the above table but found to be significant in Demontis et al. (2019) are used in the analysis.

Table A2: Descriptive Statistics—Latent Factor Measures of Cognitive Skills and Mental Health

	Analysis Sample		Siblings Sample		Diagnosed Sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Panel A: Cognitive Skills						
English Grade	2.842	0.959	2.874	0.965	2.508	0.986
Math Grade	2.671	1.026	2.724	1.018	2.414	1.059
Social Studies Grade	2.864	0.984	2.892	0.995	2.591	1.001
Science Grade	2.817	0.978	2.832	0.993	2.575	1.003
Panel B: Mental Health						
Bothered by things that don't usually bother you	0.518	0.708	0.530	0.720	0.570	0.765
Didn't feel like eating	0.471	0.713	0.492	0.720	0.446	0.719
Could not shake off the blues	0.409	0.712	0.417	0.736	0.495	0.832
Felt you were not just as good as other people	1.101	1.002	1.146	1.008	0.970	0.959
Trouble keeping mind on what you were doing	0.825	0.809	0.865	0.816	1.073	0.953
Felt depressed	0.522	0.752	0.529	0.743	0.608	0.835
Felt too tired to do things	0.739	0.745	0.752	0.745	0.761	0.780
Did not feel hopeful about the future	1.162	0.971	0.981	0.720	1.094	0.937
Thought life had been a failure	0.195	0.513	0.208	0.538	0.253	0.565
Felt fearful	0.327	0.569	0.327	0.575	0.387	0.677
Not happy	0.857	0.801	0.874	0.810	0.901	0.800
Talked less than usual	0.558	0.731	0.586	0.745	0.595	0.807
Felt lonely	0.465	0.702	0.483	0.704	0.551	0.706
People were unfriendly	0.412	0.627	0.426	0.612	0.540	0.728
Did not enjoy life	0.754	0.849	0.775	0.860	0.772	0.846
Felt sad	0.583	0.682	0.580	0.676	0.613	0.712
Felt disliked by people	0.415	0.636	0.433	0.637	0.476	0.721
Hard to get started doing things	0.636	0.696	0.648	0.711	0.726	0.734
Life not worth living	0.147	0.474	0.157	0.497	0.180	0.516
Observations	6,987		1,608		372	

Notes: The above table presents summary statistics for different sub-samples of the Add Health survey. The analysis sample refers to those who agreed to take part in the Wave IV DNA sample. We maintain this restriction in the next two groupings as well. The sibling sample refers to those with at least one full sibling linkage or one twin linkage in Add Health. The diagnosed sample includes adolescents that are coded as having ADHD through at least one of the following ways: taken medication for ADHD in the past 12 months (Wave III) or have had a doctor say they have ADHD (Wave IV). Panel A grades are measured from 0 to 4. Panel B presents the 19 depression survey questions used to construct our mental health latent factor score. These are measured on a scale of 0 to 3, with 3 indicating strongly agree.

Table A3: Descriptive Statistics—Latent Factor Measures of Genetic Risk

	Analysis Sample		Siblings Sample		Diagnosed Sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
rs1438898	1.611	0.563	1.613	0.563	1.538	0.602
rs2886697	1.423	0.661	1.412	0.664	1.368	0.709
rs7613360	0.703	0.679	0.703	0.671	0.798	0.745
rs17718444	1.460	0.626	1.461	0.627	1.444	0.631
rs114142727	1.982	0.138	1.983	0.140	1.968	0.205
rs17576733	1.795	0.434	1.820	0.404	1.788	0.435
rs6537401	0.526	0.639	0.507	0.639	0.608	0.646
rs4916723	0.669	0.679	0.687	0.690	0.707	0.710
rs77960	0.489	0.632	0.489	0.629	0.581	0.678
rs2025286	0.970	0.716	0.952	0.718	0.879	0.688
rs73145587	1.804	0.424	1.812	0.419	1.777	0.466
rs4925811	0.821	0.719	0.846	0.722	0.960	0.720
rs11255890	1.045	0.743	1.020	0.730	0.898	0.712
rs11596214	1.264	0.702	1.247	0.701	1.196	0.674
rs2582895	1.176	0.709	1.178	0.715	1.247	0.671
rs704061	0.966	0.708	1.001	0.705	0.949	0.685
rs76284431	0.271	0.485	0.284	0.491	0.266	0.478
rs1162202	1.330	0.693	1.308	0.696	1.258	0.695
rs76857496	1.778	0.449	1.789	0.441	1.739	0.469
rs11420276	1.474	0.629	1.428	0.640	1.392	0.662
rs5886709	0.831	0.704	0.772	0.697	0.860	0.702
Made careless mistakes	0.992	0.783	1.012	0.778	1.522	0.939
Difficulty sustaining attention	0.615	0.790	0.626	0.800	1.495	0.989
Didn't listen when spoken to	0.537	0.720	0.551	0.730	1.011	0.934
Failed to finish work	0.588	0.740	0.603	0.755	1.202	0.960
Difficulty organizing tasks	0.598	0.730	0.606	0.729	1.285	1.012
Avoided work with sustained focus	0.497	0.701	0.509	0.707	1.027	0.957
Lost necessary things	0.695	0.760	0.698	0.760	1.250	1.009
Easily distracted	0.997	0.896	0.999	0.896	1.906	0.965
Forgetful	0.765	0.782	0.818	0.791	1.317	1.015
Squirmed in seat	1.221	0.971	1.254	0.953	1.957	0.945
Left seat in classroom	0.412	0.719	0.409	0.709	0.987	1.037
Felt restless	0.788	0.806	0.799	0.802	1.551	0.980
Difficult to do fun things quietly	0.633	0.819	0.642	0.826	1.196	1.070
Felt "on the go"	0.956	0.927	0.945	0.929	1.522	1.018
Talked too much	1.140	1.069	1.055	1.034	1.608	1.150
Blurted answers before question complete	0.871	0.870	0.844	0.837	1.255	1.034
Difficulty waiting turn	0.665	0.806	0.639	0.795	1.228	1.040
Felt spiteful/vindictive	0.441	0.689	0.438	0.670	0.720	0.915
DRD4	0.399	0.569	0.404	0.575	0.360	0.529
DAT1	1.500	0.619	1.507	0.612	1.484	0.607
Observations	6,987		1,608		372	

Notes: The above table presents summary statistics for different sub-samples of the Add Health survey. The analysis sample refers to those who agreed to take part in the Wave IV DNA sample. We maintain this restriction in the next two groupings as well. The sibling sample refers to those with at least one full sibling linkage or one twin linkage in Add Health. The diagnosed sample includes adolescents that are coded as having ADHD through at least one of the following ways: taken medication for ADHD in the past 12 months (Wave III) or have had a doctor say they have ADHD (Wave IV). SNPs are measured from 0 to 2. The inattentive and hyperactive measures are retroactively asked about the respondent when they were ages 5-12. Each question asks how often the respondent did the stated activity/pattern and is measured from 0 to 3, with 0 indicating "never or rarely" and 3 indicating "very often." The two additional genes are also measured from 0 to 2.

Table A4: Factor Loadings for Latent Genetic Risk of ADHD

Variable	Factor Loading
rs1438898	−0.024
rs2886697	−0.015
rs7613360	0.020
rs17718444	0.010
rs114142727	−0.010
rs17576733	−0.018
rs6537401	0.034
rs4916723	0.014
rs77960	0.031
rs2025286	−0.020
rs73145587	−0.001
rs4925811	0.035
rs11255890	−0.026
rs11596214	−0.005
rs2582895	0.020
rs704061	0.000
rs76284431	0.010
rs1162202	−0.005
rs76857496	0.003
rs11420276	−0.025
rs5886709	0.016
Made careless mistakes	0.582
Difficulty sustaining attention	0.661
Didn't listen when spoken to	0.612
Failed to finish work	0.680
Difficulty organizing tasks	0.605
Avoided work with sustained focus	0.587
Lost necessary things	0.603
Easily distracted	0.752
Forgetful	0.573
Squirmed in seat	0.546
Left seat in classroom	0.615
Felt restless	0.640
Difficult to do fun things quietly	0.604
Felt "on the go"	0.417
Talked too much	0.484
Blurted answers before question complete	0.473
Difficulty waiting turn	0.643
Felt spiteful/vindictive	0.474
DRD4	−0.012
DAT1	−0.017
Observations	6,987

Notes: The table reports factor loadings for the observed indicators used to construct the latent measure of genetic risk for ADHD. Indicators include 21 single-nucleotide polymorphisms (SNPs) identified in Demontis et al. (2019) and Demontis et al. (2023), the 18 behavioral symptoms used to diagnose ADHD under DSM-5 criteria, and allelic indicators for the DRD4 and DAT1 genes. Factor loadings are obtained via the Bartlett scoring procedure. The factor loadings on individual SNPs are smaller in magnitude than those on the behavioral symptoms, consistent with the well-documented small effect sizes of individual genetic variants on complex neurodevelopmental traits such as ADHD in the genome-wide association literature.

Table A5: Descriptive Statistics—Latent Factor Measures of Parental Investment

	Analysis Sample		Siblings Sample		Diagnosed Sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Gone shopping	0.742	0.438	0.735	0.441	0.691	0.463
Played a sport	0.091	0.287	0.088	0.284	0.091	0.289
Gone to religious service	0.397	0.489	0.427	0.495	0.349	0.477
Talked about dating/party	0.484	0.500	0.463	0.499	0.484	0.500
Gone to movie, play, etc.	0.270	0.400	0.272	0.445	0.315	0.465
Had talk about personal problem	0.403	0.490	0.386	0.487	0.409	0.492
Had serious argument about behavior	0.343	0.475	0.305	0.460	0.435	0.496
Talked about school work/grades	0.641	0.480	0.659	0.474	0.629	0.484
Worked on project for school	0.135	0.342	0.142	0.349	0.164	0.371
Talked about other things in school	0.552	0.497	0.563	0.496	0.554	0.498
Observations	6,987		1,608		372	

Notes: The above table presents summary statistics for different sub-samples of the Add Health survey. The analysis sample refers to those who agreed to take part in the Wave IV DNA sample. We maintain this restriction in the next two groupings as well. The sibling sample refers to those with at least one full sibling linkage or one twin linkage in Add Health. The diagnosed sample includes adolescents that are coded as having ADHD through at least one of the following ways: taken medication for ADHD in the past 12 months (Wave III) or have had a doctor say they have ADHD (Wave IV). Investment measures are binary.

Table A6: Within-Family Estimates for Parental Investment, by Latent Factor Measure

	(1)	(2)	(3)	(4)	(5)	(6)
Latent Genetic Endowment	−0.081	−0.081	−0.103	−0.045	−0.071	−0.068
	(0.047)	(0.050)	(0.045)	(0.044)	(0.064)	(0.062)
Observations	1,608	1,608	1,608	1,608	1,608	1,608
Various Components of the Latent Genetic Endowment Measure:						
SNPs	✓	✓	✓	✓	✓	✓
DAT1 and DRD4	✓	✗	✓	✓	✓	✗
Hyperactivity Phenotypes	✓	✓	✗	✓	✗	✗
Inattentiveness Phenotypes	✓	✓	✓	✗	✗	✗

Notes: This table reports within-family fixed effects regression estimates, where the dependent variable is latent parental investment and the key independent variable is the genetic endowment measure associated with ADHD risk. The included controls are fixed across columns. Column (1) reproduces the results using our preferred specification. Column (2) omits the DAT1 and DRD4 gene realizations. Columns (3) and (4) omit the stated subset of phenotypes. Column (5) omits all phenotypes.

Table A7: Within-Family Estimates for Parental Investment and Latent Genetic Endowment for ADHD

	(1)	(2)	(3)	(4)
Panel A: Total Investment				
Full Sample	−0.049	−0.047	−0.081	−0.081
	(0.044)	(0.044)	(0.047)	(0.047)
Non-Black Sample	−0.075	−0.073	−0.119	−0.120
	(0.047)	(0.047)	(0.049)	(0.050)
Black Sample	0.061	0.069	0.125	0.115
	(0.116)	(0.111)	(0.109)	(0.109)
Panel B: Emotional Investment				
Full Sample	−0.055	−0.044	−0.088	−0.088
	(0.053)	(0.052)	(0.056)	(0.056)
Non-Black Sample	−0.073	−0.066	−0.122	−0.124
	(0.058)	(0.057)	(0.061)	(0.061)
Black Sample	0.021	0.044	0.093	0.079
	(0.124)	(0.111)	(0.115)	(0.116)
Panel C: Activities Investment				
Full Sample	−0.026	−0.039	−0.055	−0.055
	(0.057)	(0.058)	(0.058)	(0.057)
Non-Black Sample	−0.064	−0.070	−0.090	−0.089
	(0.064)	(0.064)	(0.063)	(0.063)
Black Sample	0.134	0.106	0.149	0.150
	(0.133)	(0.135)	(0.130)	(0.131)
Cognitive Control	✓	✓	✓	✓
Mental Health Control	✗	✓	✓	✓
Demographic Controls	✗	✗	✓	✓
Twin Control	✗	✗	✗	✓

Notes: This table presents within-family fixed effects regression results with latent parental investment as the dependent variable and our latent measure of (negated) ADHD risk as the independent variable of interest. "Cognitive Control" refers to the latent factor for cognitive skills, measured in Wave I. "Mental Health Control" refers to the latent factor for depression (mental health), measured in Wave I. "Demographic Controls" include gender, age, and a quadratic term for age. "Twin Control" includes an indicator for whether the respondent is a twin. Standard errors are clustered at the family level. The full sample contains 1,608 observations while the non-Black sample contains 1,304 observations and the Black sample contains 304 observations.

Table A8: Within-Family Estimates for Parental Investment and ADHD PGS

	(1)	(2)	(3)	(4)
Panel A: Total Investment				
Full Sample	−0.007	−0.008	−0.003	−0.003
	(0.070)	(0.070)	(0.071)	(0.071)
Non-Black Sample	0.070	0.069	0.083	0.083
	(0.077)	(0.077)	(0.079)	(0.079)
Black Sample	−0.337	−0.336	−0.352	−0.373
	(0.165)	(0.166)	(0.170)	(0.169)
Panel B: Emotional Investment				
Full Sample	0.011	0.011	0.023	0.024
	(0.078)	(0.078)	(0.078)	(0.079)
Non-Black Sample	0.069	0.068	0.088	0.089
	(0.090)	(0.090)	(0.092)	(0.092)
Black Sample	−0.213	−0.209	−0.208	−0.236
	(0.155)	(0.152)	(0.155)	(0.154)
Panel C: Activities Investment				
Full Sample	−0.028	−0.028	−0.040	−0.040
	(0.080)	(0.080)	(0.079)	(0.080)
Non-Black Sample	0.065	0.066	0.062	0.061
	(0.085)	(0.085)	(0.084)	(0.084)
Black Sample	−0.472	−0.478	−0.522	−0.524
	(0.225)	(0.222)	(0.223)	(0.223)
Cognitive Control	✓	✓	✓	✓
Mental Health Control	✗	✓	✓	✓
Demographic Controls	✗	✗	✓	✓
Twin Control	✗	✗	✗	✓

Notes: This table presents within-family fixed effects regression results with latent parental investment as the dependent variable and the (negated) ADHD PGS from Demontis et al. (2019) as the independent variable of interest. "Cognitive Control" refers to the latent factor for cognitive skills, measured in Wave I. "Mental Health Control" refers to the latent factor for depression (mental health), measured in Wave I. "Demographic Controls" include gender, age, and a quadratic term for age. "Twin Control" includes an indicator for whether the respondent is a twin. Standard errors are clustered at the family level. The full sample contains 1,608 observations while the non-Black sample contains 1,304 observations and the Black sample contains 304 observations.

Table A9: Pooled Estimates by Investment Type and Family Type

	Investment		Investment (Activities)		Investment (Emotional)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Analysis Sample						
Latent Factor Score	−0.036		−0.050		−0.029	
	(0.016)		(0.020)		(0.017)	
Polygenic Score		0.035		0.019		0.036
		(0.015)		(0.019)		(0.016)
Observations	6,987		6,987		6,987	
Panel B: Black Sample						
Latent Factor Score	−0.076		−0.075		−0.071	
	(0.037)		(0.046)		(0.038)	
Polygenic Score		0.074		0.034		0.079
		(0.033)		(0.041)		(0.035)
Observations	1,490		1,490		1,490	
Panel C: Non-Black Sample						
Latent Factor Score	−0.025		−0.043		−0.017	
	(0.017)		(0.022)		(0.019)	
Polygenic Score		0.023		0.013		0.023
		(0.017)		(0.021)		(0.018)
Observations	5,497		5,497		5,497	

Notes: This table presents OLS regression results with the dependent variable the stated type of investment and the variable of interest either the latent factor or polygenic score for ADHD. Each regression includes controls for child age, age squared, gender, race, initial skills, initial mental health, twin status, and mother's education. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Panels B and C present the same results for Black and non-Black children, respectively. These regressions omit race controls. Robust standard errors are presented in parentheses.

Table A10: Within-Family Estimates by Investment Type and Family Type

	Investment		Investment (Activities)		Investment (Emotional)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Analysis Sample						
Latent Factor Score	−0.081		−0.055		−0.088	
	(0.047)		(0.057)		(0.056)	
Polygenic Score		−0.004		−0.041		0.022
		(0.071)		(0.079)		(0.079)
Observations	1,608		1,608		1,608	
Panel B: Black Sample						
Latent Factor Score	0.115		0.150		0.079	
	(0.109)		(0.131)		(0.116)	
Polygenic Score		−0.373		−0.524		−0.236
		(0.169)		(0.223)		(0.154)
Observations	304		304		304	
Panel C: Non-Black Sample						
Latent Factor Score	−0.120		−0.089		−0.124	
	(0.050)		(0.063)		(0.061)	
Polygenic Score		0.083		0.061		0.089
		(0.079)		(0.084)		(0.092)
Observations	1,304		1,304		1,304	

Notes: This table presents OLS regression results with the dependent variable the stated type of investment and the variable of interest either the latent factor or polygenic score for ADHD. Each regression includes controls for child age, age squared, gender, initial skills, initial mental health, and twin status. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Panels B and C presents the same results for Black and non-Black children, respectively. These regressions omit race and maternal controls. Robust standard errors clustered at the family level are presented in parentheses.

Table A11: OLS Estimates of ADHD on Long-Run Outcomes

Panel A: Education/Labor Measures								
	Future Skills		Education (Years)		Employed		(Log) Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latent Factor Score	0.050		0.162		0.001		0.065	
	(0.011)		(0.023)		(0.003)		(0.031)	
Polygenic Score		0.022		0.167		0.005		0.070
		(0.010)		(0.022)		(0.003)		(0.031)
Observations	4,949		6,987		6,957		6,739	
Panel B: Social Measures								
	Ever Married		Ever Arrested					
	(1)	(2)	(3)	(4)				
Latent Factor Score	0.001		−0.038					
	(0.006)		(0.006)					
Polygenic Score		−0.009		−0.016				
		(0.006)		(0.005)				
Observations	6,983		6,950					
Panel C: Health Measures								
	ADHD Diagnosis		Future Depression		Illicit Drug Use			
	(1)	(2)	(3)	(4)	(5)	(6)		
Latent Factor Score	−0.059		0.078		−0.110			
	(0.004)		(0.013)		(0.015)			
Polygenic Score		−0.007		0.027		−0.024		
		(0.003)		(0.012)		(0.013)		
Observations	6,987		5,480		6,981			

NOTES: Each regression includes controls for child age, age squared, gender, race, initial skills, initial mental health, twin status, and maternal education. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Robust standard errors are presented in parentheses. Future skills and future depression are measured in the Wave II, illicit drug use is a latent factor score that uses multiple measures of illicit drug use in Wave III, and all other outcomes are measured by Wave IV.

Table A12: Within-Family Estimates of ADHD on Long-Run Outcomes for Black Families

Panel A: Education/Labor Measures								
	Future Skills		Education (Years)		Employed		(Log) Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latent Factor Score	0.107		0.311		0.026		0.195	
	(0.078)		(0.149)		(0.025)		(0.245)	
Polygenic Score		0.087		−0.005		−0.024		−0.459
		(0.075)		(0.171)		(0.047)		(0.538)
Observations	246		304		302		283	
Panel B: Social Measures								
	Ever Married		Ever Arrested					
	(1)	(2)	(3)	(4)				
Latent Factor Score	0.037		−0.072					
	(0.053)		(0.042)					
Polygenic Score		0.020		0.096				
		(0.070)		(0.051)				
Observations	304		302					
Panel C: Health Measures								
	ADHD Diagnosis		Future Mental Health		Illicit Drug Use			
	(1)	(2)	(3)	(4)	(5)	(6)		
Latent Factor Score	−0.039		0.158		0.019			
	(0.018)		(0.122)		(0.096)			
Polygenic Score		−0.030		0.092		0.082		
		(0.032)		(0.140)		(0.065)		
Observations	304		277		304			

Notes: Each regression includes controls for child age, age squared, gender, initial skills, initial mental health, and twin status. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Robust standard errors clustered at the family level are presented in parentheses. Future skills and future depression are measured in the Wave II, illicit drug use is a latent factor score that uses multiple measures of illicit drug use in Wave III, and all other outcomes are measured by Wave IV.

Table A13: Within-Family Estimates of ADHD on Long-Run Outcomes for Non-Black Families

Panel A: Education/Labor Measures								
	Future Skills		Education (Years)		Employed		(Log) Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latent Factor Score	0.003		0.030		0.001		0.079	
	(0.038)		(0.082)		(0.013)		(0.133)	
Polygenic Score		−0.036		0.118		0.019		0.169
		(0.052)		(0.125)		(0.018)		(0.186)
Observations	1,069		1,304		1,300		1,273	
Panel B: Social Measures								
	Ever Married		Ever Arrested					
	(1)	(2)	(3)	(4)				
Latent Factor Score	0.013		−0.034					
	(0.023)		(0.020)					
Polygenic Score		−0.003		−0.004				
		(0.028)		(0.027)				
Observations	1,303		1,302					
Panel C: Health Measures								
	ADHD Diagnosis		Future Depression		Illicit Drug Use			
	(1)	(2)	(3)	(4)	(5)	(6)		
Latent Factor Score	−0.072		0.018		−0.134			
	(0.013)		(0.038)		(0.055)			
Polygenic Score		−0.026		0.081		0.043		
		(0.014)		(0.056)		(0.069)		
Observations	1,304		1,231		1,304			

Notes: Each regression includes controls for child age, age squared, gender, initial skills, initial mental health, and twin status. The polygenic score is provided by Add Health and is derived from Demontis et al. (2019). Robust standard errors clustered at the family level are presented in parentheses. Future skills and future depression are measured in the Wave II, illicit drug use is a latent factor score that uses multiple measures of illicit drug use in Wave III, and all other outcomes are measured by Wave IV.

Table A14: Group-Level Differences in ADHD Latent Factor Measures

	SNP Count (1)	Hyperactive Count (2)	Inattentive Count (3)
Black Female	1.201	1.378	0.911
Black Male	1.204	1.906	1.472
Non-Black Female	1.153	1.716	0.955
Non-Black Male	1.142	2.216	1.716

Notes: The above table displays the group-level means for the total counts of each type of measure used in the construction of our latent ADHD factor score. Column (1) denotes the average number of SNPs that an adolescent has across the 21 relevant SNPs obtained from Demontis et al. (2019) and Demontis et al. (2023). The hyperactive and inattentive counts refer to the total number of the nine respective behaviors included in the retrospective ADHD questionnaire that the respondent answered "Often" or "Very Often" to.

Table A15: Initial Condition Parameters by Race and Gender Estimated Outside the Model

	Black Female (1)	Non-Black Female (2)	Black Male (3)	Non-Black Male (4)
Panel A: Means				
Cognitive Skills (μ_C)	−0.108	0.231	−0.519	−0.108
	(0.071)	(0.037)	(0.081)	(0.044)
Mental Health (μ_H)	−0.443	0.034	−0.177	0.180
	(0.091)	(0.036)	(0.113)	(0.045)
Genetic Endowment (μ_g)	0.194	0.113	−0.325	−0.342
	(0.078)	(0.045)	(0.089)	(0.049)
Panel B: Variance-Covariance				
Variance of Cognitive Skills (σ_C^2)	0.767	0.968	0.746	1.048
	(0.076)	(0.049)	(0.096)	(0.044)
Covariance of Cognitive Skills and Mental Health ($\sigma_{C,H}$)	0.212	0.374	0.148	0.203
	(0.074)	(0.045)	(0.088)	(0.048)
Covariance of Cognitive Skills and Genetic Endowment ($\sigma_{C,g}$)	0.070	0.266	0.199	0.275
	(0.088)	(0.044)	(0.123)	(0.055)
Variance of Mental Health (σ_H^2)	1.250	1.057	1.144	0.671
	(0.147)	(0.079)	(0.209)	(0.084)
Covariance of Mental Health and Genetic Endowment ($\sigma_{H,g}$)	0.128	0.257	0.196	0.189
	(0.081)	(0.066)	(0.127)	(0.053)
Variance of Genetic Endowment (σ_g^2)	0.836	0.998	1.091	1.171
	(0.129)	(0.090)	(0.174)	(0.086)

Notes: The above table reports the unconditional distribution of initial (log) skills by race and gender. Panel A reports group-specific means while Panel B reports pairwise covariances. Each entry is the group-specific value obtained from the original analytic sample in Add Health. Standard errors in parentheses are the standard deviations across 100 bootstrap replications of the Add Health data.

Table A16: Initial Condition Parameters by Race and Gender Estimated Inside the Model (SMM)

	Black Female (1)	Non-Black Female (2)	Black Male (3)	Non-Black Male (4)
Panel A: Means				
Unobserved Heterogeneity Mean (μ_ψ)	0 (normalized)	0.635 (0.133)	0.120 (0.019)	-0.589 (0.072)
Panel B: Variance and Correlation of Family-Level Heterogeneity with Endowment Skills and Genes (common across groups)				
Correlation of Cognitive Skill and Family Unobserved Heterogeneity ($\rho_{C,\psi}$)		0.633 (0.032)		
Correlation of Mental Health and Family Unobserved Heterogeneity ($\rho_{H,\psi}$)		0.426 (0.031)		
Correlation of Genetic Endowment and Family Unobserved Heterogeneity ($\rho_{g,\psi}$)		0.660 (0.044)		
Variance of Family Unobserved Heterogeneity (σ^2_ψ)		1 (normalized)		

Notes: The table reports estimated race-by-gender group-specific means for the unobserved family-level heterogeneity (μ_ψ) and scalar correlations ($\rho_{\ell,\psi}$) between trait ℓ and the unobserved family-level heterogeneity. These correlations are used to construct the covariance terms $\sigma^{(r,s)}_{C,\psi}$, $\sigma^{(r,s)}_{H,\psi}$, $\sigma^{(r,s)}_{g,\psi}$. The mean for the unobserved family-level heterogeneity for Black female adolescents is normalized to 0. Standard errors are obtained from 100 family-clustered non-parametric bootstraps.

Table A17: Model Fit – Evolution of Skills Across Adolescence

	Data (1)	Simulation (2)
Cognitive Skills		
Intercept	−0.105	−0.100
Slope	−0.001	−0.034
Mental Health		
Intercept	−0.082	−0.090
Slope	−0.046	−0.034

Notes: The above table compares the first four targeted moments in our Simulated Method of Moments (SMM) procedure. These moments are obtained via pooled regressions with each skill type as the dependent variable and adolescents' age as the independent variable. Simulation values refer to the moments generated at the model solution.

Table A18: Model Fit – Auxiliary Regressions for Parental Investment

	Data (1)	Simulation (2)
Panel A: Pooled Specification		
Cognitive Skills	0.204	0.242
Mental Health	0.051	0.040
Genetic Endowment	−0.035	−0.029
Panel B: Within-Family Specification		
Cognitive Skills	0.157	0.086
Mental Health	−0.002	0.011
Genetic Endowment	−0.081	−0.072

Notes: The above table compares the six targeted moments in our Simulated Method of Moments (SMM) procedure related to investment decisions. The moments in Panel A are derived from a pooled specification of the estimating equation, whereas those in Panel B are derived from a within-family version. Both specifications feature parental investment as the dependent variable, with cognitive skills, mental health, genetic endowment, race, gender, age, and age squared as the independent variables. Within-family regressions omit race controls. Simulation values refer to the moments generated at the model solution.

Table A19: Model Fit – Auxiliary Regressions for the Technology of Skill Formation (Cognitive Skills)

	Data (1)	Simulation (2)
Panel A: Pooled Specification		
Cognitive Skills	0.628	0.590
Mental Health	0.044	0.040
Genetic Endowment	0.061	0.114
Investment	0.031	0.118
Investment × Cognitive Skills	0.020	0.023
Investment × Mental Health	0.003	0.015
Investment × Genetic Endowment	−0.017	−0.025
Cognitive Skills × Mental Health	0.042	−0.012
Cognitive Skills × Genetic Endowment	0.043	0.067
Mental Health × Genetic Endowment	0.002	−0.000
Panel B: Within-Family Specification		
Cognitive Skills	0.450	0.513
Mental Health	0.044	0.038
Genetic Endowment	0.028	0.069
Investment	0.058	0.082
Cognitive Skills × Genetic Endowment	0.040	0.067

Notes: The above table compares the fifteen targeted moments in our Simulated Method of Moments (SMM) procedure related to the technology of skill formation for cognitive skills. The moments in Panel A are derived from a pooled specification of the estimating equation, whereas those in Panel B are derived from a within-family version. Both specifications feature next period cognitive skills as the dependent variable, with the listed variables, as well as race, gender, age, and age squared, as the independent variables. Within-family regressions omit race controls. Simulation values refer to the moments generated at the model solution.

Table A20: Model Fit – Auxiliary Regressions for the Technology of Skill Formation (Mental Health)

	Data (1)	Simulation (2)
Panel A: Pooled Specification		
Cognitive Skills	0.072	0.032
Mental Health	0.533	0.507
Genetic Endowment	0.082	0.092
Investment	0.026	0.018
Investment × Cognitive Skills	0.007	0.010
Investment × Mental Health	−0.011	−0.038
Investment × Genetic Endowment	−0.015	−0.034
Cognitive Skills × Mental Health	0.011	0.005
Cognitive Skills × Genetic Endowment	−0.003	0.000
Mental Health × Genetic Endowment	0.015	0.018
Panel B: Within-Family Specification		
Cognitive Skills	0.077	0.075
Mental Health	0.460	0.496
Genetic Endowment	0.064	0.099
Investment	0.009	0.044
Mental Health × Genetic Endowment	0.072	0.014

Notes: The above table compares the fifteen targeted moments in our Simulated Method of Moments (SMM) procedure related to the technology of skill formation for mental health. The moments in Panel A are derived from a pooled specification of the estimating equation, whereas those in Panel B are derived from a within-family version. Both specifications feature next period mental health as the dependent variable, with the listed variables, as well as race, gender, age, and age squared, as the independent variables. Within-family regressions omit race controls. Simulation values refer to the moments generated at the model solution.

Table A21: Model Fit – Heterogeneity in Investment

	Data	Simulation
	(1)	(2)
Female	0.404	0.197
Black	0.080	0.105
Female × Black	−0.173	−0.247

Notes: The above table compares the three targeted moments in our Simulated Method of Moments (SMM) procedure related to investment heterogeneity on the basis of gender and race. The moments are derived from a pooled specification of the estimating equation that features parental investment as the dependent variable, with the listed variables as the independent variables. Simulation values refer to the moments generated at the model solution.

Table A22: Auxiliary Regression Coefficients, by Genetic Measure

	Pooled		Within-Family	
	Latent Factor (1)	PGS (2)	Latent Factor (3)	PGS (4)
Panel A: Endogenous Investment Decisions				
Cognitive Skills	0.204	0.191	0.157	0.139
Mental Health	0.051	0.045	−0.002	−0.015
Genetic Endowment	−0.035	0.042	−0.081	−0.003
Panel B: Technology of Skill Formation (Cognitive)				
Cognitive Skills	0.628	0.637	0.450	0.451
Mental Health	0.044	0.055	0.044	0.050
Genetic Endowment	0.061	0.024	0.028	−0.027
Investment	0.031	0.029	0.058	0.060
Investment × Cognitive Skills	0.020	0.016	—	—
Investment × Mental Health	0.003	0.000	—	—
Investment × Genetic Endowment	−0.017	0.003	—	—
Cognitive Skills × Mental Health	0.042	0.049	—	—
Cognitive Skills × Genetic Endowment	0.043	0.027	0.040	−0.038
Mental Health × Genetic Endowment	0.002	−0.010	—	—
Panel C: Technology of Skill Formation (Mental Health)				
Cognitive Skills	0.071	0.088	0.077	0.101
Mental Health	0.533	0.544	0.460	0.443
Genetic Endowment	0.082	0.025	0.064	0.080
Investment	0.026	0.024	0.009	0.010
Investment × Cognitive Skills	0.007	0.003	—	—
Investment × Mental Health	−0.011	−0.014	—	—
Investment × Genetic Endowment	−0.015	0.004	—	—
Cognitive Skills × Mental Health	0.011	0.018	0.072	0.053
Cognitive Skills × Genetic Endowment	−0.003	−0.004	—	—
Mental Health × Genetic Endowment	0.015	−0.005	—	—

Notes: The above table compares the targeted moments used in the Simulated Method of Moments that depend, in part, on the measure of genetic risk. Column (1) displays the values for our preferred specification whereas Column (2) displays the same when relying on the polygenic score in Demontis et al. (2019). Pooled regressions include controls for gender, race, age, and age squared. Within-family regressions omit race controls.

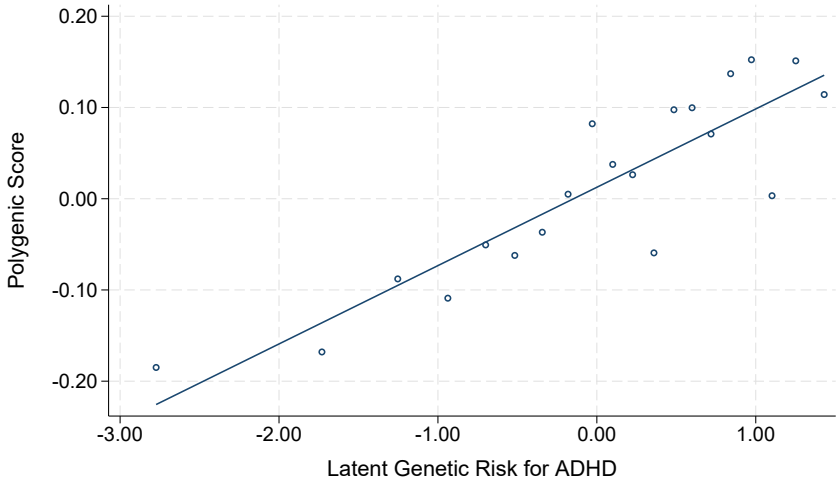
Table A23: Additional Parameter Estimates by Genetic Measure

	Latent Factor (1)	PGS (2)
Mean for Non-Black Males ($\mu_{m,nb}$)	0.414	0.479
Mean for Black Females ($\mu_{f,b}$)	1.121	0.924
Mean for Black Males ($\mu_{m,b}$)	−0.384	−0.307
Correlation with Family Effect for Non-Black Males ($\rho_{m,nb}$)	0.424	0.545
Correlation with Family Effect for Black Females ($\rho_{f,b}$)	0.128	0.217
Correlation with Family Effect for Black Males ($\rho_{m,b}$)	0.545	0.574

Notes: The above table shows the additional parameter estimates of the model, including the race-by-gender-specific means ($\mu_{r,s}$) and scalar correlations ($\rho_{r,s}$) between initial conditions and the unobserved family fixed effect. Standard errors in parentheses are obtained from 100 family-clustered non-parametric bootstraps of the estimation procedure.

B Additional Figures

Figure B1: Comparison of PGS and Latent Factor Measure of Genetic Risk for ADHD



Notes: This figure displays a binscatter plot of the Demontis et al. (2019) polygenic score (PGS) against our latent factor measure of genetic risk for ADHD. Each bin represents an equal number of observations. The weak correlation between the two measures is consistent with our argument that the PGS captures only a fraction of heritable ADHD risk, while our latent factor incorporates both SNP-level genotypes and the 18 retrospective behavioral symptoms used to diagnose ADHD under DSM-5 criteria, thereby recovering a broader share of the underlying genetic risk.

C Diagnostic Criteria for ADHD

Below is a reproduction of the corresponding diagnostic criteria outlined in the *Diagnostic and Statistical Manual of Mental Disorders: DSM-V* (American Psychiatric Association, 2013).

1. A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as characterized by (1) and/or (2):

   (a) *Inattention*: Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities: *Note*: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.

      i. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).

      ii. Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).

      iii. Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).

      iv. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).

      v. Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).

      vi. Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).

      vii. Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).

      viii. Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).

      ix. Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).

(b) *Hyperactivity and impulsivity*: Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities: *Note*: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or a failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.

i. Often fidgets with or taps hands or feet or squirms in seat.

ii. Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).

iii. Often runs about or climbs in situations where it is inappropriate. (*Note*: In adolescents or adults, may be limited to feeling restless.)

iv. Often unable to play or engage in leisure activities quietly.

v. Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).

vi. Often talks excessively.

vii. Often blurts out an answer before a question has been completed (e.g., completes people's sentences; cannot wait for turn in conversation).

viii. Often has difficulty waiting his or her turn (e.g., while waiting in line).

ix. Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

2. Several inattentive or hyperactive-impulsive symptoms were present prior to age 12 years.

3. Several inattentive or hyperactive-impulsive symptoms are present in two or more settings (e.g., at home, school, or work; with friends or relatives; in other activities).

4. There is clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning.

5. The symptoms do not occur exclusively during the course of schizophrenia or another psychotic disorder and are not better explained by another mental disorder (e.g., mood disorder, anxiety disorder,

dissociative disorder, personality disorder, substance intoxication or withdrawal).