American Economic Review 2024, 114(2): 385–425
https://doi.org/10.1257/aer.20220456

The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills[†]

By Mikkel Aagaard Houmark, Victor Ronda, and Michael Rosholm*

This paper studies the interplay between genetics and family investments in the process of skill formation. We model and estimate the joint evolution of skills and parental investments throughout early childhood. We document three genetic mechanisms: the direct effect of child genes on skills, the indirect effect of child genes via parental investments, and family genetic influences captured by parental genes. We show that genetic effects are dynamic, increase over time, and operate via environmental channels. Our paper highlights the value of integrating biological and social perspectives into a single unified framework. (*JEL* I24, I26, J12, J13, J24)

It is widely recognized that individuals have different abilities, that differences in individuals' abilities surface early in life, and that early differences in abilities explain variation in socioeconomic outcomes later in life (see, e.g., Cunha, Heckman, and Navarro 2005; Heckman and Mosso 2014). It is also well understood that inequality in family resources translates into inequality in children's outcomes and that early skills are partly determined by the genotype realized at conception (Polderman et al. 2015; Plomin and von Stumm 2018; Silventoinen et al. 2020; Cesarini and Visscher 2017; Branigan, McCallum, and Freese 2013). This substantial body of work documents the relative importance of genetics and family resources for skill formation. However, the two are often treated as separate factors, where the higher importance of one implies a lesser role of the other. Such a framework tends to overlook how genes and family resources are closely interrelated during skill formation.

* Houmark: Department of Economics and Business Economics and TrygFonden's Centre for Child Research, Aarhus University (email: mhoumark@econ.au.dk); Ronda: TrygFonden's Centre for Child Research, Aarhus University (email: victor.ronda.econ@gmail.com); Rosholm: Department of Economics and Business Economics and TrygFonden's Centre for Child Research, Aarhus University and IZA (email: rom@econ.au.dk). Chinhui Juhn was the coeditor for this article. We gratefully acknowledge helpful comments from Phillip Heiler, Elena Mattana, Helena Skyt Nielsen, Nicholas Papageorge, and seminar participants at Aarhus University, the Society of Labor Economists Annual Meeting, the MRC Integrative Epidemiology Unit at the University of Bristol, Conference on Genes, Social Mobility, and Inequalities across the Life-Course, the Center for Economic and Social Research, University of Southern California, the Demography Workshop at the University of Chicago, and the Integrating Genetics and the Social Sciences (IGSS) Conference. This research was funded by a seed money grant from TrygFonden's Centre for Child Research. We are extremely grateful to all the families who took part in this study; the midwives for their help in recruiting them; and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, and they will serve as guarantors for the contents of this paper.

[†] Go to https://doi.org/10.1257/aer.20220456 to visit the article page for additional materials and author disclosure statements.

We incorporate genetics into a model of skill formation during early childhood (ages zero to seven years) and document how the interplay between genes and family resources arises from two mechanisms. First, we show that parents respond to their children's genotype by investing more in children with higher genetic factors. We call this the *nurture of nature* effect. Second, we show that parents who themselves have higher genetic factors also invest more resources in their children. Analogously, this mechanism reflects the nature of nurture, which we proceed by calling *family genetic associations* for clarity. By incorporating these two mechanisms into an economic model, we formalize ideas in the genetics literature going back to at least Plomin, DeFries, and Loehlin (1977). The formal treatment of genetics in an economic model of skill formation allows us to understand better the role of genes in the skill formation process and to learn what is missed from excluding genes from such models.

The empirical estimation relies on detailed genetic and survey data from a longitudinal British dataset, the Avon Longitudinal Study of Parents and Children (ALSPAC)(Boyd et al. 2012; Fraser et al. 2012). We observe detailed molecular genetic data from the child participants and both of their parents. We identify genetic influences using multiple polygenic indexes related to educational attainment and cognitive performance. Variations of these indexes are widely used in the literature and have been shown to predict a wide range of economic and social outcomes, including early childhood skills (Belsky et al. 2016), school achievement (Ward et al. 2014), educational attainment (Rietveld et al. 2013; Domingue et al. 2015; Okbay et al. 2016; Lee et al. 2018; Ronda et al. 2022), as well as earnings and wealth (Papageorge and Thom 2019; Belsky et al. 2018; Barth, Papageorge, and Thom 2020).

A particular challenge in this literature is identifying the independent effect of an individual's genes from that of her family. To overcome this challenge, we explicitly control for parental genes in the model. Conditional on parental genes, variation in children's genes is random, allowing us to identify the independent effect of the child's genes on skill formation. This is important. Although genes are fixed at conception and thus predetermined, they are not exogenous. Rather, they are determined entirely by parental genes, which also affect the environment in which the child grows up and forms its skills (see, e.g., Kong et al. 2018; Young et al. 2018). Thus, any observed association between genes and socioeconomic outcomes may be partially driven by the individual's childhood environment. One common solution to this challenge is to control for observable differences in family environment as in Barth, Papageorge, and Thom (2020) and Papageorge and Thom (2019). Another is to exploit random genetic variation across siblings as in Ronda et al. (2022). Our approach allows us to directly control for the influence of parental genes on children's skills, whether or not this works through observable aspects of the family environment. In addition, our approach allows us to document the importance of family genetic influences captured by parental genes, which is of interest on its own.

Another challenge is measurement error. Because child skills, parental investments, and the underlying genetic factors are all latent variables that are imperfectly measured, correlations of observed proxies only provide a biased signal of the underlying relationship that we are interested in. To identify genetic influences on skill formation, we, therefore, incorporate genetics into a dynamic latent factor

model, as in Cunha and Heckman (2008). As far as we are aware, we are the first to do so. The model allows us to control for measurement error and identify the genetic factor, latent skills, and investments. In addition, it allows us to decompose the different genetic mechanisms and compare genetic influences across child ages. We rely on multiple measures of skills and investments collected by the ALSPAC and standard latent factor model assumptions to identify the underlying skills and investments in each period. We extend the traditional model to also control for measurement error in genetic factors and identify the underlying latent genetic factor.

Our approach allows us to gain additional insights into the process of skill formation. We find that genetic influences accumulate over time and gradually increase over the early childhood period. Genetic influences on initial skills are small, but by ages six to seven, a one standard deviation increase in the child's genetic factor leads to almost a 0.2 standard deviation increase in skills. This pattern is consistent with earlier findings on the increasing importance of genes over the life-span (Bouchard 2013; Tucker-Drob, Briley, and Harden 2013; Tucker-Drob and Briley 2014; Belsky et al. 2016). Unlike previous work, our approach allows us to rule out several potential sources of bias, including confounding from the environment and differences in measurement error over time, and, at the same time, to gain additional insight into why this pattern appears. We find that the increase is due to two main mechanisms. First, conditional on their current stock of skills and parental investments, genetics make some children better able to retain and acquire new skills, *the direct effect of genes*. Second, parents reinforce initial genetic differences by investing more in children with higher genetic factors and higher stock of skills, *the nurture of nature effect*.[1] We show that the second mechanism is more important at early ages, and the first is more important at later ages.

We also document a strong association between parents' genes and children's skills. This association captures the effect of family genetics and unobserved environmental factors correlated with parental genes on the environment experienced by the child. These influences explain between 40 and 82 percent of the association between the child's genetic factor and her skills. We show that these genetic influences are completely mediated by parental educational attainment. This suggests that controlling for parental education may be enough to capture the family background in analyses of child development that have access to a child's genetic data but not their parents'.

Another key contribution of our paper is documenting what child development models that ignore genetic influences miss. First, we show that neglecting genes leads to an overestimation of the returns to parental investment, although adding parental controls eliminates most of this bias. Second, we identify significant heterogeneity in the returns to investments across the child's genetics that

[1] In addition to this gene-by-environment correlation, child genes could also influence skill formation by what is called gene-by-environment interactions, referring to genetic differences in sensitivity to the environment. A number of recent studies in economics have shown that such genetic differences are quite important for educational attainment, health, and labor market outcomes (see, e.g., Barcellos, Carvalho, and Turley 2018; Papageorge and Thom 2019; Barth, Papageorge, and Thom 2020; Ronda et al. 2022). In our context, this interaction could come about in many different ways. For example, it could come about if genes changed the returns to parental investments in skill formation. In online Appendix D, we consider this possibility. However, we do not find support in the data for these interactions being important during early skill formation.

is not captured by models that ignore genes. This heterogeneity is only partially captured by observable family characteristics. Third, the genetic heterogeneity that is captured in models without genes is misattributed to observable nongenetic factors, such as family income, that do not reflect the underlying causal mechanism. This exercise highlights the importance of incorporating genes into models of skill formation.

We thus contribute to the literature on genetics and skill formation by studying genetic effects in an integrated framework, allowing us to consider multiple potential mechanisms simultaneously and learn more about how each works. While it is well-known that parental genes matter for children's development beyond the genes transmitted to the children (Kong et al. 2018; Ronda et al. 2022; Wertz et al. 2020), we show directly that parental genes affect how much parents invest in their children. And while others have investigated whether parental investments tend to be reinforcing or compensating with respect to early childhood skills or birth endowments (with some conflicting findings; see, e.g., Hsin 2012; Grätz and Torche 2016; Nicoletti and Tonei 2020), we relate such behavior directly to the child's genes, which are not influenced by later investments.[2] Our approach allows us to simultaneously identify the influence of genes via the nurture of nature and family genetics channels and show that both are important. Moreover, we show the importance of considering genetic influences for the identification of structural parameters in models of child development. Otherwise, genetic influences and genetic heterogeneity will be misattributed to nongenetic factors.

The paper is organized as follows. In Section I, we outline the theoretical framework describing the various channels through which genes influence skill accumulation. In Section II, we introduce the ALSPAC dataset, discuss our measures of the genetic factor, and conduct a preliminary descriptive analysis. In Section III, we describe our empirical model, including the measurement system used to identify the genetic factor, latent skills, and investments, along with the estimation procedure. Section IV presents our main results. In Section V, we study mechanisms and highlight the implications of our findings for our understanding of the skill formation process. Section VI offers brief concluding remarks, discusses the limitations of our work, and makes suggestions for future research.

I. Conceptual Framework

We incorporate genetics into a model of early skill formation in the spirit of Cunha and Heckman (2007, 2008). The model allows us to identify different channels through which the child and parental genetics influence the process of skill formation. The model considers a family with a single child and two parents. Thus, for simplicity, we abstract from the influence of siblings.[3] We model the evolution

[2] In a descriptive analysis, Breinholt and Conley (2020) document that parental investments are positively related to children's polygenic index for educational attainment. Sanz-de Galdeano and Terskaya (2019) also investigate whether parental investments are compensating or reinforcing but do so by comparing siblings in a static framework. We go a step further by incorporating the genetics of both children and parents into a structural model to quantify the different mechanisms of genetic influence and how they develop over time.

[3] The conceptual model could quite easily be extended to consider families with several children, but data limitations prevent us from learning much empirically by doing so. It would be interesting for future research to utilize genetic information on families with multiple children.

of skills from birth ($t = 0$) until the end of the child's early development in period T. Skills are complex traits jointly determined by the child's genetic makeup and interactions and experiences determined by parents, which we refer to as parental investments.

We highlight three main mechanisms relating genes to skills. First, genes can have a direct effect on skills, capturing individual heterogeneity in learning ability. Second, the child's genes may influence skill formation via its effect on parental investments. We call this mechanism the nurture of nature effect. It may capture the fact that children with different genetic makeup are more likely to induce parents to invest more in them, e.g., by enjoying being read stories or enjoying cognitively stimulating play. Alternatively, it may also capture parental compensating or reinforcing responses to initial endowments, which are genetically determined. Third, the investment decision will also depend on parental genotype. We refer to this mechanism as family genetic associations.[4] It captures genetic differences in the quantity and quality of parental interactions with their children.

We describe the three mechanisms in more detail in the following sections. We start by describing what we mean by genetic endowments and how such endowments are inherited from parents to children in Section IA. In Section IB, we formally describe how children's skills evolve as a function of parental investments and genetic factors. In Section IC, we describe how parental investments are determined. Lastly, in Section ID, we summarize the three main mechanisms through which genes can influence the process of skill formation.

A. *The Genotype*

The child's genotype is realized at conception and remains fixed throughout her life. The genotype is described as a vector of individual base pairs, the fundamental structure of the DNA. DNA consists of two sets of 23 chromosomes each, one inherited from the mother and one from the father. Each set of chromosomes contains approximately 3 billion nucleotide base pairs located at specific addresses in the genome. The bases are adenine (A), thymine (T), guanine (G), and cytosine (C). The majority of the base pairs are invariant across the entire human population. A typical genome differs from the reference human genome at only 4–5 million of these addresses (1000 Genomes Project Consortium 2015).[5] Most of this variation consists of single base pair changes called single nucleotide polymorphisms or SNPs for short.[6] Genotypes can then be described as a vector of SNPs.[7]

[4] For a play of words, we call this the nature of nurture in the title of the paper. This is also sometimes called genetic nurture (see, e.g., Kong et al. 2018). More generally, this is an example of passive gene-environment correlation where an association comes about between the child's genes and the environment that the child is exposed to, independently of the genes that the child inherits (see Plomin, DeFries, and Loehlin 1977 for an early discussion of the term).

[5] The reference human genome is defined as "an accepted representation of the human genome sequence that is used by researchers as a standard for comparison to DNA sequences generated in their studies" (https://www.genome.gov/genetics-glossary/Human-Genome-Reference-Sequence).

[6] The remaining variation, not captured by SNPs, consists of rare single base pair variants (rare-variants), insertion or deletion of a sequence of base pairs (indels), and larger variations affecting multiple bases (structural variants). See 1000 Genomes Project Consortium (2015) for an overview of the variation in the human genome.

[7] The vector of SNPs is a proxy for genotypes, which is commonly used in the literature due to data availability. SNPs are correlated with (and hence pick up some of the) non-SNP genetic variation.

Formally, let $\mathbf{g}_i$ be the genotype of individual i; hence, $\mathbf{g}_i$ is a vector of nucleotide base pairs:

$$\mathbf{g}_i = \{g_{i1}, \ldots, g_{iS}\}, \tag{1}$$

where g_{is} is the base pair variant for individual i at position s, and S is the total number of SNPs.

While there are four different nucleotide base pairs, the vast majority of SNPs in the human genome are biallelic, meaning that only two types of base pairs are observed at that location. Therefore, we can summarize the variation in a specific SNP using three values $\{0, 1, 2\}$. These values correspond to the number of minor (less common) alleles present at the base pair. Formally,

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

For example, imagine that at base pair s, the common variant is guanine (G), and individual i inherited the guanine (G) variant from her mother and the less common cytosine (C) variant from her father.[8] Individual i has one minor allele (C) at position s, and we would say that the genotype of individual i at position s has a value of 1 ($g_{is} = 1$). Alternatively, if individual i had inherited the cytosine (C) variant from both parents, its genotype at position s would have a value of 2 ($g_{is} = 2$). Similarly, if individual i had inherited the common guanine (G) variant from both parents, its genotype would have the value 0 ($g_{is} = 0$).

The child's genotype is randomly determined from the parental genotypes, where in expectation, for each base pair s, we have that

$$E[g_{is}] = 0.5g_{is}^{f} + 0.5g_{is}^{m}, \tag{3}$$

where g_{is}^{f} (g_{is}^{m}) is the minor allele frequency for the child's father (mother) at position s. In expectation, the child's number of minor alleles will be an average of the number of minor alleles in the parental genotypes. For example, if both parents have zero minor alleles ($g_{is}^{f} = g_{is}^{m} = 0$), the process is deterministic since the child has no minor alleles to inherit and $g_{is} = 0$. Similarly, if both parents have two minor alleles ($g_{is}^{f} = g_{is}^{m} = 2$), then $g_{is} = 2$. The randomness of the process comes into play when one or both parents have exactly one minor variant. For example, in case the father has zero minor alleles and the mother has one ($g_{is}^{f} = 0$ and $g_{is}^{m} = 1$), the child will inherit one or zero maternal alleles with equal probability, so $g_{is} = 1$ or $g_{is} = 0$ and $E[g_{is}] = 0.5$. The likelihood that the child will inherit one allele or another, when both are present, is random by nature. This process creates a truly natural experiment that potentially allows for the identification of the causal effect of genes on a variety of outcomes.[9]

[8] Here, we disregard the fact that variants are base pairs and consider only one of the DNA strands and one of the bases. This is commonly done in the literature for simplicity since one base in the pair can be directly inferred from the other.

[9] The law of independent assortment implies that genetic inheritance occurs independently for each base pair g_{is}. In practice, however, some DNA sequences are more often inherited together. In particular, because the genetic reshuffling happens at random locations on the chromosome, genetic markers that are physically near to each other

Most socioeconomic outcomes (e.g., educational attainment, intelligence, personality, earnings, etc.)—as well as most other outcomes (e.g., height, BMI, several psychiatric disorders)—are highly polygenic, meaning that they are influenced by a large number of SNPs. Such polygenicity is analyzed in genome-wide association studies (GWAS). For example, Lee et al. (2018) show that at least 1,271 independent SNPs significantly influence educational attainment. For these outcomes, we are not interested in the effect of a particular variant but instead in the combined effect of all genotyped SNPs. The combined effect can be captured by the genetic factor

$$G_i = f_g(\mathbf{g}_i), \tag{4}$$

where G_i is the genetic factor for skill formation and f_g is an unknown function that combines the individual variants into the index. In Section IIIB, we show how we can identify the genetic factor using multiple independent measures called polygenic indexes.

This part of the model highlights some important concepts. First, genotypes are multidimensional, comprising millions of individual genetic variants that vary across the human population. Second, the child's genotype is not exogenous but determined by a random draw from the parental genotypes. The inheritance process induces a correlation between the child's and her parents' genotypes. Third, the randomness in the inheritance process allows for the identification of the effect of the child's genes that is independent of its parents. Finally, the association between the genome and an outcome of interest can be summarized in a genetic factor. We revisit these points throughout the paper.

B. *Technology of Skill Formation*

Skill Endowments: The child is born in period 0 with a set of initial skill endowments. We assume that skills may be described by a unidimensional measure (e.g., cognitive ability). We acknowledge that skills are multidimensional in nature and other (e.g., noncognitive) skills may also matter, but we prefer to restrict attention to a single dimension to focus instead on the different channels through which genes may influence the skill formation process. The model may be extended to include multiple skills as in Cunha, Heckman, and Schennach (2010). Let θ_{i0} be the skill endowment of child i at birth. We allow the initial skill endowment to be influenced by investments in utero U_i, including maternal health behaviors, such as smoking, drinking, and taking nutritional supplements.

We extend the traditional model and allow both the child's genetic factor (G_i) and parental genetic factors (G_i^m and G_i^f) to influence the child's development in utero: It is possible, for the same level of the mother's health behaviors, for some children to be more able to extract nutrients and other resources from their mother. Similarly, some mothers may biologically provide a better environment for fetal growth.

on the same chromosome are less likely to be separated—a tendency known as genetic linkage. Further, even without genetic linkage, some alleles may be more likely to appear together. Such "linkage disequilibrium" can be caused by various factors, for example, assortative mating and population stratification.

Formally, the child's initial skill endowment may be described by

$$\theta_{i0} = f_0^{\theta}\left(U_i, G_i, G_i^m, G_i^f\right), \tag{5}$$

where f_0^{θ} is a function describing how genes interact with the environment in utero in determining the child's initial skills.

We assume a log-linear specification for initial skills, so that

$$\ln\theta_{i0} = \alpha_1 G_i + \alpha_2 G_i^m + \alpha_3 G_i^f + \alpha_x \mathbf{X}_{i0} + \epsilon_{i0}, \tag{6}$$

where ε_{i0} is an i.i.d., mean zero, and normally distributed shock to early skills and $\mathbf{X}_{i0}$ is a vector of individual controls, both capturing differences in in utero investments (U_i). The α_1, α_2, and α_3 parameters capture the effect of the child's genetic factor on her initial stock of skills (i.e., effects during development in utero) and the association between her parents' genetic factors and the child's initial skills. These effects correspond to the direct effect on initial skills and the family genetic associations with initial skills.[10]

Skill Formation: Skills develop over time in response to external inputs. As in Cunha, Heckman, and Schennach (2010), the child's skills in period $t+1$, θ_{it+1}, are determined by its current skills, θ_{it}, and parental investments I_{it}. In addition, we allow the child's genetic factor, G_i, and the parents' genetic factors, G_i^m and G_i^f, to enter the production function of skills.

The model allows for some children to be better at learning on their own and improving their own skills. G_i captures this individual heterogeneity by influencing skill acquisition conditional on parental investments I_{it} and the current stock of skills θ_{it}. Similarly, for a given level of parent-child interaction, some parents may be better able to improve their children's skills than others. This heterogeneity is captured by the direct association with parental genes, G_i^m and G_i^f, in the technology of skill formation.

Formally, at each developmental stage t, let θ_{it} denote the child's skill stock. The technology of production of skills at stage t is

$$\theta_{it+1} = f_t^{\theta}\left(\theta_{it}, I_{it}, G_i, G_i^m, G_i^f, \mathbf{X}_{it}\right) \tag{7}$$

for $t = 1, 2, \ldots, T$. f_t^{θ} is a function that describes how genes interact with parental investments in determining the child's accumulation of skills.

We assume a translog technology specification in the form

$$\begin{aligned}\ln\theta_{it+1} = {} & \ln A_t + \delta_{1,t}\ln\theta_{it} + \delta_{2,t}\ln I_{it} + \delta_{3,t}\ln\theta_{it}\times\ln I_{it} \\ & + \delta_{4,t}G_i + \delta_{5,t}G_i^m + \delta_{6,t}G_i^f + \gamma_{x,t}\mathbf{X}_{it} + \epsilon_{it},\end{aligned} \tag{8}$$

[10] In practice, skills are not observable right at birth. We measure initial skills when the child is one year old. Hence, the genetic influences on initial skills include influences during the first year of life.

where $\ln A_t$ is the total factor productivity (TFP) parameter at period t and ϵ_{it} is a stochastic technology shock, which we assume is i.i.d. across individuals and time periods and is normally distributed with mean zero and variance σ_ϵ^2.

Genes influence skill accumulation via a variety of mechanisms. The child's genes can have a direct effect on skill accumulation, as captured by $\delta_{4,t}$. The direct effect captures genetic heterogeneity in children's ability to accumulate skills, irrespective of parental investments. These can change over time. The child's genes may also have indirect effects on skill accumulation via early skills and investments ($\delta_{1,t}$ and $\delta_{2,t}$). Lastly, family genetics may also matter for children's skill accumulation beyond the effects via investments ($\delta_{5,t}$ and $\delta_{6,t}$). These can capture a variety of familial influences, including the idea that some parents might be more successful at increasing their children's skills irrespective of the amount of interaction they have with their children. This can potentially capture genetic heterogeneity in the quality of parenting.

The model captures the idea that genetic influences may change in the course of the life cycle. This is motivated by Belsky et al. (2016), who demonstrate that genetic associations with academic ability increase from age 3 to age 13 years. Similarly, it is well documented that the heritability of IQ increases with age, a phenomenon known as the *Wilson Effect* (Bouchard 2013). In addition, the model captures the idea that returns to genetic factors may be different for individuals growing up in different environments as described in Papageorge and Thom (2019) and Ronda et al. (2022).

C. *Investment Policy Function*

In the model, parents invest in their children either due to altruism, paternalistic interest in having well-educated children, or some other motivation. Such investments may work either through direct interactions or through environmental changes (e.g., sending the child to swimming classes). Importantly, we allow the parental investment decision to depend on both the child's and the parents' genetic factors.

The investment policy function is modeled as follows:

$$I_{it} = f_t^I\left(\theta_{it}, G_i, G_i^m, G_i^f, \mathbf{X}_{it}\right) \tag{9}$$

for $t = 0, 1, 2, \ldots, T$. f_t^I is a function that describes how genes and the child's stock of skills influence parental investments.

The empirical specification for the investment policy function is

$$\ln I_{it} = \gamma_{1,t} \ln \theta_{it} + \gamma_{2,t} G_i + \gamma_{3,t} G_i^m + \gamma_{4,t} G_i^f + \gamma_{x,t} \mathbf{X}_{it} + \eta_{it}, \tag{10}$$

where η_{it} are i.i.d., mean zero, and normally distributed shocks. The investment policy function above is a reduced-form approximation of the parental behavior and follows previous work by Attanasio, Meghir, and Nix (2015); Agostinelli and Wiswall (2020); and Attanasio et al. (2020).[11]

[11] Parental investment choices depend on parental preferences for child quality, parental budget constraints, and parents' beliefs about both the child's current skills and the technology parameters. All of these components

The model allows us to decompose the association between the child's genes and parental investments into three distinct components. First, parents make investment decisions in response to the child's existing stock of skills ($\gamma_{1,t}$), which in turn is partially determined by the child's genetics. Second, different children might elicit different responses from their parents because of preferences and behavior not captured by current skills (e.g., enjoying being read stories), and these can also be partially determined by the child's genetics ($\gamma_{2,t}$). These two channels together capture the nurture of nature effect, as they describe how parental investments respond to the child's genetic makeup. Lastly, different parents face different constraints and have different preferences for investments, which might also be genetically determined ($\gamma_{3,t}$ and $\gamma_{4,t}$). This is another channel capturing the family genetic associations.

D. *Genetic Mechanisms*

The model highlights the idea that genes influence complex traits through a variety of mechanisms. We focus on three main mechanisms: the *direct effect*, the *nurture of nature effect*, and the *family genetic associations*.[12] We describe each in detail below.

The Direct Effect: First, we have the *direct* effect of children's genes on skill accumulation. The direct effect captures genetic heterogeneity in children's ability to retain new concepts, absorb information, and learn from their environment. It captures the idea that, for any given level of parental investments, some children may be better at taking advantage of their environment to improve their skills. The model captures this mechanism in two ways: α_1 and $\delta_{4,t}$, where the first parameter describes the effect of a child's genes on early skills and the second its effect on skill accumulation.

The Nurture of Nature Effect: A second way that genes may influence skill accumulation is via parental investments. We call this mechanism the *nurture of nature effect*. It captures how individuals, in general, interact with their own environments and how the parents respond to and invest in the child based on its genetic makeup. Such interactions may come about through two interactions, often called reactive and active genotype-environment correlation (Plomin, DeFries, and Loehlin 1977). First, in the reactive type, parents react to the child's existing stock of skills when deciding how to allocate resources within the family. The child's existing stock of skills may influence the price of investing in children, as in Becker and Tomes

could be influenced by parents' genetic factors. In principle, we could identify the separate genetic influences on investment choices using a structural model. However, a structural specification would either require detailed data on parental beliefs or assume that parents know the true production function, which goes against recent evidence (see, e.g., Cunha, Elo, and Culhane 2013; Boneva and Rauh 2018). Moreover, our current specification is consistent with multiple structural models of parental investments (see Attanasio, Meghir, and Nix 2015).

[12] In addition to the three mechanisms we focus on here, genes can also influence skill formation by what is often called gene-by-environment interactions. The conceptual model allows for these interactions. For example, gene-by-environment interactions could come about if genes changed the returns to parental investments in skill formation, captured by $\partial^2 f_t^{\theta}\left(\mathbf{g}_i, \bar{\theta}_{it}, I_{it}, \bar{\mathbf{g}}_i^m, \bar{\mathbf{g}}_i^f\right)/\partial \mathbf{g}_i \partial I_{it}$ in the model. In our preferred empirical specification, we do not consider such interactions. In online Appendix D, we consider this possibility and find little support in the data for these interactions being important during early skill formation.

(1976). In addition, in multiple-child families, parents might respond to one of the children's stock of skills due to aversion to inequality in children's outcomes (Behrman, Pollak, and Taubman 1982). Second, in the active type of interaction, different children might elicit different responses from their parents due to preferences and behavior not captured by current skills (e.g., enjoying being read stories), which are partially determined by the child's genetics. We cannot separately identify the two types of interaction and instead, refer to the overall mechanism as the *nurture of nature effect*. In the model, this effect is then captured in two ways: $\gamma_{2,t}$ and $\gamma_{1,t}$, where the first parameter describes how parents' investment decisions depend on the child's genetic factor and the second how it depends on the child's stock of skills.

Family Genetic Associations: Lastly, parental genes may be associated with the environment in which the child develops her skills. For the play of words, we also refer to it as *nature of nurture* in the title. Empirically, however, parental genes can capture a variety of environmental effects. For example, grandparents' genes may also affect child development directly (insofar as grandparents help raise the child) or indirectly through the parents. Since we don't observe grandparents' genes, parental genes will capture these effects. Thus, family genetic associations reflect the combined genetic effects of previous generations captured by parental genes. In addition, this channel may also reflect other social mechanisms, including assortative mating and population stratification effects that are correlated with the family genetic makeup. Since we cannot causally identify the specific mechanism behind these interactions, we loosely refer to this mechanism broadly as *family genetic associations*.

There are a variety of ways that parental genes enter the model. First, parental genes may be related to skill formation directly via the quality of parental interactions in utero (α_2 and α_3). In addition, parental genes may be associated with the quantity of investments in the child's skill formation ($\gamma_{3,t}$ and $\gamma_{4,t}$). Lastly, parental genes may be related to skill formation directly via the quality of parent-child interactions during the child's development ($\delta_{5,t}$ and $\delta_{6,t}$).

A better understanding of the different ways genes affect skill formation, either directly, via parental investments (nurture of nature), or via family genetic influences (nature of nurture), can help us better understand heterogeneity in environmental effects and thereby enable us to design policies directly aimed at decreasing the effects of various types of disadvantage. For instance, the existence of a positive nurture of nature effect would imply that parents invest more in genetically advantaged children. This would mean that environmental and genetic differences interact to enlarge existing inequalities in skills. It would also mean that the existing genetic effects may increase the value of policies aimed at reducing inequality in parental investments. We come back to this point in Section V, where we discuss the relative importance of the different mechanisms and the implications of our findings.

II. Data

In this section, we introduce the ALSPAC dataset and the key variables used in our analysis. We also present some reduced-form results on the relationship between child genes, parental genes, child skills, and family investments.

A. *ALSPAC*

To investigate the relationship between genetics and the development of child skills and family investments during childhood, we need a comprehensive dataset. For our purpose, the Avon Longitudinal Study of Parents and Children (Boyd et al. 2012; Fraser et al. 2012) provides a compelling resource. ALSPAC is a British birth cohort study initially composed of 14,541 women recruited during pregnancy between April 1991 and December 1992, resulting in 14,062 live births. Data were collected by epidemiologic researchers from the University of Bristol to aid the study of the environmental and genetic factors affecting human health and development (Boyd et al. 2012; Fraser et al. 2012).[13]

Questionnaires were sent to the primary caregiver (usually the mother) at regular intervals, starting before the child's birth.[14] The caregiver responds to questions about the child's development and behavior as well as parenting, activities, and the home environment. We focus on the first seven years of the child's life, as this allows us to follow the child's development through a set of similar measures.

One attractive feature of the ALSPAC is the large set of child developmental data. Another crucial feature for our study is the availability of genetic information. DNA samples have been collected and genotyped for many of the mothers and children as well as some of the fathers. The maternal and child biological samples consist of blood samples repeatedly collected as part of routine prenatal care and follow-up at clinics. Paternal samples started being collected only recently, and thus, only a few of the fathers have been genotyped.[15]

Our main sample includes families where the child and either of its parents were genotyped. We excluded individuals of non-European ancestry and those with missing information on many skill and investment measures. We describe the sample selection procedure in more detail in online Appendix A. The resulting sample includes 4,510 children from the original sample of 14,062 children.

Measures of Skills and Investments.—From the wide range of questions put to the mother, we selected the subset of questions most closely related to child skill development and family investments in the child. Measurements of child skills were obtained from the "milestones" and "abilities and disabilities" sections of the primary caregiver questionnaires. In these sections, the primary caregiver was given a list of things children gradually learn to do as they get older and asked to indicate whether the child (i) "Can do it well" or "Does it often," (ii) "Can do it but not very well" or "Has done it once or twice," or (iii) "Has not yet done it." We selected a subset of the measures that relate to children's ability to process new information and perform various tasks and their capacity to learn abstract concepts such as language. The selected measures are displayed in Table 1.

[13] The study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/. Ethical approval for this study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

[14] Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

[15] Consent for biological samples was collected in accordance with the Human Tissue Act 2004.

TABLE 1—MEASURES OF CHILD SKILLS

Measure	Period / Age	0 0–2	1 2–3	2 3–4	3 4–5	4 5–6	5 6–7
1	Can build tower of 8 bricks	X	X	X			
2	Plays cards (or board games)		X	X	X	X	X
3	Plays peek-a-boo	X					
4	Can focus eyes on small object	X					
5	Can build tower of 4 bricks	X					
6	Freq. names things	X					
7	Combines two different words		X				
8	Can copy vertical line with pencil		X				
9	Can copy and draw a circle		X	X			
10	Uses plurals		X	X			
11	Uses possessives		X	X			
12	Adds -ing to words		X	X			
13	Adds -ed to words		X	X			
14	Can copy and draw a plus sign/cross			X			
15	Can copy and draw a square			X	X		
16	Can write their name				X		
17	Can write any numbers				X		
18	Knows at least 10 letters				X		
19	Can read simple words				X		
20	Can read a story with <10 words per page				X		
21	Can count up to 20				X		
22	Can read a story with >10 words per page				X	X	X
23	Can count up to 100				X	X	X
24	Can play any board games				X	X	X

Notes: This table reports the individual measures of child skills. An X indicates that the measure is available in that period and is used in the estimation.

Similarly, we obtained the measurements of family investments from the "you and your child" sections from the primary caregiver questionnaires. We selected a subset of measures capturing aspects of the family environment relating to behavior and activities involving the child and her parents. For such activities, we strove to achieve a balance between the parents by selecting several measurements relating specifically to both the mother and the father (in addition to a number of neutral measurements). For these measurements, the primary caregiver indicates whether the parent does certain activities with the child (e.g., "Frequency child goes to the library") and at which frequency: (i) "Nearly every day," (ii) "2–5 times per week," (iii) "Once per week," (iv) "Once per month," (v) "A few times per year," or (vi) "Never." The selected measures are displayed in Table 2.

B. *Measures of the Genetic Factor*

We rely on recent advances in molecular genetics to construct different measures of the genetic factor. We measure the genetic factor using multiple polygenic indexes (PGI). Polygenic indexes are a linear combination of the SNP count variables weighted by the strength of association with the outcome of interest, educational attainment and cognitive performance in our case. The weights are derived from large, external genome-wide association studies.

Table 2—Measures of Investments

Measure	Period / Age	0 0–2	1 2–3	2 3–4	3 4–5	4 5–6
1	Freq. goes to places of interest	X	X	X	X	X
2	Freq. goes to library	X	X	X	X	X
3	Freq. mum reads to child	X	X	X	X	X
4	Freq. partner sings to child	X	X	X	X	X
5	Freq. child taken to park	X	X	X		
6	Freq. mum shows child picture books	X		X		
7	Freq. partner shows child picture books	X		X		
8	Freq. partner plays with toys with child	X		X		
9	Freq. partner reads to child	X		X	X	X
10	Freq. goes to swimming pool or sports area				X	X
11	Freq. goes to special classes or clubs				X	X

Notes: This table reports the individual measures of child investments. An X indicates that the measure is available in that period and is used in the estimation.

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

$$pgi_i^w = \sum_{s=1}^{S} \beta_s^w g_{is}, \tag{11}$$

where g_{is} is individual i's SNP count at location s and β_s^w is the GWAS weight for variant s and outcome w. We provide more information on the construction of these indexes and the literature on polygenic indexes in online Appendix B.

We construct three polygenic indexes based on different samples and outcomes separately for children, their mothers, and their fathers. In Section IIIB, we show how three different indexes enable us to correct for measurement error in the underlying genetic factor. We construct two polygenic indexes of educational attainment (EA). The first is based on the GWAS conducted using 23andMe participants, and the second is based on the GWAS sample in Lee et al. (2018) excluding 23andMe participants. For our purpose, the difference between the EA PGI (23andMe) and the EA PGI (w.o. 23andMe) is not important, except that they are based on two different nonoverlapping samples. We supplement these with the GWAS for cognitive performance also based on the GWAS sample in Lee et al. (2018) excluding 23andMe participants. [16] We impute missing parental genotypes when possible before constructing the indexes. This is done using the method outlined in Young

[16] We use the publicly available summary statistics at the SSGAC website, which includes the summary statistics of all meta-analyses of all discovery cohorts except 23andMe, as well as private summary statistics provided to us by 23andMe directly.

et al. (2020), which infers a missing parental genotype from a parent-offspring pair with observed genotypes.[17,18]

Polygenic indexes have several appealing features but some important limitations. It is appealing that we can measure the genetic factor at the individual level. It is also appealing that we can construct multiple and independent measures of the genetic factor. Another appealing feature is that we can directly compare children's genetic factors to that of their parents. We can exploit the natural experiment created by the inheritance process to estimate the effect of the child's PGI that is independent of her parents' PGI.

However, the PGI has two important limitations. First, it measures the genetic potential for educational attainment and cognition, which means that the index does not capture genetic variation unrelated to these two traits. This may be a problem if genes unrelated to cognitive performance and educational attainment also influence children's skill formation. For example, altruistic parents may invest more in their children, and at the same time, the genetic propensity for altruism might not be well captured by these indexes. That said, given that the polygenic index is constructed as the best linear predictor of cognitive performance and educational attainment in cross-sectional data (i.e., not controlling for parental genotypes), it should partially capture any genetically influenced characteristics of parents that promote skills in their children. Second, polygenic indexes are estimated with error. We can accurately measure the SNPs, but the weights (β_j^ws in equation (11)) are always measured with some error. This is due to the estimation process. An increase in sample sizes and improved quality control in the original GWAS can improve the estimation error. Our approach corrects for this error under some reasonable assumptions. We detail our method and assumptions in Section III.

C. *Summary Statistics*

We present summary statistics for the main variables used in our study in online Appendix A. As explained above, we restrict our sample to families for whom genetic information is available for full trios—child, mother, and father—after imputing missing genotypes. Imputing missing genotypes has two advantages. First, it increases statistical power by substantially increasing the estimation sample size. Second, it helps with issues of sample selection. The missingness in parental genotype is not random since both parents had to be present in the household and willing to participate in the genetic sample collection for their genotype to be observed. This restriction induces a strong positive selection on the genetic factor, especially for the father (who is usually the missing parent). We show this in Table A2 in online Appendix A. Imputing the missing parental genotype makes our main sample much

[17] The method exploits known patterns of genetic inheritance to infer the genotype of the missing parent. For each SNP, one allele on the missing parental genotype can be inferred exactly unless both the child and the observed parent are heterozygous (have exactly one minor allele). For example, suppose for some SNP, the mother has two minor alleles (say, CC), and the child has only one minor allele (say, GC). The major allele (G) must have been inherited from the father. The other paternal allele is not known exactly but is known in expectation. It can be imputed using the information on the allele frequencies at different SNPs from individuals from a similar ancestry group.

[18] In online Appendix C, we show that linear estimators using imputed paternal polygenic scores are consistent both in the simple OLS case and with an IV estimator, as the one used in this paper.

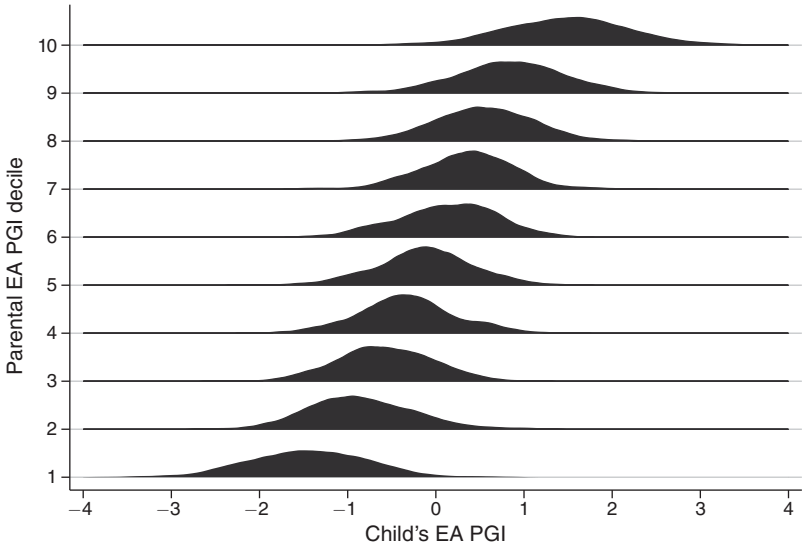

Figure 1. Distribution of Children's EA PGI by Parental EA PGI Decile

Notes: This figure plots the density of the standardized EA PGI of the child, separately for each decile of the parental EA PGI. The figure highlights the dependence between the child's and its parents' genetic potential for education. It also highlights the variation in the child's potential for education even after conditioning on the parents' potential. There is an overlap in the distribution of genetic potential for education across all parental PGI deciles.

more similar to the full sample (before applying the other sample selection criteria). However, it remains significantly different with regard to the polygenic indexes and birth order. It is not surprising that some positive selection remains because another restriction comes from whether parents fill out sufficiently many of the items in the questionnaire. But importantly, we reduce the extent of this selection through the imputation method.

The child's genotype is randomly determined from the SNPs of the mother and the father. The inheritance process can be thought of as a large series of coin tosses. For that reason, the possible values the child's genetic factor can take vary considerably. The child's genetic factor can be significantly different from the parental indexes. Since there is substantial variation in how strongly each SNP is associated with skill formation, some children will be lucky and inherit the important genes from both the mother and the father, even to the point that the child's genetic factor may be higher than that of both its mother and father. Naturally, the reverse may also be true.

This point is illustrated in Figures 1 and 2. Figure 1 plots the distributions of the child educational attainment polygenic index and how it relates to parental (the average of maternal and paternal) PGI.[19] For each decile of parental PGI, we plot the density of polygenic indexes of the children of those parents. The figure illustrates the randomness of genetic inheritance. Although parents with a high PGI

[19] For the exercises in this section, we rely on the polygenic index for educational attainment constructed from the results in Lee et al. (2018) excluding 23andMe participants. Results are similar when relying on the other two polygenic indexes.

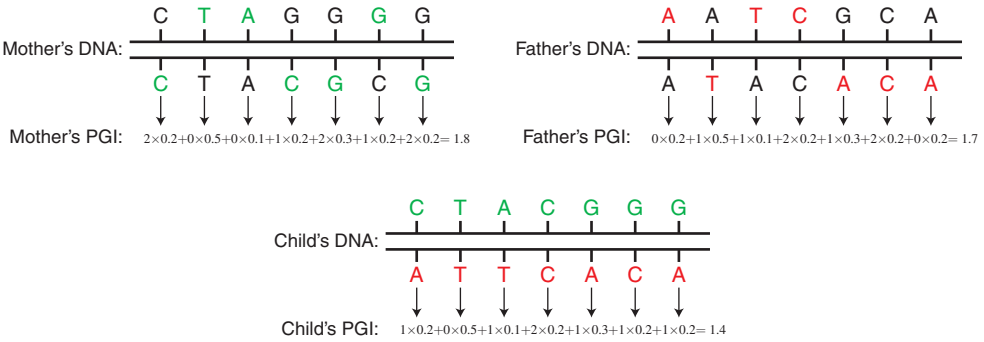

FIGURE 2. FAMILY GENETIC DATA

Notes: Using a stylized example, this figure depicts how maternal and paternal DNA is transmitted to the child and how the child's polygenic index might differ from that of its parents. The child's DNA is composed of variants inherited at random from the mother (in green) and from the father (in red). The randomness in the inherited process allows for significant variation in the variants that are inherited by the child and in the child's polygenic index.

naturally tend to have children with a high PGI, the different decile groups overlap extensively, even between children of parents in the top and bottom deciles of the PGI distribution. Figure 2 offers a stylized example of how that can happen. We demonstrate graphically how the child's DNA is inherited from the mother (the green variants) and from the father (the red variants). Using only seven SNPs, we also show how the child's genetic factor can be below that of both parents.

D. *Preliminary Evidence*

While the above shows the substantial variation in children's indexes even after accounting for their parents' genetic variation, strong dependence remains between the two, and controlling for this dependence is important when trying to understand the effect of genes on skills and investments. To make this point, we present preliminary reduced-form results where we test for the association between the child's PGI, without correcting for measurement error and using a crude measure of skills and investments before and after we control for the parental PGI. We construct the crude measures of skills and investments by averaging the standardized set of measures available at each age.

This exercise is also helpful as it clarifies the variation in the data that identifies the genetic mechanisms estimated using the structural model. A correlation between the child's PGI and the child's skills would suggest that genes matter for skill development. This could reflect any of the three mechanisms we have described in our framework (the direct effect of genes, the nurture of nature effect, and family genetic associations). If an association then remains between the child's PGI and skills after controlling for parental genes, this would reflect an actual effect of the child's genes (which are random conditional on parental genes), either through the direct channel or the nurture of nature channel. Similarly, if the child's PGI is associated with parental investments, this could be indicative of either the nurture of nature effect or of family genetic associations (given that investments also affect skill accumulation). If we regress parental investments on both the child's and the

Table 3—EA PGI and Skills by Age

Ages	[0–2]	[2–3]	[3–4]	[4–5]	[5–6]	[6–7]	[Pooled]
Panel A							
Child's PGS	0.029	0.041	0.067	0.121	0.132	0.091	0.080
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.010)
R^2	0.001	0.002	0.004	0.015	0.016	0.008	0.006
Observations	4,510	4,510	4,510	4,510	4,510	4,510	27,060
Panel B							
Child's PGS	0.015	0.007	−0.013	0.071	0.076	0.050	0.034
	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.018)
Mother's PGS	0.032	0.037	0.069	0.055	0.057	0.041	0.049
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.013)
Father's PGS	−0.006	0.019	0.061	0.028	0.034	0.027	0.027
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.015)
R^2	0.006	0.066	0.047	0.039	0.035	0.019	0.032
Observations	4,510	4,510	4,510	4,510	4,510	4,510	27,060

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

parents' PGI, a larger association between the child's PGI and investments would suggest that the nurture of nature channel is more important, while a larger association between parental PGI and investments would suggest that the family genetic associations are more important.

The results of these reduced-form analyses are shown in Tables 3 and 4. Panel A of Table 3 demonstrates that the child's EA PGI is positively associated with her skills across childhood, starting at ages zero to two years and lasting until ages six to seven years. Moreover, the association tends to increase over time and triples in size from ages zero to two years to ages six to seven years. However, because each time period uses a different combination of skill measures and because a repeated measure may not be equally precise in different periods, this association could be an artifact of the limitations of this reduced-form approach. Furthermore, these preliminary associations capture a combination of all three mechanisms described in Section ID. Our main empirical model, described in the next section, solves both of these concerns.

In panel A of Table 4, we relate the child's EA PGI and parental investments across the child's development. We document a strong and roughly age-invariant association between the child's genetics and investments across the entire age span from zero to seven years. These results provide preliminary evidence of the importance of genes for parental investment, one of the key findings in this paper. However, this association captures both nurture of nature effects and family genetic associations.

In panel B of both tables, we control for maternal and paternal PGI. Controlling for the parents' PGI reduces the association between the child's EA PGI and skills substantially. We document an even larger reduction when looking at the association between the child's EA PGI and parental investments. These results are a

Table 4—EA PGI and Investments by Age

Ages	[0–2]	[2–3]	[3–4]	[4–5]	[5–6]	[6–7]	[Pooled]
Panel A							
Child's PGS	0.120	0.156	0.128	0.146	0.140	0.161	0.142
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)
R^2	0.015	0.025	0.016	0.022	0.020	0.025	0.020
Observations	4,510	4,510	4,510	4,510	4,510	4,510	27,060
Panel B							
Child's PGS	0.063	0.043	0.051	0.008	0.003	0.024	0.032
	(0.028)	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.021)
Mother's PGS	0.056	0.130	0.069	0.132	0.104	0.133	0.104
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.016)
Father's PGS	0.038	0.059	0.058	0.094	0.116	0.093	0.076
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.017)
R^2	0.021	0.039	0.019	0.043	0.039	0.041	0.034
Observations	4,510	4,510	4,510	4,510	4,510	4,510	27,060

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

strong indication of the presence and importance of family genetic influences on children's development and emphasize the need to control for parental genes in our main empirical analysis. While informative, these initial findings have several unattractive features that are improved in the empirical model we present in the next section. These estimates are significantly attenuated due to measurement error, which can be different across time periods. For this reason, the crude measures of skills and investments do not allow us to compare genetic influences over time or separately identify the three mechanisms described in Section ID. The empirical approach described in the next section addresses these concerns.

III. Empirical Model: Identification and Estimation

In this section, we discuss how we take the empirical model to the data. We also explain the empirical challenges we face and the estimation procedure. We described our model in detail in Section I. The model involves estimating three dynamic equations. The child's initial skills (equation (6)), the technology of skill formation (equation (8)), and the investment policy function (equation (10)). The key challenge is that there are many latent and unobserved factors in the model. We do not observe skills and investments directly but proxies of these latent factors that change over time. We also do not observe the genetic factor of children and their parents and instead only observe the constructed polygenic indexes. We rely on multiple measures of skills, investments, and genetic factors in a dynamic measurement system to identify all these unobserved latent factors. We follow the estimation approach detailed in Agostinelli and Wiswall (2020) to recover all structural parameters.

A. *Identification of Latent Skills and Investments*

We observe multiple measures of children's skills and parental investments in each period. These measures are imperfect proxies of underlying skills and investments. In this section, we describe how the multiple measures may be used to identify these underlying latent variables and how we can track the evolution of skills and investments dynamically. We formalize the measurement error system in a factor analytic approach as in Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010).

Formally, in each period t, we observe J measurements of the child's skills and K measurements of parental investments. Let m^{θ}_{ijt} denote the jth measurement of child i's skill at period t, and let m^{I}_{ikt} denote the kth measurement of child i's parental investment at period t. Following Attanasio, Meghir, and Nix (2015) and Agostinelli and Wiswall (2020), we assume a linear-log relationship between each measurement, the latent child skills θ_{it}, and latent parental investments I_{it}:

$$m^{\theta}_{ijt} = \mu^{\theta}_{jt} + \lambda^{\theta}_{jt} \cdot \ln \theta_{it} + \nu^{\theta}_{ijt}, \tag{12}$$

$$m^{I}_{ikt} = \mu^{I}_{kt} + \lambda^{I}_{kt} \cdot \ln I_{it} + \nu^{I}_{ikt}, \tag{13}$$

where λ^{θ}_{jt} and λ^{I}_{kt} are the factor loading for skill measurement j and investment measurement k and ν^{θ}_{ijt} and ν^{I}_{ikt} are i.i.d. measurement errors.[20] As in Agostinelli and Wiswall (2020), we make no further assumptions on the distribution of the measurement errors.

Identifying Assumptions.—We array the measurement errors for skills ν^{θ}_{jt} in a vector $\boldsymbol{\nu}^{\theta}$ and the measurement errors for investments ν^{I}_{kt} in a vector $\boldsymbol{\nu}^{I}$ and assume that

$$\nu^{\theta}_{jt} \perp\!\!\!\perp \nu^{\theta}_{j't}, \ \forall j \neq j', \tag{14}$$

$$\nu^{\theta}_{jt} \perp\!\!\!\perp \nu^{\theta}_{jt'}, \ \forall t \neq t', \tag{15}$$

$$\nu^{I}_{kt} \perp\!\!\!\perp \nu^{I}_{k't}, \ \forall k \neq k', \tag{16}$$

$$\nu^{I}_{kt} \perp\!\!\!\perp \nu^{I}_{kt'}, \ \forall t \neq t', \tag{17}$$

$$\boldsymbol{\nu}^{\theta} \perp\!\!\!\perp \boldsymbol{\nu}^{I}. \tag{18}$$

The assumptions in equations (14) to (18) maintain that the measurement errors are independent of each other and independent across time. These assumptions imply that conditional on latent skills and investments, the residual information

[20] This assumption says that conditional on latent skills and investments, the measurement error in the skill and investment measures is independent across measures and developmental periods. It also means that conditional on the latent skills and investments, the measurement errors are independent of child and parental genes. That means we assume that genes influence the latent skills and investments and not the measures themselves, which is consistent with our model.

in the measurements is uninformative about the process of skill formation. While this is a standard assumption in the literature, it can be strong in certain contexts. For example, it can fail if one of the measurements captures information on additional skills not captured in θ but important for the development process. Similarly, it would fail if there are additional dimensions of parental investments not captured by I but relevant for some of the measurements.

Identification of the measurement system and the latent skills and investments requires further restrictions. Without further normalization, neither location nor scale of the latent skills and investments can be identified. Agostinelli and Wiswall (2020) discuss the implications of different normalizing assumptions. In particular, they show that the production function can be estimated without further restrictions only if a particular measurement (or a combination of several overlapping measurements) is available at all periods for both skills and investments, the measurement thus being age-invariant. That way, the location and scale of the latent skills or investments at any period can be identified relative to the age-invariant measurement.

We are fortunate to have several such measures for investments in our data. Table 2 lists all the measures we use to identify the latent investments at the different periods. We have four measures that are asked at all periods: "Frequency the child goes to places of interest," "Frequency the child goes to a library," "Frequency the mother reads to the child," and the "Frequency the mum's partner sings to the child." In our benchmark specification, we chose the "Frequency the child goes to places of interest" as our age-invariant measure, denoted by $k = 1$. Results are similar when we use one of the other three measures. We make the following normalizing assumption on the age-invariant measure of investments:

$$m^I_{i1t} = 0 + 1 \cdot \ln I_{it} + \nu^I_{i1t} \quad \text{for} \quad t \in \{0,1,2,3,4\}. \tag{19}$$

Unfortunately, we do not have a measure that is asked at all periods for the latent skill. Our measures of skills capture different child development achievements, such as being able to use plurals or read simple words. These achievements are age specific since most children are able to complete some of the tasks after a certain age, and few young children can complete other tasks. For this reason, no question is put to the child in all six periods. Identification is then obtained from two separate measures that are asked at many but not all periods (see Table 1). The survey asks whether the child "Can build a tower of 8 bricks" in periods 0, 1, and 2. Similarly, the survey asks the mother if the child "Can play card games (or board games)" in periods 1, 2, 3, 4, and 5. Since the two measures overlap at some periods and cover all periods together, we use them to identify the location and scale of the latent skills across periods. Other combinations are possible and do not alter our main findings. Formally, letting the measure "Can build a tower of 8 bricks" be described by $j = 1$ and "Can play card games (or board games)" by $j = 2$, we make the following normalizing assumption on the two measures:

$$m^\theta_{i1t} = 0 + 1 \cdot \ln \theta_{it} + \nu^\theta_{i1t} \quad \text{for} \quad t \in \{0,1,2\}, \tag{20}$$

$$m^\theta_{i2t} = \mu_{21} + \lambda_{21} \cdot \ln \theta_{it} + \nu^\theta_{i2t} \quad \text{for} \quad t \in \{1,2,3,4,5\}, \tag{21}$$

where μ_{21} and λ_{21} are identified in period 1 using the normalization on the first measure. We describe how these are estimated in the next section.

Estimation of the Measurement System.—Given the assumptions in equations (14) to (18) and the normalization restrictions, we can estimate all parameters of the measurement system for latent skills and investments and the means and distribution of the latent variables. The parameters of the measurement system include the factor loadings (λ_{jt}^{θ} and λ_{kt}^{I}), the measurement means (μ_{jt}^{θ} and μ_{kt}^{I}), and the variance of the measurement errors ($\sigma_{jt,\theta}^{2}$ and $\sigma_{kt,I}^{2}$). These parameters can be estimated directly from ratios of the covariance between different measurements, from the measurement means, and from the measurement variance.

Consider three measurements of latent investments in period 1 (m_{11}^{I}, m_{21}^{I}, and m_{31}^{I}). Recall that we assume $\lambda_{11}^{I} = 1$ and that the measurement errors are independent (equation (16)), so we can write the covariance between each pair of measurements as

$$\begin{aligned}
\operatorname{cov}\left(m_{11}^{I}, m_{21}^{I}\right) &= 1 \cdot \lambda_{21}^{I} \cdot \operatorname{var}\left(\ln I_1\right),\\
\operatorname{cov}\left(m_{11}^{I}, m_{31}^{I}\right) &= 1 \cdot \lambda_{31}^{I} \cdot \operatorname{var}\left(\ln I_1\right),\\
\operatorname{cov}\left(m_{21}^{I}, m_{31}^{I}\right) &= \lambda_{21}^{I} \cdot \lambda_{31}^{I} \cdot \operatorname{var}\left(\ln I_1\right).
\end{aligned}$$

As first shown in Carneiro, Hansen, and Heckman (2003), we can use these three identities to identify the three unknowns (λ_{21}^{I}, λ_{31}^{I}, and $\operatorname{var}\left(\ln I_1\right)$). To see this, note that

$$\begin{aligned}
\operatorname{var}\left(\ln I_1\right) &= \frac{\operatorname{cov}\left(m_{11}^{I}, m_{21}^{I}\right) \cdot \operatorname{cov}\left(m_{11}^{I}, m_{31}^{I}\right)}{\operatorname{cov}\left(m_{21}^{I}, m_{31}^{I}\right)},\\
\lambda_{21}^{I} &= \frac{\operatorname{cov}\left(m_{21}^{I}, m_{31}^{I}\right)}{\operatorname{cov}\left(m_{11}^{I}, m_{31}^{I}\right)},\\
\lambda_{31}^{I} &= \frac{\operatorname{cov}\left(m_{21}^{I}, m_{31}^{I}\right)}{\operatorname{cov}\left(m_{11}^{I}, m_{21}^{I}\right)}.
\end{aligned}$$

We can extend this procedure to include additional measurements beyond the first three. When the model is overidentified, we take the means of different combinations of measurements as our estimates. The procedure can be applied to all periods to identify all factor loadings (λ_{kt}^{I}). The factor loadings for the latent skills can be identified in a similar manner, with the additional step that we must first estimate λ_{21} before estimating the factor loadings in the later periods.

Once the variance of the latent variable ($\operatorname{var}\left(\ln I_t\right)$ and $\operatorname{var}\left(\ln \theta_t\right)$) and the factor loadings are identified, we can also identify the mean of the latent variables ($E\left[\ln I_t\right]$ and $E\left[\ln \theta_t\right]$) and then the measurement means (μ_{kt}^{I} and μ_{jt}^{θ}). To see this, note that since we assume $\mu_{1t}^{I} = 0$, we have that

$$E\left[\ln I_t\right] = E\left[m_{1t}^{I}\right].$$

Similarly, we have that

$$\mu_{kt}^{I} = E\left[m_{kt}^{I}\right] - \lambda_{kt}^{I} \cdot E[\ln I_t].$$

The estimation procedure for the latent skill is similar but with the additional step that we need to set $\mu_{2t}^{\theta} = \mu_{21}^{\theta}$, which can be identified in period 1 from the assumption that $\mu_{11}^{\theta} = 0$.

Lastly, once all other parameters are identified, we can identify the variance of the measurement errors ($\sigma_{jt,\theta}^2$ and $\sigma_{kt,I}^2$) from each measurement variance. These follow from the following identity:

$$\sigma_{kt,I}^2 = \operatorname{var}\left(m_{kt}^{I}\right) - \left(\lambda_{kt}^{I}\right)^2 \cdot \operatorname{var}(\ln I_t),$$

$$\sigma_{jt,\theta}^2 = \operatorname{var}\left(m_{jt}^{\theta}\right) - \left(\lambda_{jt}^{\theta}\right)^2 \cdot \operatorname{var}(\ln \theta_t).$$

Signal to Variance Ratio.—The information relative to measurement error contained in each of the measures is described by the signal to variance ratio. It measures the fraction of the variance in a given measure that is explained by the latent factor. Formally, for the *j*th measure of child's skill in period *t*, the ratio is defined as

$$s_{jt}^{\theta} = \frac{\left(\lambda_{jt}^{\theta}\right)^2 \times \operatorname{var}(\ln \theta_t)}{\left(\lambda_{jt}^{\theta}\right)^2 \times \operatorname{var}(\ln \theta_t) + \operatorname{var}\left(\nu_{jt}^{\theta}\right)}. \tag{22}$$

We report the signal to variance ratio for each skill measure at each period in Table A5 and for each investment measure at each period in Table A6, both in online Appendix A. The degree of measurement error varies substantially, both across measures and over time. In particular, we see that the first period is characterized by low signal to variance ratios for all items. This shows how measures of skills at early ages are poor proxies of children's underlying skills. It also highlights the importance of accounting for measurement error in a formal skill formation model. Without this approach, the estimates of the parameters in the structural model would be biased.

B. *Identifying the Genetic Factor*

A similar logic can be applied to identify the genetic factor of children and their parents. To do so, we need to first assume that the latent genetic factor is a linear combination of individual *i*'s SNPs. At first, this seems like a very strong assumption. However, there is much empirical and theoretical evidence that most genetic variance for polygenic phenotypes can be explained by the additive (linear) component (see, for example, the discussion in Hill, Goddard, and Visscher 2008). Formally, let g_{ij} correspond to the *j*th SNP of individual *i*, measuring the number of minor alleles at the *j*th location of individual *i*'s genome, and *J* be the number of relevant SNPs in the genome. We can then write the underlying genetic factor of individual *i* as

$$G_i = \sum_{j=1}^{J} \beta_j g_{ij}, \tag{23}$$

where β_j corresponds to the direct effect of SNP j.

The key identification issue is that G_i and the corresponding weights β_j are unobserved. However, we do observe measures of G_i corresponding to the polygenic indexes. Under classical measurement error assumptions, we can use the latent factor model to identify G_i as we did with skills and investments. Formally, let pgi_i^k be a polygenic index for a trait related to G_i and $\hat{\beta}_j^k$ be the estimated weight from a GWAS. We assume that

$$pgi_i^k = \sum_{j=1}^{J} \hat{\beta}_j^k g_{ij} \tag{24}$$

$$= \lambda_k^G \sum_{j=1}^{J} \beta_j g_{ij} + \sum_{j=1}^{J} \eta_j^k g_{ij} \tag{25}$$

$$= \lambda_k^G G_i + \zeta_i^k, \tag{26}$$

where under the linearity assumption, we can decompose the PGI into the genetic factor and a measurement error component $\zeta_i^k = \sum_{j=1}^{J} \eta_j^k g_{ij}$, which aggregates the estimation error for the coefficient of association across SNPs.

Identification of the genetic factor relies on the assumption that the measurement error is independent across PGIs. That is, it relies on the assumption that

$$\zeta_i^k \perp\!\!\!\perp \zeta_i^{k'}, \ \forall k \neq k'. \tag{27}$$

This assumption is plausible for PGIs constructed from independent GWAS with no sample overlap if we assume that η_j^k captures mainly estimation error in the GWAS. However, this assumption would not hold if the β_js are systematically misestimated, for example, due to not correctly controlling for population stratification in the GWAS. In this case, the bias for each SNP would be similar across GWAS, and assumption (27) would fail. As a general rule, any bias in the original GWAS will carry over to downstream analyses and cannot be addressed ex post using standard measurement error methods. This highlights the importance of a proper GWAS design for any downstream analyses. Since there are only two large independent GWAS of educational attainment, we rely on a PGI constructed from the cognition GWAS as our third measure of the genetic factor. Suppose we again assume that η_j^k captures only the estimation error in the GWAS. In that case, it is also plausible that assumption (27) will hold for PGIs constructed using two distinct, but related, outcomes since there will be overlap in the genetic signal from the two outcomes, but the estimation error will be independent across the two estimates.

The availability of three polygenic indexes that satisfy assumption (27) is sufficient to identify G_i. To see that, without loss of generality, assume that $\text{var}(G_i) = 1$ and $E[G_i] = 0$. Then the λ_k^G can be identified by the covariances between the three indexes:

$$\text{cov}(pgi_i^1, pgi_i^2) = \lambda_1^G \lambda_2^G \text{var}(G_1) + \text{cov}(\zeta_i^1, \zeta_i^2) \tag{28}$$

$$= \lambda_1^G \lambda_2^G. \tag{29}$$

So

$$\lambda_1^G = \frac{\text{cov}\left(pgi_i^1, pgi_i^2\right) \times \text{cov}\left(pgi_i^1, pgi_i^3\right)}{\text{cov}\left(pgi_i^2, pgi_i^3\right)}. \tag{30}$$

Signal to Variance Ratio.—We report the signal to variance ratio for the genetic factor of children and their parents in Table A7 in online Appendix A. The EA PGI computed using the summary statistics from Lee et al. (2018) without the 23andMe data shows the strongest signal, followed by the EA PGI computed using the 23andMe GWAS summary statistics, and the cognitive PGI. The pattern is similar for each of the family members.

Note that the genetic factor identified by this approach captures genes that share a common signal across the different polygenic indexes. In our case, this means that we can identify the genetic factor for cognitive skills relating to educational attainment.

C. *Estimating the Technology of Skill Formation and the Investment Policy Function*

Once the parameters of the measurement system are identified, we can estimate the remaining parameters in the technology of skill formation (equation (8)), the investment function (equation (10)), and in the early skills function (equation (6)). To do so, we again follow Agostinelli and Wiswall (2020) and construct "residual" measures of skills and investments. The residual measures can be used in a regression framework to identify the remaining parameters in the model. Formally, for each measure of latent skills, investments, and genetic factor, we construct "residual measures" by subtracting the estimated measurement mean and dividing by the estimated factor loading, such that

$$\tilde{m}_{ijt}^{\theta} = \frac{m_{ijt}^{\theta} - \mu_{jt}^{\theta}}{\lambda_{jt}^{\theta}} = \ln\theta_{it} + \frac{\nu_{ijt}^{\theta}}{\lambda_{jt}^{\theta}}, \tag{31}$$

$$\tilde{m}_{ikt}^{I} = \frac{m_{ikt}^{I} - \mu_{kt}^{I}}{\lambda_{kt}^{I}} = \ln I_{it} + \frac{\nu_{ikt}^{I}}{\lambda_{kt}^{I}}, \tag{32}$$

$$\tilde{pgi}_i^k = \frac{pgi_i^k}{\lambda_k^G} = G_i + \frac{\zeta_i^k}{\lambda_k^G}. \tag{33}$$

We use these residual measures to estimate the remaining parameters. For example, to estimate the investment policy function (equation (10)), we can use the *k*th residual measurement for the latent investment, the *j*th residual measurement for the latent skill, and the *k*th residual genetic factor for the latent genetic factor instead of the true unobserved latent variables.[21]

[21] In practice, we can use all possible combinations of investments and skill measurements to estimate the model parameters. There are many possible ways to use this large amount of measures. In our preferred specification, our parameters are averages of all possible combinations of measures for each period.

$$\tilde{m}^{I}_{ikt} = \gamma_{1,t}\tilde{m}^{\theta}_{ijt} + \gamma_{2,t}\tilde{pgi}^{k}_{i} + \gamma_{3,t}\tilde{pgi}^{k,m}_{i} + \gamma_{4,t}\tilde{pgi}^{k,f}_{i} + \gamma_{x,t}\mathbf{X}^{I}_{it} + \tilde{\eta}_{it}, \tag{34}$$

where $\tilde{\eta}_{it} = \eta_{it} + \nu^{I}_{ikt}/\lambda^{I}_{kt} - \gamma_{1,t}\left(\nu^{\theta}_{ijt}/\lambda^{\theta}_{jt}\right) - \gamma_{2,t}\left(\zeta^{k}_{i}/\lambda^{G}_{k}\right) - \gamma_{3,t}\left(\zeta^{k,m}_{i}/\lambda^{G,m}_{k}\right) - \gamma_{4,t} \times \left(\zeta^{k,f}_{i}/\lambda^{G,f}_{k}\right)$.

Estimation of equation (34) by OLS would yield inconsistent estimates of the γ coefficients because the residual measures are correlated with their measurement errors, which are included in the residual term $\tilde{\eta}_{it}$. A common solution in the literature, which we follow here, is to use an instrumental variables estimator with the vector of excluded measurements $\left[m^{\theta}_{ij't}\right]$ as instruments for $\tilde{m}^{\theta}_{ijt}$ and $\left[pgi^{k'}_{i}\right]$ as instruments for $\tilde{pgi}^{k}_{i}$. This instrumental variables strategy yields consistent estimators of the γ coefficients. A similar approach is used to estimate the parameters of the technology of skill formation (equation (8)) and early skills function (equation (6)). Since this is an innovation, we also prove the consistency of our IV estimator when using imputed parental polygenic indexes. We show the proof in online Appendix C.

The key identifying assumption is that all shocks and measurement errors are independent of each other and across time. Formally, we array the skill formation shocks ϵ_t in a vector $\boldsymbol{\epsilon}$, the investment shocks η_t in a vector $\boldsymbol{\eta}$, and assume that

$$\epsilon_t \perp\!\!\!\perp \epsilon_{t'}, \ \forall t \neq t', \tag{35}$$

$$\eta_t \perp\!\!\!\perp \eta_{t'}, \ \forall t \neq t', \tag{36}$$

$$\boldsymbol{\epsilon} \perp\!\!\!\perp \boldsymbol{\eta}, \tag{37}$$

$$(\boldsymbol{\epsilon}, \boldsymbol{\eta}) \perp\!\!\!\perp \left(\boldsymbol{\nu}^{\theta}\boldsymbol{\nu}^{I}, \zeta\right). \tag{38}$$

Assumptions (35) and (36) maintain the independence of the shocks over time, and (37) maintains the independence between shocks to investments and skills. This means that we treat shocks and innovations to the investment policy function as exogenous. This is a potentially strong assumption that is commonly made in the literature (see, e.g., Agostinelli and Wiswall 2020). Relaxing this assumption is possible if instruments are available, as in Attanasio et al. (2020). Common instruments are price variations across regions and across time. Unfortunately, our sample was born in the same year and region, making it difficult for the same strategy to be implemented. In addition, we need to assume that the measurement errors are independent of the shocks to investments and skills (assumption (38)). This assumption means that conditional on the latent investments (I_{it}), skills (θ_{it}), and genetic factors (G_i, G^{m}_{i}, G^{f}_{i}), the remaining information on the measurements is unrelated to the process of skill formation.

There are other important assumptions in how we specify our model. In particular, it is worth highlighting that we assume that the relationship between log-latent investments and skills and the measures of those latent variables is linear and homogeneous across families (equation (13)). This assumption could be violated in a variety of ways. One such violation could happen if a given parent-child interaction does not reflect the underlying latent investment in the same way across different families. Say, for example, that for high-PGI parents, taking their child to a "place

of interest" is a strong signal of investment in child skill formation, whereas for low-PGI parents, going to a "place of interest" is a less strong signal of investment, perhaps because the different families go to different places. In this scenario, λ_k^I would be different for different families, which is a violation of our model. Unfortunately, we cannot test for this type of violation, and allowing for heterogeneity in λ_k^I would significantly complicate our estimation and identification strategies.

D. *Inference*

We rely on a bootstrap procedure for inference. We resample the individuals from our initial sample at random with replacement and redo all estimation steps to obtain new model parameters under each new bootstrap sample. The entire procedure is replicated 1,000 times. Using the bootstrap procedure, we compute the 95 percent confidence intervals that are reported in the paper. The procedure takes into account the estimation error at all the steps. Similarly, when comparing estimates from different models, as in Section VC, we do the comparison for each bootstrap sample to compute the 95 percent confidence intervals.

IV. Estimation Results

We present the parameter estimates from the model. We first discuss the estimates for the initial skills function (Section IVA). This is followed by the estimates for the technology of skill formation (Section IVB). Next, we discuss how parental investments are determined (Section IVC). In the following sections, we use empirical simulations to describe the patterns captured by our model. We first show that the effect of genes increases during early childhood and then how the growth of skills is different across the genetic spectrum (Section IVD).

A. *Genetic Influences on Initial Skills*

We present the parameter estimates of equation (6) in the first column of Table 5. We document small and insignificant effects of the child's genes on its initial skills. This means that any documented genetic effects on later skills cannot be explained by the effect of genes on initial skill endowments. We also document positive and significant associations with maternal genes but not with paternal genes for the child's initial skill endowments. These endowments are measured at ages zero to two years, and they are therefore affected by conditions in utero as well as by very early investments after birth. We interpret the association with maternal genes as possibly being driven by more favorable in utero and birth conditions influencing early child development.

B. *Genetic Influences on the Technology of Skill Formation*

The estimated parameters of the technology of skill formation (equation (8)) are presented in the second to sixth columns of Table 5. The parameter on $\ln \theta_{it}$ is the self-productivity parameter, which has the interpretation of an elasticity, capturing the influence of past skills on current skills. The parameter on $\ln I_{it}$ captures the

Table 5—Technology of Skill Formation

	Ages 0–2 (1)	Ages 2–3 (2)	Ages 3–4 (3)	Ages 4–5 (4)	Ages 5–6 (5)	Ages 6–7 (6)
G_i	0.002 [−0.015, 0.019]	0.001 [−0.008, 0.008]	0.001 [−0.016, 0.017]	0.030 [0.014, 0.046]	0.022 [0.008, 0.036]	0.008 [−0.001, 0.017]
G_i^m	0.012 [−0.001, 0.024]	0.003 [−0.002, 0.009]	0.005 [−0.005, 0.017]	0.011 [−0.000, 0.023]	0.011 [0.001, 0.021]	0.001 [−0.005, 0.008]
G_i^f	0.004 [−0.010, 0.019]	0.004 [−0.002, 0.011]	0.015 [0.003, 0.029]	0.003 [−0.009, 0.016]	0.006 [−0.006, 0.018]	0.003 [−0.004, 0.011]
$\ln\theta_{it}$		0.224 [0.088, 0.351]	1.844 [1.214, 2.724]	0.583 [0.340, 0.829]	1.009 [0.565, 1.486]	1.988 [1.175, 2.819]
$\ln I_{it}$		0.098 [0.051, 0.140]	0.696 [0.386, 1.100]	0.329 [0.127, 0.542]	0.931 [0.390, 1.525]	2.122 [1.107, 3.233]
$\ln\theta_{it} \times \ln I_{it}$		−0.005 [−0.031, 0.024]	−0.216 [−0.366, −0.101]	−0.056 [−0.122, 0.010]	−0.184 [−0.345, −0.028]	−0.552 [−0.855, −0.263]
$\ln A$	1.409 [1.391, 1.428]	1.818 [1.604, 2.038]	−2.375 [−4.855, −0.652]	0.936 [0.187, 1.706]	−0.808 [−2.525, 0.736]	−3.939 [−7.011, −1.051]
$E[\ln\theta_{it+1}]$	1.433	2.599	3.034	3.278	3.428	3.505
$\text{var}(\ln\theta_{it+1})$	0.033	0.031	0.061	0.090	0.090	0.034

Notes: This table reports the parameter estimates for the initial skill function (equation (6)) in the first column and the parameter estimates for the technology of skill formation (equation (8)) at different child ages in columns 2–6. We report the associated 95 percent bootstrap confidence intervals in brackets.

returns to investments. The complementarity between skills and investments is captured by the parameter on $\ln\theta_{it} \times \ln I_{it}$. All three parameters vary significantly across periods and tend to increase over time. Our estimates for the later periods of development are similar to those documented in Agostinelli and Wiswall (2020), including the negative interaction (substitutability) between skills and investments. Agostinelli and Wiswall (2020) use a similar translog specification and focus on children aged five to six years onward, making our estimates at the same ages and using a similar specification comparable. In online Appendix D.2, we also estimate a full translog model that allows for interactions between all the terms. However, as this specification does not significantly change any of our findings, and most of the additional interaction terms are insignificant, we prefer the more restrictive model as our main specification.

We extend the traditional model to allow genetic effects to influence skill formation. We find that both children's and their parents' genes matter for the children's skill development. The direct effect on child skills is negligible in the initial periods of development but increases significantly starting at age four. The effect of the child's own genetic factor captures the ability to acquire new skills, given the current stock of skills and parental investments (the direct effect in Section ID). Even after allowing parental genes to affect investments, parental genes still have a nonnegligible association with skill formation that is not captured by investments. We interpret these as capturing the quality of the family environment that can be explained by family genetics. It could include, for example, genetic influences on the quality of parenting. These influences accumulate over time. In Section IVD, we provide more insights into the importance of genes by analyzing the importance of the cumulative genetic mechanisms in explaining skill formation over time as well as variation in skill formation across individuals.

Table 6—Investment Policy Function

	Ages 0–2 (1)	Ages 2–3 (2)	Ages 3–4 (3)	Ages 4–5 (4)	Ages 5–6 (5)
G_i	0.026 [0.006, 0.045]	0.022 [−0.005, 0.049]	0.012 [−0.002, 0.026]	−0.002 [−0.008, 0.005]	−0.001 [−0.007, 0.005]
G_i^m	0.038 [0.023, 0.054]	0.079 [0.058, 0.100]	0.030 [0.020, 0.041]	0.019 [0.013, 0.025]	0.013 [0.008, 0.018]
G_i^f	0.022 [0.004, 0.039]	0.036 [0.013, 0.058]	0.019 [0.007, 0.031]	0.015 [0.009, 0.021]	0.016 [0.009, 0.021]
$\ln\theta_{it}$	0.339 [0.268, 0.405]	0.719 [0.597, 0.862]	0.182 [0.134, 0.230]	0.091 [0.067, 0.115]	0.107 [0.076, 0.133]
Constant	4.144 [4.047, 4.248]	2.650 [2.269, 2.976]	3.005 [2.846, 3.159]	2.493 [2.402, 2.582]	2.416 [2.314, 2.537]
$E[\ln I_{it}]$	4.643	4.528	3.557	2.809	2.796
$\text{var}(\ln I_{it})$	0.112	0.182	0.058	0.010	0.007

Notes: This table reports the parameter estimates for the investment policy function (equation (10)) at different child ages in columns 1–5. We report the associated 95 percent bootstrap confidence intervals in brackets.

C. *Genetic Influences on Parental Investments*

The estimated parameters for the investment policy function (equation (10)) are presented in Table 6. The most interesting results are those relating to genes and their interpretation. The impact of the child's genes on parental investments is the direct nurture of nature effect, reflecting how parents respond to their children's genetics conditional on their current stocks of skills. This effect is large and significant in the first period but then decreases gradually in magnitude over time. Children with higher genetic factors behave in ways that elicit their parents to spend more time with them, even after conditioning on their cognitive skill levels.

However, we also find that the child's genes indirectly influence parental investments through the stock of skills. We find that parents reinforce initial skill differences and invest more in high-skilled children. The reinforcing behavior is an indirect form of the nurture of nature effect since the stock of skills is, in part, determined by the child's genetics. Hence, children with a higher genetic factor will tend to have a higher stock of skills, leading parents to invest more in their skill development. This reinforcing behavior is also stronger earlier in life.

We also document a large association between parental genetic factors and parental investments. This association captures a variety of family genetic influences on the environment experienced by the child. The association is significant in all periods, and it is stronger for maternal genes than for paternal genes, in particular until around the age of four. In Section VB, we show that most of these influences are mediated through parental education.

D. *Genetic Influences across Early Ages*

To make better sense of the magnitude and importance of the genetic influences, we graphically depict the relationship between the child's genetic factor and the child's predicted skills at different ages. To do so, we simulate the evolution of skills and investments 1,000 times for each individual using the model parameters. We

then standardize the stock of skills at each period for ease of comparison. The results are shown in Figure 3.

Panel A of Figure 3 shows the effect of the child's own genetic factor on the standardized stock of skills at different ages conditional on parental genes, which are held fixed. Two important results are evident; first, the effects are relatively large, with a one standard deviation increase in the genetic factor leading to about a 0.18 standard deviation increase in the stock of skills at ages six to seven years. Second, the effect of genes accumulates over time. Before age four, the same increase in the genetic factor leads to a negligible increase in the stock of skills. This is consistent with Belsky et al. (2016), who find that the association between EA PGI and cognitive ability increases from age 3 to age 11, and with earlier twin and adoption studies showing an increasing heritability of IQ with age (Bouchard 2013; Tucker-Drob, Briley, and Harden 2013). Our contribution is that the dynamic factor model allows us to control for differences in measurement error across periods, thereby allowing us to rule out that this pattern is simply an artifact of skills being less precisely measured at earlier ages. In addition, our approach also allows us to control for measurement error in the genetic factors. In online Appendix D.1, we show that not performing this correction would make us underestimate the effects of child genetics.

It is not only child genes that influence skill formation; parental genes are also associated with skill formation. Panels B and C of Figure 3 show the relationship between the maternal and paternal genetic factors and the child's stock of skills by age. Once again, the associations are relatively large. One standard deviation increase in the maternal genetic factor is related to a 0.12 increase in the child's stock of skills at ages six to seven. For fathers, the association is slightly smaller, at 0.08. We also observe an increasing importance of parental genes by age, although the increase is less clear than that for the child's own genes.

V. Mechanisms and Implications

The model also allows us to better understand genetic mechanisms and what is gained from including genetics in models of skill formation. We demonstrate this with three different counterfactual simulations. First, to illustrate the importance of the different genetic mechanisms, we simulate what happens to the link between child genes, investments, and the distribution of skills when we remove nurture of nature effects and family genetic associations. Second, to better understand the role of family genetic influences, we reestimate our model with different family controls ($\mathbf{X}_{it}$) that may mediate the influence of parental genes on skill formation. Third, to better understand what is gained from including genetics in models of skill formation, we compare our baseline model to the same model without the child and parental genes. We then compare the estimated effect and genetic heterogeneity of parental investments in both models. This exercise helps us understand what is missed and quantifies the possible biases from neglecting genetics in traditional models of skill formation.

A. *Mechanism Decomposition*

We perform the first simulations to illustrate the relative contribution of the nurture of nature and family genetic associations. We assess how the relationship

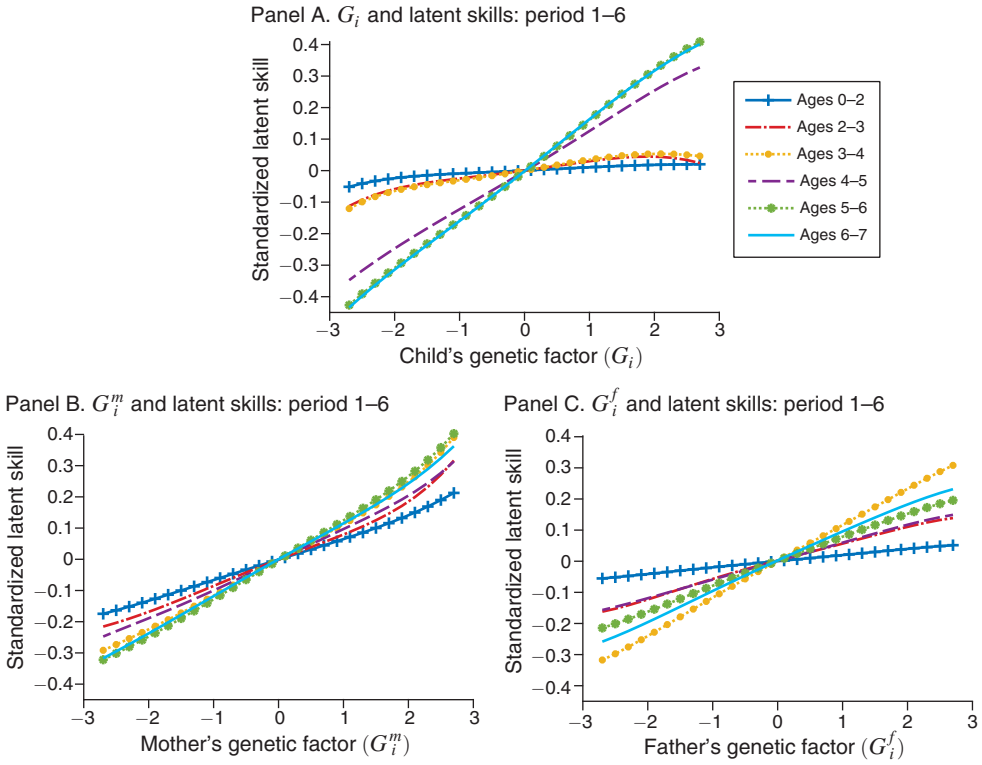

FIGURE 3. ASSOCIATIONS BETWEEN GENETIC FACTORS AND LATENT SKILLS ACROSS CHILD DEVELOPMENT

Notes: These figures plot the relationship between the child's and its parents' genetic factors and the child's latent skill at different ages. Using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's, the mother's, and the father's genetic factors while keeping the remaining indexes constant. This figure highlights how the associations between skills and both parental and child genes increase over time.

between genes, investments, and the distribution of skills change in counterfactual worlds where the two mechanisms do not exist.

Our main findings show that parental behavior is reinforcing, as parents invest more in children with higher genetic factors (the nurture of nature effect). However, this pattern may be different in other contexts, as suggested by the literature on whether parents reinforce or compensate for initial differences in skills (see, e.g., Hsin 2012; Grätz and Torche 2016; Nicoletti and Tonei 2020). In one specification, we, therefore, simulate the distribution of investments and skills in a counterfactual world with no nurture of nature, i.e., where parental investments are independent of child endowments.

Our main findings also show a significant influence of parental genes on skill formation, which captures family genetic influences. While this pattern is likely to hold across most contexts, there is at least one particular case where this link is eliminated, namely when children are adopted or placed in out-of-home care. In another specification, we, therefore, simulate the link between genes, investments, and skills in a counterfactual world with no family genetic influences, i.e., where parental genes only matter by being transmitted to the child.

Panels A and B of Figure 4 plot the standardized simulated latent investments at ages zero to two and five to six years under the different specifications as a function of the child's genetic factor. The blue line corresponds to the predicted association between the child's genetic factor and latent investments from our empirical model. It is not a causal effect, as it captures all three mechanisms, including family genetic associations. The figures show that children with higher genetic factors receive higher levels of parental investments in earlier and later periods. The predicted associations are large; each one point increase in the standard deviation of the genetic factor is associated with a 0.2 to 0.3 standard deviation increase in parental investments. Controlling for family genetic associations significantly decreases the overall association, as shown by the red line. As shown in panel B of Table 7, between 57 and 90 percent of the association between the genetic factor and parental investments is explained by parental genes. The importance of family genetic associations is larger at later stages of development. The remaining association is due to the nurture of nature channel, as shown by the change from the red to the yellow line in panels A and B of Figure 4. We thus show that both mechanisms contribute to the observed inequality in investments.

Panels C and D of Figure 4 present the corresponding relationship between the child's genetic factor and standardized latent skills at ages two to three and six to seven years under the different scenarios. Again, the blue line depicts the predicted association from our empirical model, and we see that the predicted skill gradient is substantial. Next, we illustrate to what extent the inequality in skills can be explained by the inequality due to family genetic associations. The red line illustrates the relationship between the child's genetic factor and skills in the counterfactual world without the influence of parental genes. Eliminating the association between parental genes and investments and skills reduces the association between the child's genetic factor and skills by between 41 to 82 percent, as shown in panel A of Table 7. In contrast to investments, the influence of family genetics on skills is stronger in earlier developmental periods. The yellow line in panels C and D of Figure 4 depicts the predicted relationship between the child's genetic factor and skills when we further control for the nurture of nature effect. Doing so, we see a significant decrease in the effect of genes in earlier periods of about 18 percent but a small decrease in later periods of only about 3 percent. The remaining relationship is due to direct genetic effects. These are more important at later stages of development, accounting for up to 56 percent of the association between the genetic factor and skills.

This exercise serves two purposes. First, it gives a sense of the relative importance of the different channels of genetic influence. We see that family genetic influences and the nurture of nature both contribute to the gradient between child genes and parental investments, and child skills. Family genetic associations are relatively more important earlier on for skills and later on for parental investments. Second, the simulations illustrate a more general point, namely that genetic effects are not independent of investments or the environment in general. Rather, genes and investments interact, and the reason that we see large genetic inequalities is partly that parents respond to (and reinforce) such inequalities.

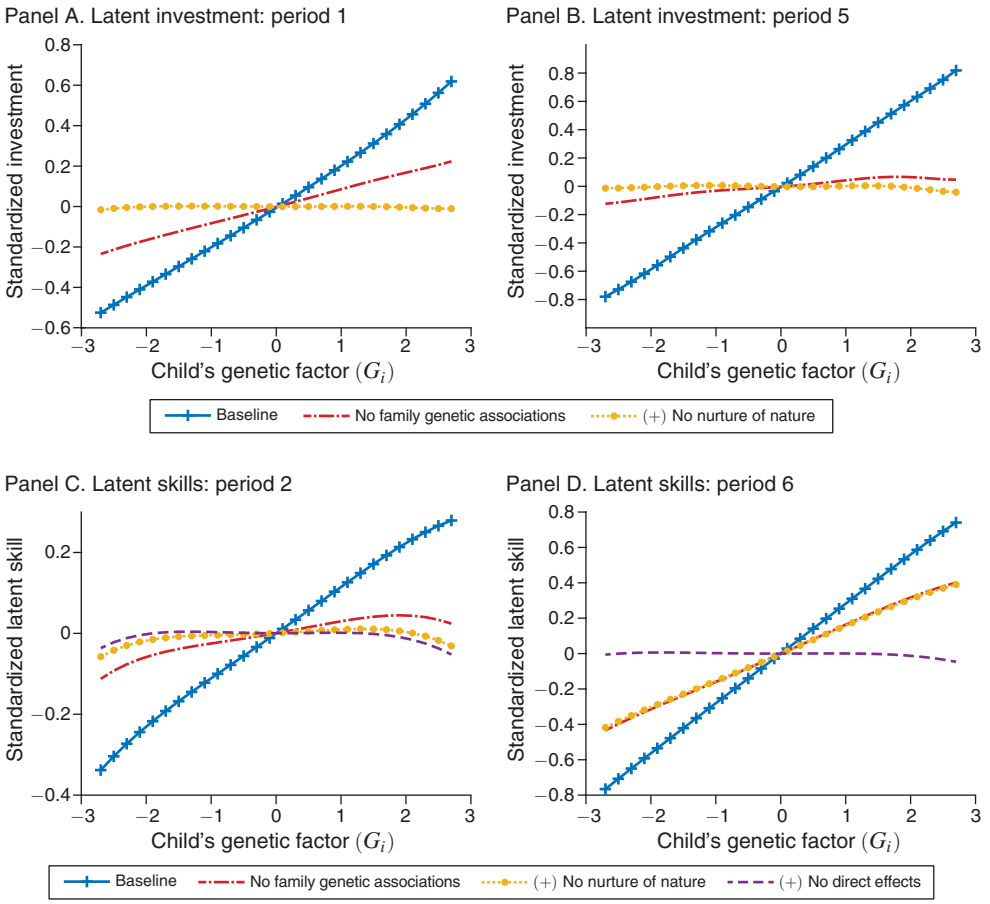

FIGURE 4. MECHANISMS DECOMPOSITION

Notes: This figure illustrates the importance of the different genetic channels for the relationship between the child's genetic factor and latent child skills and parental investments in the first and last periods. We first plot the observed relationship (baseline). Using the estimated model parameters, we then simulate expected latent skills/investments when we exogenously increase the child's genetic factor while holding the mother's and father's genetic factor constant (no family genetic associations). We then additionally shut down the link between child genes and investments (no nurture of nature), and, for latent skills, the link between child genes and skills (no direct effects).

B. *Decomposing Family Genetic Associations*

In our second set of simulations, we aim to understand the mechanisms by which family genetics influence children's skill formation. To do so, we reestimate our model using different sets of control variables in $\mathbf{X}_{it}$. In our baseline model, $\mathbf{X}_{it}$ includes only information about the child's sex. We reestimate our model, adding different observed family variables to $\mathbf{X}_{it}$ one by one. The full set of controls also includes dummies for different parental education levels, dummies for parental occupational groups, and information on family size and family income.

Table 7—Mechanisms Decomposition by Age

	Ages 0–2	Ages 2–3	Ages 3–4	Ages 4–5	Ages 5–6	Ages 6–7
Panel A. Child's skills						
Direct effects	18.09%	6.94%	4.83%	51.76%	55.82%	55.44%
Nurture of nature	0.00%	18.08%	13.33%	7.50%	2.90%	2.72%
Family genetic associations	81.91%	74.98%	81.84%	40.75%	41.28%	41.84%
Panel B. Parental investments						
Nurture of nature	42.86%	27.90%	37.14%	9.88%	18.74%	
Family genetic associations	57.14%	72.10%	62.86%	90.12%	81.26%	

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

This is of interest for three reasons. First, a better understanding of the mechanisms of transmission is key if one decides to design policies to decrease genetic influences on skills. Second, understanding what observed factors account for family genetic influences can help researchers perform similar types of analyses in datasets where parental genes are not observed. Lastly, this acts as a validation exercise to our identification strategy. We assume that variation in the child's genetic factor is random conditional on the parents' genetic factors. If this is true, we should expect the estimated effect of the child's genetic factor to be unchanged, even though parental genetic influences are expected to change once we add different family controls to the model.

We summarize the results from this exercise in Figure 5 and present the estimated coefficients in online Appendix D.3. The figure plots the estimated relationship between the child's and her parents' genetic factors and her skills at ages six to seven. The baseline model (in blue) shows the estimated relationships for each of the three genetic factors while keeping the remaining indexes constant as in Figure 3. We then allow maternal education to enter the control set ($\mathbf{X}_{it}$) and plot the estimated relationships in red. The effect of the child's genetic factor remains unchanged, but the associations between the maternal and paternal genetic factors and child skills decrease substantially. We further control for paternal education and plot the estimated relationships in yellow. Again, the effect of the child's genetic factor is unchanged, but now the lines for the parental genetic factors are flat. Controlling for the education of both parents completely mediates the association between parental genes and child skills. Including additional family controls only changes the relationship between genes and skills marginally.

This exercise shows that the estimated family genetic associations mainly operate via the association between parental genes and parental educational attainment. Other factors seem unimportant once we control for parental education. This exercise also shows that controlling for parental education, as is commonly done in the literature when parental genes are not available, seems to be a valid approach. Lastly, the fact that the child's genetic effects are unchanged as we vary the control set ($\mathbf{X}_{it}$) validates our identification strategy.

C. *Ignoring Genes in Models of Skill Formation.*

In this last set of simulations, we want to understand better the consequences of not including genetic data in models of skill formation. To do so, we reestimate the

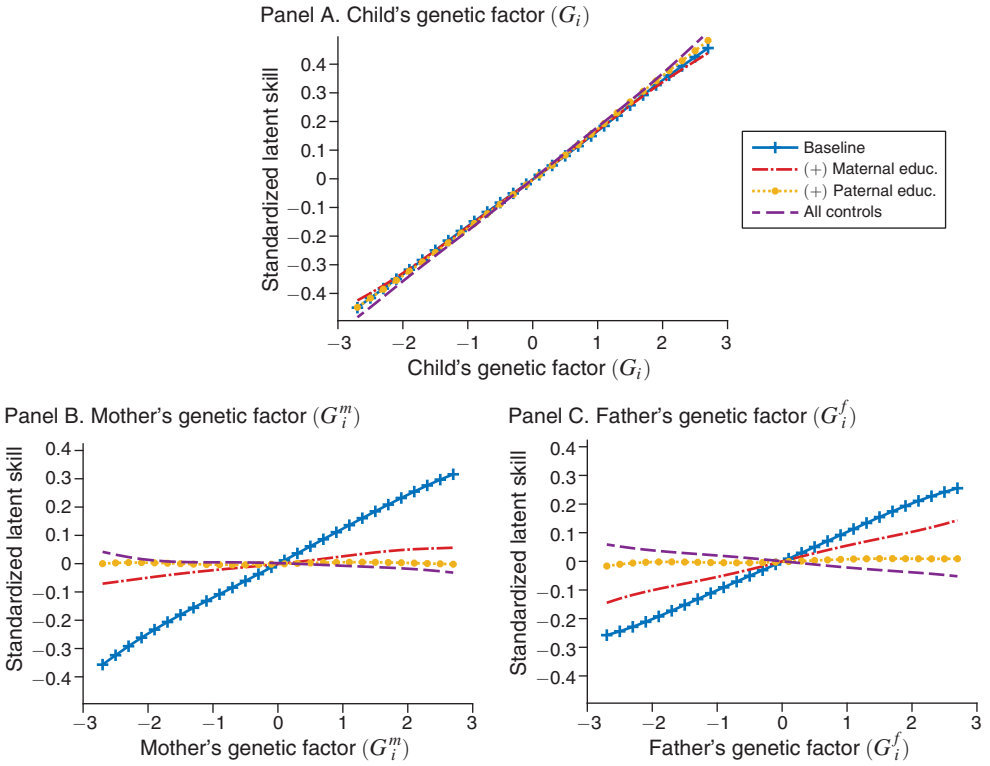

FIGURE 5. PARENTAL CONTROLS

Notes: These figures compare how the relationship between the child's and its parents' genetic factors and the child's latent skill at ages six to seven changes once different family controls are added to our estimation model. The baseline set of controls $\mathbf{X}_{it}$ includes only the child's sex. We add different family controls one by one to the analysis. The full set of controls includes dummies for different parental education levels, dummies for parental occupational groups, and information on family size and family income. For each set of controls, we reestimate the model, and using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's and the parental genetic factor while keeping the other constant. This figure highlights how parental education explains the majority of the family genetic associations but does not explain the effect via the child's genes.

model without any genetic factors. We do this both for the baseline model (without controls) and the model from Section VB with the full set of parental controls. We present the full set of estimates in online Appendix D.4 (Tables D14–D17).

To compare the models, we focus on the estimate of the returns to investments. The sensitivity of skills with respect to investments is a key parameter in the skill formation literature (Cunha and Heckman 2007). Cunha and Heckman (2008) illustrate differences in sensitivity across different skills and periods by simulating the effect of an exogenous increase of 10 percent in investments. We follow their simulation exercise to understand whether omitting genes from models of skill formation can lead to bias in the estimates of the sensitivity of parental investments.[22]

[22] While we follow the simulation approach in Cunha and Heckman (2008), our estimates and goals are not directly comparable. First, while we are interested in understanding the consequence of omitting genes from our model, they are interested in understanding which periods are sensitive for investments in cognitive and noncognitive skills. Second, the estimates themselves are not directly comparable since they anchor the returns to skills to adult outcomes, which we cannot do because of a lack of data availability.

Table 8—Effect of a 10 Percent Increase in Investments

	Ages 2–3	Ages 3–4	Ages 4–5	Ages 5–6	Ages 6–7
Panel A. Without controls					
Baseline model	0.92%	2.19%	2.47%	4.48%	3.86%
	[0.75, 1.07]	[1.77, 2.70]	[2.03, 2.99]	[3.62, 5.71]	[3.12, 5.01]
Model without genes	0.93%	2.27%	2.66%	4.93%	4.23%
	[0.76, 1.08]	[1.84, 2.81]	[2.21, 3.23]	[4.01, 6.25]	[3.41, 5.47]
Percentage difference	1.25	3.79	7.74	10.03	9.38
	[0.35, 2.53]	[2.07, 6.77]	[5.70, 10.97]	[7.14, 13.52]	[6.16, 13.20]
Panel B. With controls					
Baseline model	0.80%	1.47%	1.44%	2.79%	2.37%
	[0.65, 0.94]	[1.16, 1.85]	[1.13, 1.79]	[2.09, 3.77]	[1.78, 3.18]
Model without genes	0.80%	1.47%	1.47%	2.85%	2.42%
	[0.65, 0.94]	[1.15, 1.86]	[1.14, 1.82]	[2.15, 3.86]	[1.82, 3.23]
Percentage difference	0.03	0.27	1.61	2.19	2.09
	[−0.37, 0.55]	[−0.83, 1.29]	[0.24, 3.33]	[0.50, 4.21]	[0.19, 4.34]

Notes: This table compares the effects of a 10 percent increase in investments on latent child skills at different periods, implied by the baseline model and the model without genes. The last row shows in percentages how much the model without genes overestimates the returns to investments. We report the associated 95 percent bootstrap confidence intervals in brackets.

The results are reported in Table 8. In panel A, we compare the model without controls, while in panel B, we compare the model with the full set of controls. In the model without controls, we see that a permanent increase in investments of 10 percent leads to a 3.86 percent increase in average skills at ages 6–7. In comparison, the corresponding estimate without genes is 4.23 percent. The estimated returns using the model without genes are significantly overestimated in all periods by up to 10 percent. However, when adding the full set of controls, the estimates become more similar, and the model without genes only overestimates the returns to investments by up to 2 percent. Hence, this again shows the importance of detailed family controls to minimize the bias in standard models of skill formation when genetic data are unavailable.

Another important question in the literature is whether the returns to investments differ across individuals and whether some relevant child or family characteristic can capture it. This heterogeneity is important in describing how inequality in skills is generated. For example, Agostinelli and Wiswall (2020) estimate heterogeneity in the returns to investments by household income and show that the returns to investments are highest for children from low-income families. Such estimates are important for assessing how different policies will affect the distribution of skills and the overall degree of inequality in the population. An additional advantage of our model is that we can estimate the heterogeneous returns to investments across the child's genetic factor. While observable characteristics such as household income capture many different correlated factors, we can causally attribute the heterogeneity to the genetic factor because variation in the child's genetics is random after conditioning on parental genetic factors.

Figure 6 shows how returns to investments are heterogeneous across children's genetics. The blue line plots the percentage change in skills resulting from the permanent investment increase as a function of the child's genetic factor. Returns to

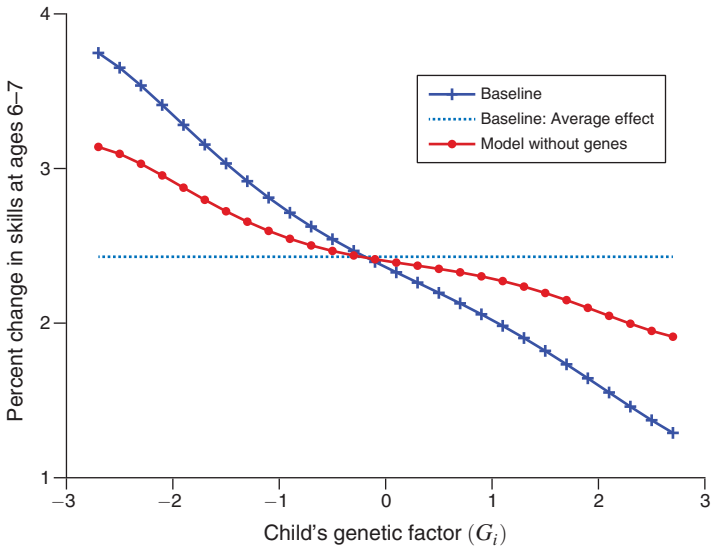

FIGURE 6. RETURNS TO INVESTMENTS BY G_i

Notes: This figure plots the returns to investments implied by the baseline model and the model without genes across the child's genetic factor. For the baseline model, the dotted blue line illustrates the average effect, while the solid blue line illustrates the actual effect, which is heterogeneous across the child's genetic factor. The red line shows how this effect is misestimated by the model without genes.

investments are heterogeneous across children's genetics, and children with lower genetic factors face the highest returns to investments. This finding is particularly interesting given the fact that parents in our sample tend to invest more in the high–genetic factor children, even though the returns are lower for them. However, if investments were to increase proportionally for all children—for example, through social policy—the distribution of skills would become more equal. This suggests that there is no trade-off between equity and efficiency with respect to investments in children's skills at early ages.

Again, we can demonstrate what happens when we omit genes from the model. The red line in Figure 6 plots the observed correlation between the returns to investments and the child's genetic factor using the parameter estimates from the model without genes. There are two main points to highlight. First, the fact that this line is not flat indicates that the genetic heterogeneity is misattributed to environmental variables included in the model (e.g., parental income) that are correlated with the (unobserved) genetic factor. Second, the fact that the red line is flatter than the blue line shows that a significant part of the heterogeneity in the returns to investments is missed when genes are omitted from child development models. That is, omitting genes from models of child development will underestimate the heterogeneity in the returns to investments and misattribute genetic heterogeneity to environmental variables.

VI. Discussion and Conclusion

To better understand the interplay between genetics and family resources for skill formation and its relevance for policy, we incorporated genetic factors into a

dynamic model of skill formation. We modeled and estimated the joint evolution of skills and parental investments throughout early childhood (ages zero to seven years). We observed both child and parental genotypes, allowing us to estimate the independent effect of the child's genes on skill formation and to estimate the association between family genetics and child development. Furthermore, by incorporating (child and parent) genetics into a formal model, we were able to evaluate the mechanisms through which genes influence skill formation and describe what is missed from excluding genes from models of skill formation.

Using the model, we document the importance of both parental and child genes for child development. We show that the effect of genes increases over the child's early life course and that a large fraction of these effects operate via parental investments. Genes directly influence the accumulation of skills; conditional on their current stock of skills and parental investments, genetics make some children better able to retain and acquire new skills (the direct effect). In addition, we show that genes indirectly influence investments, as parents reinforce genetic differences by investing more in children with higher skills (the nurture of nature effect). We also find that family genetic influences matter and that these influences are mediated through parental education. The impact of genes on parental investments is especially significant, as it implies an interplay between genetic and environmental effects. These findings illustrate that nature and nurture jointly influence children's skill development, a finding that highlights the importance of integrating biological and social perspectives into a single framework. We also demonstrate the importance of considering genetic influences for the identification of structural parameters in models of child development. We show that both the average return to investments and the heterogeneity in these returns will be misestimated when genes are ignored in skill formation models since genetic influences will be either completely missed or misattributed to nongenetic factors.

A limitation of our work is that genetic factors are measured using polygenic indexes for educational attainment and cognition. It is possible that genes unrelated to these outcomes nevertheless also influence children's early life skill formation. For example, genetic variation related to mental health and altruism may potentially be unrelated to education but might influence how parents interact with their children. If this is true, we are missing the parts of the genetic contribution to skill formation that work through such alternative channels. Another limitation of our work is that we only observe children's cognitive skills. It is possible that the mechanisms for the formation of other child skills are different. It is also possible that some of our estimated genetic mechanisms can be explained by the effect of the child's genes on early socioemotional skills. Future work should explore these other dimensions.

Lastly, it is important to recognize that our findings may not generalize to other contexts and other groups of individuals. In particular, our sample is restricted to individuals of European ancestry. This issue is common to the broader genetics literature since the majority of polygenic indexes are constructed from GWAS performed in Europeans, and their transferability to other populations depends on many factors (see Martin et al. 2017 for a discussion of the transferability issues of GWAS results and Mostafavi et al. 2020 for an empirical comparison of PGIs across ethnic groups). This also illustrates a problem of inequity in research, where the only individuals being studied are those of European ancestry. This opens the possibility

that other ancestry groups will not benefit from the advances in genetics research (see the discussion in Martin et al. 2019). While the key insights from our research apply to all ancestry groups, we cannot test for any differences in the role of genetics across groups until we have solved the transferability issue. We hope future work will address these issues and lead to a more inclusive research agenda.

REFERENCES

1000 Genomes Project Consortium. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526 (7571): 68–74.

Agostinelli, Francesco, and Matthew Wiswall. 2020. "Estimating the Technology of Children's Skill Formation." NBER Working Paper 22442.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2015. "Human Capital Development and Parental Investment in India." NBER Working Paper 21740.

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.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2018. "Education can Reduce Health Differences Related to Genetic Risk of Obesity." *Proceedings of the National Academy of Sciences* 115 (42): E9765–72.

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. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84 (4): S143–62.

Behrman, Jere R., Robert A. Pollak, and Paul Taubman. 1982. "Parental Preferences and Provision for Progeny." *Journal of Political Economy* 90 (1): 52–73.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-course Development." *Psychological Science* 27 (7): 957–72.

Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arseneault, Jason D. Boardman, Avshalom Caspi, Dalton Conley, et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31): E7275–84.

Boneva, Teodora, and Christopher Rauh. 2018. "Parental Beliefs about Returns to Educational Investments - The Later the Better?" *Journal of the European Economic Association* 16 (6): 1669–1711.

Bouchard, Thomas J. 2013. "The Wilson Effect: The Increase in Heritability of IQ with Age." *Twin Research and Human Genetics* 16 (5): 923–30.

Boyd, Andy, Jean Golding, John Macleod, Debbie A. Lawlor, Abigail Fraser, John Henderson, Lynn Molloy, Andy Ness, Susan Ring, and George Davey Smith. 2012. "Cohort Profile: The 'Children of the 90s' - The Index Offspring of the Avon Longitudinal Study of Parents and Children." *International Journal of Epidemiology* 42 (1): 111–27.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* 92 (1): 109–40.

Breinholt, Asta, and Dalton Conley. 2020. "Child-Driven Parenting: Differential Early Childhood Investment by Offspring Genotype." NBER Working Paper 28217.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "2001 Lawrence R. Klein Lecture: Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2): 361–422.

Cesarini, David, and Peter M. Visscher. 2017. "Genetics and Educational Attainment." *NPJ Science of Learning* 2: 4.

Cunha, Flávio, Irma Elo, and Jennifer Culhane. 2013. "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation." NBER Working Paper 19144.

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

Cunha, Flávio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (4): 738–82.

Cunha, Flávio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2): 191–261.

Cunha, Flávio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

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): 1–13.

Fraser, Abigail, Corrie Macdonald-Wallis, Kate Tilling, Andy Boyd, Jean Golding, George Davey Smith, John Henderson, et al. 2012. "Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers Cohort." *International Journal of Epidemiology* 42 (1): 97–110.

Grätz, Michael, and Florencia Torche. 2016. "Compensation or Reinforcement? The Stratification of Parental Responses to Children's Early Ability." *Demography* 53 (6): 1883–1904.

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

Hill, William G., Michael E. Goddard, and Peter M. Visscher. 2008. "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." *PLoS Genetics* 4 (2): 1–10.

Houmark, Mikkel, Victor Ronda, and Michael Rosholm. 2024. "Replication data for: The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." American Economic Association [Publisher], Inter-university Consortium for Political and Social Research [Distributor]. https://doi.org/10.3886/E192844V1.

Hsin, Amy. 2012. "Is Biology Destiny? Birth Weight and Differential Parental Treatment." *Demography* 49 (4): 1385–1405.

Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdottir, et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374): 424–28.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, M. Johannesson, et al. 2018. "Gene Discovery and Polygenic Prediction from a 1.1-Million-Person GWAS of Educational Attainment." *Nature Genetics* 50: 1112–21.

Martin, Alicia R., Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, and Eimear E. Kenny. 2017. "Human Demographic History Impacts Genetic Risk Prediction Across Diverse Populations." *American Journal of Human Genetics* 100 (4): 635–49.

Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51 (4): 584–91.

Mostafavi, Hakhamanesh, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K. Pritchard, and Molly Przeworski. 2020. "Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group." *ELife* 9: 1–52.

Nicoletti, Cheti, and Valentina Tonei. 2020. "Do Parental Time Investments React to Changes in Child's Skills and Health?" *European Economic Review* 127: 103491.

Okbay, Aysu, Jonathan P. Beauchamp, Mark Alan Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, Patrick Turley, et al. 2016. "Genome-wide Association Study Identifies 74 loci Associated with Educational Attainment." *Nature* 533 (7604): 539–42.

Papageorge, Nicholas W., and Kevin Thom. 2019. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3): 1351–99.

Plomin, Robert, and Sophie von Stumm. 2018. "The New Genetics of Intelligence." *Nature Reviews Genetics* 19 (3): 148–59.

Plomin, Robert, John C. DeFries, and John C. Loehlin. 1977. "Genotype-Environment Interaction and Correlation in the Analysis of Human Behavior." *Psychological Bulletin* 84 (2): 309–22.

Polderman, Tinca J. C., Beben Benyamin, Christiaan A. de Leeuw, Patrick F. Sullivan, Arjen van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies." *Nature Genetics* 47 (7): 702–09.

Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W. Martin, Harm-Jan Westra, et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340 (6139): 1467–71.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm. 2022. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." *Scandinavian Journal of Economics* 124 (2): 550–78.

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

Silventoinen, Karri, Aline Jelenkovic, Reijo Sund, Antti Latvala, Chika Honda, Fujio Inui, Rie Tomizawa, et al. 2020. "Genetic and Environmental Variation in Educational Attainment: An Individual-Based Analysis of 28 Twin Cohorts." *Scientific Reports* 10 (1): 1–11.

Tucker-Drob, Elliot M., and Daniel A. Briley. 2014. "Continuity of Genetic and Environmental Influences on Cognition across the Life Span: A Meta-Analysis of Longitudinal Twin and Adoption Studies." *Psychological Bulletin* 140 (4): 949–79.

Tucker-Drob, Elliot M., Daniel A. Briley, and K. Paige Harden. 2013. "Genetic and Environmental Influences on Cognition across Development and Context." *Current Directions in Psychological Science* 22 (5): 349–55.

Ward, Mary E., George McMahon, Beate St Pourcain, David M. Evans, Cornelius A. Rietveld, Daniel J. Benjamin, Philipp D. Koellinger, David Cesarini, George Davey Smith, and Nicholas J. Timpson. 2014. "Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children." *PloS One* 9 (7): 1–7.

Wertz, Jasmin, Terrie E. Moffitt, Jessica Agnew-Blais, Louise Arseneault, Daniel W. Belsky, David L. Corcoran, Renate Houts, et al. 2020. "Using DNA from Mothers and Children to Study Parental Investment in Children's Educational Attainment." Unpublished.

Young, Alexander I., Michael L. Frigge, Daniel F. Gudbjartsson, Gudmar Thorleifsson, Gyda Bjornsdottir, Patrick Sulem, Gisli Masson, Unnur Thorsteinsdottir, Kari Stefansson, and Augustine Kong. 2018. "Relatedness Disequilibrium Regression Estimates Heritability Without Environmental Bias." *Nature Genetics* 50 (9): 1304–10.

Young, Alexander I., Seyed Moeen Nehzati, Chanwook Lee, Stefania Benonisdottir, David Cesarini, Daniel J. Benjamin, Patrick Turley, and Augustine Kong. 2020. "Mendelian Imputation of Parental Genotypes for Genome-Wide Estimation of Direct and Indirect Genetic Effects." Unpublished.