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

Mikkel Aagaard Houmark† Victor Ronda‡ Michael Rosholm§

Abstract: We incorporate genetic endowments into a dynamic model of skill formation to evaluate the mechanisms through which genes influence skill formation. We document the importance of three distinct genetic mechanisms: the direct effect of child genes on skills, the indirect effect of child genes via parental investments (nurture of nature), and the effect of parental genes on parental investments (nature of nurture). Using the model, we show the relative importance of genes depends on how parental (or public) investments are allocated across children. Our work highlights the importance of integrating biological and social perspectives into a single framework.

Keywords: Skill Formation; Dynamic Factor Analysis; Social Science Genomics; ALSPAC
JEL Classification: D10 I24 J13 J24

*Current draft: December 21, 2020. We gratefully acknowledge helpful comments from Elena Mattana, Nicholas Papageorge, and seminar participants at Aarhus University. 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. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the work of the authors and they will serve as guarantors for the contents of this paper.

†Department of Economics and Business Economics, Aarhus University (e-mail: mhoumark@econ.au.dk).

‡Department of Economics and Business Economics and TrygFonden's Centre for Child Research, Aarhus University (e-mail: ronda@econ.au.dk).

§Department of Economics and Business Economics and TrygFonden's Centre for Child Research, Aarhus University and IZA (e-mail: rom@econ.au.dk).

1 Introduction

It is widely recognised 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 genetic endowments 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.

In this paper, we incorporate genetic endowments into a model of skill formation and document how the interplay between genes and family investments arises from two mechanisms. First, we show that parents respond to their children's genetic endowments by investing more in children with a high genetic propensity for education. We call this the *nurture of nature* effect. Second, we show that parents who themselves have a high genetic propensity for education also invest more resources in their children. We call this the *nature of nurture* effect (this mechanism is also referred to as genetic nurture in the literature).[1] By incorporating these two mechanisms into an economic model, we formalise ideas in the genetics literature going back to at least Plomin, DeFries, and Loehlin (1977). The formal treatment in an economic model of skill formation allows us to better understand the role of genes in the skill formation process. In particular, it allows us to show that the importance of genes in explaining inequalities in skills later in life can differ significantly according to whether parental investments and government policies compensate for or reinforce initial inequalities. It also allows us to explain why the importance of genes can differ significantly across countries with different social policies and family cultures (Branigan, McCallum, and Freese, 2013).

The empirical estimation relies on detailed genetic and survey data from a longitudinal

[1] In addition to such gene-by-environment correlations, genes can also influence skill formation by what is called gene-by-environment interactions, which refer 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., 2020). 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 Appendix D, we consider this possibility. However, we do not find support in the data for these interactions being important during early skill formation.

British data set, the Avon Longitudinal Study of Parents and Children (ALSPAC). We observe detailed molecular genetic data from the child participants and both of their parents. We measure genetic endowments using a polygenic score for educational attainment (EA PGS). Variations of this score 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., 2020), 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 pre-determined, 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 socio-economic outcomes may be partially driven by the individual's childhood environment. One common solution to this challenge is to directly control for differences in family environment as in Barth, Papageorge, and Thom (2020); Papageorge and Thom (2019). Another is to exploit random genetic variation across siblings as in Ronda et al. (2020). Our approach allows us to directly control for the influence of parental genes, which might affect the child's skills both directly and through parental investments. 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.

To identify the genetic influences on skill formation, we incorporate genetic endowments 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 identify the different genetic mechanisms, control for measurement error, identify latent skills and investments, and compare genetic influences across child ages. This is possible since, in addition to the families' genetic data, the ALSPAC data set contains a battery of survey questions regarding the child's development and parent-child interactions at early ages. We observe multiple measures of children's skills and parental investments, which differ across periods. These measures are imperfect proxies of underlying skills and investments. Neglect of measurement error can lead to significant biases. In particular, differences in measurement error in the skill measures across periods will lead to wrong conclusions about the evolution of skills and the relative importance of

genes over time. We solve these problems with the latent factor model in two ways. First, we rely on the numerous measures in the ALSPAC data to identify the underlying skills and investments in each period. Second, we rely on repeated measures across periods to identify changes in skills and investments over time. These features allow us to identify the growth of skills and investments and understand how the importance of genes changes over the early life cycle.[2]

Our approach allows us to gain new insights into the process of skill formation. We find that genetic influences on initial skills are small. Instead, genetic influences accumulate over time and gradually increase over the life cycle. We document that variation in children's genetic potential for educational attainment explain only 1.5% of the variance in skills at age 0-2 but 7.4% of the variance in skills at age 6-7. This pattern is consistent with earlier findings on the increasing importance of genes over the life span (Bouchard, 2013). Unlike previous work, our approach allows us to rule out several potential sources of bias, including 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 a higher stock of skills, *the nurture of nature effect*. In addition, parental genes also become more important over time, as parents' genetic potential for educational attainment explain 0% of the variance in skills at age 0-2 but 5% of the variance in skills at ages 6-7. This happens in part because parents with higher genetic potential invest more in their children, *the nature of nurture effect*.

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., 2020; 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. Our approach allows us to simultaneously identify the influence of genes via the

[2]A related issue that we cannot directly address is measurement error in the genetic endowments. We rely on polygenic scores to measure child and parental endowment measures, but these are imperfect measures. Assuming measurement error is classical, all our estimated effects are lower bounds of true effects due to attenuation bias. We discuss this issue in more detail in Section 4.2.

nature of nurture and nurture of nature channels and show that both are important.

Our findings also have important implications for understanding the value of policies targeting inequality in skills in the presence of genetic effects. We highlight these implications using counterfactual simulation. We use our model to simulate early childhood skill formation in a scenario where parental investments are the same for all children. We show that the predicted relationship between child genetic potential for education and child skills is smaller in this counterfactual world than in the baseline scenario. An alternative interpretation - beyond the present model framework - would be that the public sector is able to construct an investment function that compensates for parental investment differences, say, by investing much more heavily in kindergartens in disadvantaged areas. This policy experiment shows that the existence of genetic effects is not at odds with the value of social policies in reducing inequality in skills. Typically, in a variance decomposition exercise, one would expect the importance of genes to increase if environmental differences were eliminated. Contrary to this, we show that the equalization of parental investments reduces the role of genes since it removes the nature of nurture and nurture of nature effects that operate via parental investments. This implies that the importance of genes depends crucially on how investments are allocated, whether by parents or society at large.[3]

The paper is organized as follows. In Section 2, we outline a conceptual model that describes the various channels through which genes influence skill accumulation. In Section 3, we discuss polygenic scores and the literature on genetic influences on skill formation. In Section 4, we introduce the ALSPAC data set, discuss our measure of genetic endowments, and conduct a preliminary descriptive analysis. In Section 5, we describe our empirical model, including the measurement system used to identify latent skills and investments, along with the estimation procedure. Section 6 presents our main results. In Section 7, we discuss the implications of our findings for our understanding of the skill formation process and for the relevance of policies designed to decrease inequality in skills. Section 8 offers brief concluding remarks, discusses limitations of our work and makes suggestions for future research.

[3]In a second policy simulation in Appendix F, we simulate the model assuming that parents - or the public sector - can fully compensate for, rather than reinforce, initial differences in genetic potential for education. We demonstrate how, within the context of our model, it is possible to eliminate all genetic influences on skills by changing how investments are allocated across children. This is a similar point as the one made by Manski (2011) when discussing the role of policy and technology in determining genetic effects on eyesight.

2 Model of Genetics and Skill Formation

In this section, 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 genetic endowments influence skill formation. The model considers a family with a single child and two parents. Thus, for simplicity, we abstract from the influence of siblings.[4] We model the evolution 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. The child's genes influence parental investments in two ways. The child's genes may directly influence parental investments as children differ in their ability to induce parents to invest more (or less) in them. In addition, depending on their preferences, parents may compensate for or reinforce disparities in initial endowments. Third, the investment decision will also depend on parental genetic endowments. This mechanism is called the nature of nurture effect. It captures genetic differences in the quantity and quality of parental interactions with their children. This is also sometimes called genetic nurture (see, e.g., Kong et al., 2018). Here we use the term nature of nurture to distinguish it from the above mentioned influence of the child's genes on parental investments (the nurture of nature).

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 2.1. In Section 2.2, we formally describe how children's skills evolve as a function of parental investments and genetic endowments. In Section 2.3, we describe how parental investments are determined. Lastly, in Section 2.4, we summarize the three main mechanisms through which genes can influence the process of skill formation.

2.1 Genetic Endowments

The child's genetic endowment is realized at conception and remains fixed throughout her life. The genetic endowment is described as a vector of individual base pairs, the fundamental

[4]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 utilise genetic information on families of four or more.

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 (Consortium et al., 2015). Most of this variation consists of single base pair changes called single nucleotide polymorphisms or SNPs for short.[5] Genetic endowments can then be described as a vector of SNPs.

Formally, let $\mathbf{g_i}$ be the genetic endowment of individual i; hence, $\mathbf{g_i}$ is a vector of nucleotide base pairs:

$$\mathbf{g_i} = \{g_{i1}, ..., g_{iS}\} \tag{1}$$

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 summarise the variation in a specific SNP using three values $\{0, 1, 2\}$. These values correspond to the number of minor (least 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.[6] Individual i has one minor allele (C) at position s, and we would say that the genetic endowment 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 genetic endowment 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 genetic endowment would have the value 0 ($g_{is} = 0$).

The child's genetic endowment is randomly determined from the parental genetic pool,

[5]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 Consortium et al. (2015) for an overview of the variation in the human genome.

[6]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.

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} is the minor allele frequency for the child's father at position s and g_{is}^{m} for the child's mother. In expectation, the child's number of minor alleles will be an average of the number of minor alleles in the parental genetic pool. 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.[7]

Most socio-economic 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 analysed 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 this reason, genetic influences on most outcomes are studied using polygenic scores that aggregate the information from thousands, if not millions, of SNPs into a single score. These scores are trait and individual specific. The polygenic score ideally captures the influence of the genome on the outcome of interest. For example, the Lee et al. (2018) GWAS may be used to construct a polygenic score for educational attainment.

This part of the model highlights some important concepts. First, genetic endowments are multidimensional, comprising of millions of individual genetic variants that vary across the human population. Second, genetic endowments are not exogenous but determined by a random draw from the parental genetic pool. The inheritance process induces correlation between the child's and her parents' genetic endowments. Third, the randomness in the inheritance process allows for identification of the effect of the child's genes that is independent

[7]The law of independent assortment implies that genetic inheritance occurs independently for each base-pair g_{is}. In practice, however, some DNA sequences are inherited together. For example, genetic markers that are physically near to each other on the same chromosome are more likely to be inherited because of genetic linkage. Similarly, assortative mating and population stratification can induce correlation in the parental genetic pool, which will break the independence of the inheritance across base-pairs. The resulting correlation in base-pairs located close to one another is called "linkage disequilibrium".

of their parents. Finally, the association between the genome and an outcome of interest can be summarized in a polygenic score. We revisit these points throughout the paper.

2.2 Technology of Skill Formation

Skill Endowments:

The child is born in period 0 with a set of initial skill endowments. For simplicity, we assume that skills may be described by a uni-dimensional measure (e.g., cognitive ability). We acknowledge that skills are multi-dimensional in nature and other (e.g., non-cognitive) skills may also matter, but prefer to focus on 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 et al. (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.

In addition, we extend the traditional model and allow both the child's genetic endowments ($\mathbf{g_i}$) and parental genetic endowments ($\mathbf{g_i^m}$ and $\mathbf{g_i^f}$) to influence the child's development in-utero: It is possible, for the same level of the mother's health behaviours, for some children to be more able to extract nutrients and other resources from their mother. Similarly, some mothers might biologically provide a better environment for fetal growth.

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

$$\theta_{i0} = f_0^{\theta}(U_i, \mathbf{g_i}, \mathbf{g_i^m}, \mathbf{g_i^f}) \tag{4}$$

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

Skill Formation:

Initial skills develop over time in response to external inputs. As in Cunha et al. (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 endowments, $\mathbf{g_i}$, and the parents' genetic endowments, $\mathbf{g_i^m}$ and $\mathbf{g_i^f}$, to enter the production function of skills.

The model allows for some children to be better at learning on their own and improving their skills. $\mathbf{g_i}$ 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 effect of parental genes, $\mathbf{g_i^m}$

and $\mathbf{g_i^f}$, 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

$$\theta_{it+1} = f_t^{\theta}(\theta_{it}, I_{it}, \mathbf{g_i}, \mathbf{g_i^m}, \mathbf{g_i^f}) \tag{5}$$

for $t = 0, 1, 2, ...T$. f_t^{θ} is a function that describes how genes interact with parental investments in determining the child's accumulation of skills.

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 demonstrated 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 endowments may be different for individuals growing up in different environments as described in Papageorge and Thom (2019) and Ronda et al. (2020).

2.3 Investment Policy Function

In the model, parents invest in their children because of altruism. 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 endowments. Child genetic endowments enter the investment policy function for two reasons. First, it captures the parents' decision to reinforce or compensate initial skill endowments. Second, it is possible for some children to be better at eliciting interactions from their parents than for others.

We also allow for parental genetic endowments to enter the investment policy function. This captures the idea that parental genes are determinants of parental skills and, as a result, determinants of the socio-emotional, intellectual and financial resources available to be invested in the child. It also captures biological differences in parents that could influence the quantity and quality of investments in the child.

The investment policy function is modelled as follows:

$$I_{it} = f_t^{I}(\theta_{it}, \mathbf{g_i}, \mathbf{g_i^m}, \mathbf{g_i^f}) \tag{6}$$

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

2.4 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 *nature of nurture effect*.[8] 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:

$$\frac{\partial f_0^\theta(\mathbf{g_i}, \overline{U_i}, \overline{\mathbf{g_i^m}}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i}} \quad \text{and} \quad \frac{\partial f_t^\theta(\mathbf{g_i}, \overline{\theta_{it}}, \overline{I_{it}}, \overline{\mathbf{g_i^m}}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i}} \tag{7}$$

where the first derivative describes the effect of 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 (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 on interaction, and instead refer to the overall

[8] In addition to the three mechanisms we focus 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 $\frac{\partial^2 f_t^\theta(\mathbf{g_i}, \overline{\theta_{it}}, I_{it}, \overline{\mathbf{g_i^m}}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i} \partial I_{it}}$ in the model. In our preferred empirical specification, we do not consider such interactions. In Appendix D, we consider this possibility and find little support in the data for these interactions being important during early skill formation.

mechanism as the *nurture of nature effect*. In the model, this effect is then captured in two ways:

$$\frac{\partial f_t^I(\mathbf{g_i}, \overline{\theta_{it}}, \overline{\mathbf{g_i^m}}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i}} \quad \text{and} \quad \frac{\partial f_t^I(\overline{\mathbf{g_i}}, \theta_{it}, \overline{\mathbf{g_i^m}}, \overline{\mathbf{g_i^f}})}{\partial \theta_{it}} \tag{8}$$

where the first derivative describes how parents investment decision depend on the child's genetic endowments and the second how it depends on the child's stock of skills.

The Nature of Nurture Effect:

Lastly, parental genes may influence the environment in which the child develops its skills. This mechanism is called the *nature of nurture effect*.[9] Since the child's genes are inherited from its parents (see equation 3), the existence of nature of nurture effects will induce an association between the child's genes and its skills. The same is true for earlier generations. That is, grandparents' genes may also affect child development, either directly (insofar as grandparents help to raise the child) or indirectly through the parents. Thus, the nature of nurture effect reflects the combined genetic nurture of previous generations captured by parental genes. We nevertheless refer to it as an 'effect' because it is an estimate of the hypothetical effect of being born into a family with different genetic endowments (holding own endowments fixed).

There are a variety of ways that parental genes enter the model. First, parental genes may influence skill formation directly via the quality of parental interactions in utero.

$$\frac{\partial U_i}{\partial \mathbf{g_i^m}} \ , \ \frac{\partial U_i}{\partial \mathbf{g_i^f}} \ , \ \frac{\partial f_0^\theta(\overline{\mathbf{g_i}}, \overline{U_i}, \mathbf{g_i^m}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i^m}} \ \text{ and } \ \frac{\partial f_0^\theta(\overline{\mathbf{g_i}}, \overline{U_i}, \overline{\mathbf{g_i^m}}, \mathbf{g_i^f})}{\partial \mathbf{g_i^f}} \tag{9}$$

In addition, parental genes may influence the quantity of investments on the child's skill formation

$$\frac{\partial f_t^I(\overline{\mathbf{g_i}}, \overline{\theta_{it}}, \mathbf{g_i^m}, \overline{\mathbf{g_i^f}}, \overline{\epsilon_{it}})}{\partial \mathbf{g_i^m}} \quad \text{and} \quad \frac{\partial f_t^I(\overline{\mathbf{g_i}}, \overline{\theta_{it}}, \overline{\mathbf{g_i^m}}, \mathbf{g_i^f}, \overline{\epsilon_{it}})}{\partial \mathbf{g_i^f}} \tag{10}$$

Lastly, parental genes may influence skill formation directly via the quality of parent-child interactions during the child's development

$$\frac{\partial f_t^\theta(\overline{\mathbf{g_i}}, \overline{\theta_{it}}, \overline{I_{it}}, \mathbf{g_i^m}, \overline{\mathbf{g_i^f}})}{\partial \mathbf{g_i^m}} \quad \text{and} \quad \frac{\partial f_t^\theta(\overline{\mathbf{g_i}}, \overline{\theta_{it}}, \overline{I_{it}}, \overline{\mathbf{g_i^m}}, \mathbf{g_i^f})}{\partial \mathbf{g_i^f}} \tag{11}$$

[9]This mechanism is also often described as genetic nurture (see, e.g., Kong et al., 2018) as well as passive gene-environment correlation (see Plomin, DeFries, and Loehlin, 1977, for an early discussion of the term).

A better understanding of the different ways genes affect skill formation, either directly, via parental investments (nurture of nature), or via parental genes (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 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 7, where we discuss the relative importance of the different mechanisms and the implications of our findings.

3 Polygenic Scores and Related Literature

The importance of genetics in explaining socio-economic outcomes has been well established in the behavioural genetics literature (Polderman et al., 2015; Plomin and von Stumm, 2018; Sacerdote, 2007; Silventoinen et al., 2020; Cesarini and Visscher, 2017; Branigan, McCallum, and Freese, 2013). The fraction of the variance of educational attainment that is explained by genes (also called heritability) has been estimated at around 40% in twin studies (Polderman et al., 2015). These estimates are consistent across a variety of kinship relationships (Cesarini and Visscher, 2017).[10] While useful, kinship studies tend to rely on a variety of strong assumptions about the familial relationship, which have been noted by several critics to be unappealing (for some early and recent criticisms, see, e.g., Taubman, 1976; Goldberger, 1979; Behrman and Taubman, 1989; Björklund, Jantti, and Solon, 2005; Manski, 2011; Durlauf, Kourtellos, and Tan, 2020).

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

Formally, a PGS for a particular outcome, w (pgs_i^w), is a linear combination of the SNP

[10] However, individual estimates of heritability also vary substantially, e.g. across countries (Branigan, McCallum, and Freese, 2013). Potentially, such differences could be explained by the nurture of nature effect that we find. Indeed, the extent to which genetic inequalities are reinforced by family investments will depend on the existence of egalitarian policies that provide more equal investments in children, as we show in Section 7.

count variables weighted by the strength of association between each SNP and the outcome of interest:

$$pgs_i^w = \sum_{s=1}^{S} \beta_j^w g_{is} \tag{12}$$

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

A GWAS follows an atheoretical approach to test the relationship between the outcome of interest and each SNP individually. In a GWAS, the outcome of interest is regressed on each SNP, one by one, along with a set of controls for population stratification. In contrast to earlier literature that relied on single genetic variants, the so-called candidate genes approach that has faced a severe replication crisis (see, e.g., Chabris et al., 2012; Charney and English, 2012; Hewitt, 2012), the GWAS approach has generated a series of robust findings. In the present paper, we use the results from the latest GWAS for educational attainment (Lee et al., 2018) to construct a polygenic score for educational attainment (EA PGS for short) for the individuals in our sample. This is our measure of genetic variation, which we describe in more detail in Section 4.2.

Polygenic scores for educational attainment have been widely used in economics and other social-science fields. We now know that these scores are highly predictive of education as well as many related outcomes, see e.g. Plomin and von Stumm (2018); Lee et al. (2018). Hence, the EA PGS predicts the accumulation of early childhood skills (Belsky et al., 2016), achievement in school (Ward et al., 2014), educational attainment (Rietveld et al., 2013; Domingue et al., 2015; Okbay et al., 2016; Lee et al., 2018; Ronda et al., 2020), as well as earnings, socioeconomic mobility, and wealth, over and above the direct effect of education (Papageorge and Thom, 2019; Belsky et al., 2018; Barth, Papageorge, and Thom, 2020).

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

It is important to note, however, that, while these scores are highly predictive of educational attainment, and are pre-determined, they are not exogenous in most models. This point has been established in two seminal papers. Using a newly developed technique for

studying heritability (relatedness disequilibrium regression – RDR), Young et al. (2018) show that neglect of parental genetic influences leads to overestimation of the importance of the child's genes.[11] Using genetic information on the child and both of its parents, Kong et al. (2018) demonstrate how an EA PGS of parents' non-transmitted genes affect their children's educational attainment; they call this 'genetic nurture'. This corresponds to the nature of nurture effect described in the previous section. They document that the size of the nature of nurture effect is about one third of the direct effect of the child's own genes on educational attainment, implying moreover that the latter is overestimated in most studies due to the confounding nature of parental genes.

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

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

[11] They estimate SNP heritability of educational attainment to be 17%, about 75% of the conventional estimates of around 22% (Rietveld et al., 2013; Okbay et al., 2016). The difference between SNP heritability and the 40 % heritability estimated in twin studies (Branigan, McCallum, and Freese, 2013; Cesarini and Visscher, 2017) is known as the missing heritability problem, see e.g. Plomin and von Stumm (2018).

The studies by Young et al. (2018), Kong et al. (2018), Ronda et al. (2020), and, to some extent, Wertz et al. (2020) highlight the importance of controlling for parental genetic influences, either directly (by incorporating the parents' EA PGS into the analyses) or indirectly (by using, e.g., sibling fixed effects designs) for identifying an effect of the child's own genes. The point is that the EA PGS, as it is calculated, captures not only the direct association between the child's own genes and the outcome of interest, but also an indirect association reflecting the correlation between children and their parents' DNA. Many studies in the field did not have this possibility (and acknowledge it) due to lack of appropriate data.

While the emerging socio-genomic literature is thus beginning to reveal certain partial associations, parental genetic influences, and to some extent interactions of (own and parental) genes with the environment, our understanding of whether and how genetic endowments interact with family resources in the process of human capital formation is still lacking in the deeper theoretical sense outlined above. E.g., do parental investments depend on the child's genetic endowments or are they primarily determined by the parents' genetic endowments? Do parental genetic endowments affect the child in other ways and, if so, what are the channels and mechanisms? The aim of the present study is to contribute with some first insights into these subjects.

4 Data and Preliminary Evidence

In this section, we introduce the ALSPAC data set 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.

4.1 ALSPAC

To investigate the relationship between genetics and the development of child skills and family investments during childhood, we need an extensive data set. For our purpose, the The Avon Longitudinal Study of Parents and Children (ALSPAC) 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 is 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).[12]

[12]The study website contains details of all the data that is 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

Questionnaires were sent to the primary caregiver (usually the mother) at regular intervals, starting before the child's birth. The caregiver responds to questions, among others, about the child's development and behaviour, 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.[13]

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 partners. The maternal and child biological samples consist of blood samples repeatedly collected as part of routine antenatal care and follow-up at clinics. Paternal samples started being collected only recently, and thus only a few of the fathers have been genotyped. Consent for biological samples was collected in accordance with the Human Tissue Act (2004).

Our main sample includes families where the child and both 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 Appendix A. The resulting sample includes 1,267 children from the original sample of 14,062 children. The small number of genotyped fathers is the main reason for the large reduction in sample size, as we have genetic information on only 1,722 fathers (in comparison to 8,804 genotyped children).[14]

4.1.1 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. In Appendix B, we provide a detailed description of the selected questions and the sections from which the questions were taken. In short, the skill measures should reflect the child's ability to perform tasks, process new information and learn abstract concepts. The investment measures reflect the family environment, either relating to physical objects or to child/parent activities. The specific measures are shown in Appendix Tables B5 and B6. In Section 5, we describe how

Committees.

[13]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.

[14]In Appendix D.2, we show that this restriction is essential. There, we re-estimate our main results in a larger sample without considering the influence of paternal genes. We show that by not including paternal genes in the estimation, we overestimate the impact of the child's genes. This is not surprising given that we end up attributing most of the paternal genetic effects to the child's genes. The robustness check highlights the importance of observing the genetic makeup of both parents.

we combine the information available on the large number of measures to identify latent skills and latent investments in a dynamic factor model.

4.2 Measure of Genetic Endowments

We measure genetic endowments using a polygenic score for educational attainment (EA PGS). As explained in Section 3, polygenic scores are a linear combination of the SNP count variables weighted by the strength of association with the outcome of interest, educational attainment in our case. The weights are derived from the most recent GWAS of educational attainment (Lee et al., 2018).[15] The PGS is constructed using the LDpred2 software (Privé, Arbel, and Vilhjálmsson, 2020). LDpred, and the newer version LDpred2, are two popular methods for deriving polygenic scores based on summary statistics. The method takes into account the matrix of correlation between genetic variants (the linkage disequilibrium matrix) to weigh properly each SNP in the construction of the PGS.

We construct the EA PGS in the same way but separately for children, their mothers and their fathers. The constructed scores have no meaningful location and scale. For ease of interpretation, we standardise all three scores in our empirical sample. For most of our analyses, we combine maternal and paternal PGS into a summary score of parental PGS. This is done for simplicity of exposition and because of the relatively small sample size. While the genes of the mother and the father may not matter equally, in Appendix C we show that this simplification does not change the preliminary results regarding the effect of the child's PGS. Furthermore, in Appendix D, we show that the structural estimates of all other parameters are virtually unaffected by separating the maternal and paternal PGS.[16]

Polygenic scores have several appealing features but some important limitations. It is appealing that we can measure the genetic potential for education at the individual level. It is also appealing that we can directly compare children's genetic potential to that of their parents. We can exploit the natural experiment created by the inheritance process to estimate the effect of the child's PGS that is independent from her parents' PGS. However, the PGS has two important limitations. First, it measures the genetic potential for educational attainment, which means that the score does not capture genetic variation unrelated to ed-

[15] We use the publicly available summary statistics at the SSGAC website, which includes the summary statistics of all meta-analysis of all discovery cohorts except 23andMe.

[16] These additional results suggest that the mother's genes may in fact be more important than the father's genes. However, as previously noted, while the child's genes are exogenous in our model, the same is not true for parental genes. Thus, while these results may reflect that – for biological or cultural reasons – the mother's genes are more important than the father's, they may also reflect that previous generations on the maternal side have a larger impact on the child's development. To dig deeper into this interesting question, we would need genetic data for three generations.

ucation. This may be a problem if genes unrelated to educational attainment also influence children's skill formation. For example, altruistic parents invest more in their children, and at the same time, the genetic propensity for altruism might not be well captured by the EA PGS. Second, polygenic scores are estimated with error. We can accurately measure the SNPs, but the accuracy of the weights (β_j^ws in eq. 12) is highly dependent on the estimation power of the original GWAS. For these two reasons, estimates of genetic effects from polygenic scores are lower bounds of the true genetic effects (see the discussion in Young, 2019). Classical measurement error in the polygenic scores is less of a problem for understanding the mechanisms through which genetic effects operate and their relative effects over time, which is our main interest in this paper. The idea is that if the measurement error in the PGS is driven solely by estimation error in the GWAS coefficients, then all estimated genetic effects are attenuated by the same amount, and we can identify their relative importance. However, even classical measurement error can lead to biased estimates in the other parameters in the model, including the roles of investments and skills. For example, we might over-estimate the importance of investments for skill formation because of the measurement error in the polygenic scores. These problems are also present when genes are completely neglected from models of skill formation.

4.3 Summary Statistics and Preliminary Evidence

We present summary statistics for the main variables used in our study in Appendix A. As explained above, we restrict our sample to families for whom genetic information is available for the trio - child, mother and father. This restriction is important for our estimation strategy. However, this approach induces positive selection in our sample of analysis, since the fathers needed to be present at home for the biological sample collection. Interestingly, we show that this positive selection is manifested in genetic differences across families. Table A4 compares the child's and the mother's PGS for families included in our study and all other families where this information is available. In general, families excluded from the analysis have a lower genetic propensity for education than the families included in the analysis. Moreover, the two samples do not differ significantly in terms of birth characteristics unrelated to the PGS, i.e. gender and birth order. The positive selection means that we should be careful in extending our findings to other groups outside our analytical subsample. However, the restriction to families with parental genetic information is important since the child's PGS is random only after we condition on both parents' genetic potential.

The child's genetic endowment 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 PGS can take vary considerably. The child's PGS can be significantly different from the parental PGS. Since there is considerable variation in how strongly each SNP is associated with educational attainment, 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 PGS 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 polygenic scores and how it relates to parental (the average of maternal and paternal) PGS. For each decile of parental PGS, we plot the density of polygenic scores of the children of those parents. The figure illustrates the randomness of genetic inheritance. Although parents with a high PGS naturally tend to have children with a high PGS, the different decile groups overlap extensively, even between children of parents in the top and bottom deciles of the genetic potential for education. 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 score can be below that of both parents.

We show the substantial variation in children's genetic potential for education even after accounting for their parents' genetic potential. However, 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 EA PGS and a crude measure of skills and investments before and after we control for the parental EA PGS. We construct the crude measures of skills and investments by averaging the standardized set of measures available in each period.

The results of these analyses are shown in Tables 1 and 2. Panel A of Table 1 demonstrates that the child's EA PGS is positively associated with her skills across childhood, starting at age 0-2 years and until age 6-7 years. Moreover, the association tends to increase over time and triple in size from age 0-2 years until age 6-7 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 2.4. Our main empirical model, described in the next section, solves both of these concerns.

In Panel A of Table 2, we relate the child's EA PGS and parental investments across the child's development. We document a strong and roughly age-invariant association between

the child's genetic endowments and investments across the entire age span from 0-7 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 and nature of nurture effects.

In Panel B of both tables, we control for parental EA PGS (the standardized average of the mother's and father's PGS).[17] Controlling for the parental EA PGS reduces the association between the child's EA PGS and skills by about 50%. We document an even larger reduction when looking at the association between the child's EA PGS and parental investments. These results are a strong indication of the presence and importance of the nature of nurture effect for children's development. We confirm this in our main empirical model that controls for measurement error.

These findings suggest a strong nature of nurture element 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 2.4. The empirical approach described in the next section addresses these concerns.

5 Empirical Model and Estimation

In this section, we present the empirical model taken to the data. We discuss the empirical challenges and the estimation procedure.

5.1 Empirical Model

The empirical model follows closely the conceptual model described in Section 2. It can be described by three main equations. First, we have the technology of skill formation that describes how skills evolve as a function of investments and child and parental genetic endowments. We describe the technology in Section 5.1.1. Second, we have the parental investment policy function that describes how investments are determined by parents as a response to their own genes and the child's characteristics, current skills and genetic endowments. The investment policy function is described in Section 5.1.2. To close the

[17] In Appendix C, we show that using both parents' EA PGS separately yields results that are very similar to those we find using their average. For simplicity, we therefore use the average parental EA PGS in all main specifications.

model, we also need to specify how initial skills are determined. Initial skills are a function of the child's characteristics, and both child and parental genes, as described in Section 5.1.3.

The empirical model covers the early stages of development from birth until age 7 years, when children enter school and parental investments become less important.[18] This is the most sensitive period for children's development (Cunha and Heckman, 2007; Cunha, Heckman, and Schennach, 2010). We include six different periods, denoted by t, each corresponding to a child's age. The periods are as follows: ages 0-2 ($t = 0$), 2-3 ($t = 1$), 3-4 ($t = 2$), 4-5 ($t = 3$), 5-6 ($t = 4$) and 6-7 ($t = 5$).

One difference compared with the conceptual model is that in the empirical specification, the child's genetic endowment is measured using an EA PGS, as described in Section 4.2. To make this distinction clear we now refer to the child's endowments as pgs_i instead of $\mathbf{g_i}$.[19] Another difference is that in the empirical specification, we consider the joint effect of the parental genes on children's development instead of the effect of each parents' genes separately. We aggregate maternal (pgs_i^m) and paternal (pgs_i^f) polygenic scores for educational attainment into a single parental genetic endowment score (pgs_i^p). We do this to simplify the model and the interpretation of the model parameters. Additional analyses, reported in Appendix C and Appendix D, show that this simplification does not significantly alter the results.[20]

5.1.1 Technology of Skill Formation

The technology of skill formation describes how skills evolve over time as a function of initial skills, parental investments and children's and parents' genetic endowments. We assume a Cobb-Douglas technology specification in the form:

$$\ln \theta_{it+1} = \ln A + \delta_1 \ln \theta_{it} + \delta_2 \ln I_{it} + \delta_3 \,\text{pgs}_i + \delta_4 \,\text{pgs}_i^p + \epsilon_{it} \tag{13}$$

where $\ln A$ is the total factor productivity (TFP) parameter and ϵ_{it} is a stochastic technology shock, which we assume is i.i.d. across individuals and time periods and is normally

[18] While it would be interesting to extend the model beyond age 7 years, unfortunately, our data does not allow us to do so. After age 7 years, the measures of children's development change substantially, and we only have information on their school performance. We leave to future research to extend our empirical findings beyond age 7 years.

[19] This distinction is important since $\mathbf{g_i}$ is a multidimensional vector of genetic variants, and pgs_i is a unidimensional score.

[20] In Appendix C and Appendix D, we also show that most of the parental genetic effects seem to be driven by the mother's genes rather than by the father's. Still, neglecting the father's genes will bias the estimates of the effect of the child's PGS (see Appendix C). This highlights the importance of controlling for the genes of both parents in the analysis.

distributed with mean zero and variance σ_ϵ^2.[21]

Genes influence skill accumulation via a variety of mechanisms. The child's genes can have a direct effect on skill accumulation, as captured by δ_3 (the direct effect in Section 2.4). The direct effect captures genetic heterogeneity in children's ability to accumulate skills, irrespective of parental investments. The child's genes might also have indirect effects on skill accumulation via early skills and investments (δ_1 and δ_2). Lastly, parents might have a direct influence on their children's skill accumulation that does not operate via investments captured by δ_4 (the nature of nurture effect in Section 2.4). The direct effect of parental genes captures 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 captures genetic heterogeneity in the quality of parenting.[22]

5.1.2 Parental Investments

Parental investments depend on the child's current stock of skills, and on both the child's and its parents' genetic endowments. The empirical specification for the investment policy function is:

$$\ln I_{it} = \gamma_1 \ln \theta_{it} + \gamma_2 \,\mathrm{pgs}_i + \gamma_3 \,\mathrm{pgs}_i^p + \gamma_x \, X_{it}^I + \eta_{it} \tag{14}$$

where η_{it} are i.i.d., mean zero, and normally distributed shocks and X_{it}^I includes individual controls that are independent of the child's and its parents' genetic endowments, such as gender and birth order dummies. In principle, we could add a variety of observable determinants of parental investments to X_{it}^I, such as maternal education and family income. These determinants, however, are partially determined by parental genes and mediate the relationship between parental genes and child outcomes. Since we are interested in the total effect of parental genes, we exclude observable variables that are dependent on parental genes from the model.[23]

[21] We assume that technology is constant across periods. While it is possible to allow all parameters to vary across periods, we do not do so in our preferred specification. In an alternative specification available upon request, the technology is allowed to vary across periods. The key patterns remain mostly unchanged, but uncertainty increases.

[22] In principle, it is possible that the four inputs interact with each other in the production of child skills. In Appendix D.3, we re-estimate the model under the assumption that the technology of skill formation follows a translog production function. The translog specification is an extension of the Cobb-Douglas that allows for interaction terms between the different inputs. The main findings do not change under this specification, but most interaction parameters in the technology are insignificant or display large confidence intervals. For these reasons, we prefer the more restrictive Cobb-Douglas form in our main specification.

[23] Note that SNPs on the gender (X and Y) chromosomes do not enter the PGS, which is therefore independent of gender.

The investment policy function above is a reduced-form approximation of the parental behaviour and follows previous work by Attanasio, Meghir, and Nix (2015), Agostinelli and Wiswall (2020), and Attanasio et al. (2020). 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 could be influenced by parents' genetic endowments. 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).

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 (γ_1), 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 behaviour not captured by current skills (e.g., enjoying being read stories), and these can also be partially determined by the child's genetics (γ_2). These two channels together capture the nurture of nature effect described in Section 2.4, 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 are genetically determined (γ_3). This is another channel capturing the nature of nurture effect described in Section 2.4.

5.1.3 Initial Skills

To close the model, we need to define how initial skills are determined. Following the conceptual model in Section 2, we allow the initial skill endowments to depend on the child's and its parents' genetic endowments. One important distinction is that in the empirical model, initial skills are realized at ages from 0 to 2 years and are thus affected by investments after birth. Unfortunately, we do not observe these early investments, so the influence of early parental investments will be captured by parental genes. The empirical specification for the initial skills function is:

$$\ln \theta_{i0} = \alpha_1 \, \text{pgs}_i + \alpha_2 \, \text{pgs}_i^p + \alpha_x \, X_i^{\theta_0} + \epsilon_{i0} \tag{15}$$

where ε_{i0} is an i.i.d., mean zero, and normally distributed shock to early skills, and $X_i^{\theta_0}$ includes individual controls that are independent of the child's and its parents' genetic endowments, such as gender and birth order dummies. The α_1 and α_2 parameters capture the direct effect of the child's and its parents' genetic endowments for the child's initial stock of skills; they capture direct genetic effects during development in utero and the first year of life. These effects correspond to the direct effect and the nature of nurture effect on initial skills.

5.1.4 Measurement System

We observe multiple measures of children's skills and parental investments in each period. These measures are imperfect proxies of underlying skills and investments. Using individual proxies or a simple average, as in Section 4.3, can lead to significant biases. In this section, we describe how the multiple measures may be used to identify the underlying skills and investments in each period. We formalise 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_{ijt}^{θ} denote the jth measurement of child i's skill at period t, and let m_{ikt}^{I} 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}:

$$m_{ijt}^{\theta} = \mu_{jt}^{\theta} + \lambda_{jt}^{\theta} \cdot \ln \theta_{it} + \nu_{ijt}^{\theta} \tag{16}$$

$$m_{ikt}^{I} = \mu_{kt}^{I} + \lambda_{kt}^{I} \cdot \ln I_{it} + \nu_{ikt}^{I} \tag{17}$$

where λ_{jt}^{θ} and λ_{kt}^{I} are the factor loading for skill measurement j and investment measurement k, ν_{ijt}^{θ} and ν_{ikt}^{I} are i.i.d. measurement errors.[24] As in Agostinelli and Wiswall (2020), we make no further assumptions on the distribution of the measurement errors.

[24]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.

5.1.5 Identifying Assumptions

The key identifying assumption is that all shocks and measurement errors are independent from each other and across time. Formally, we array the skill formation shocks ϵ_t in a vector ϵ, the investment shocks η_t in a vector η, and assume that,

$$\epsilon_t \perp\!\!\!\perp \epsilon_{t'} \ \forall t \neq t', \tag{18}$$

$$\eta_t \perp\!\!\!\perp \eta_{t'} \ \forall t \neq t', \tag{19}$$

$$\epsilon \perp\!\!\!\perp \eta. \tag{20}$$

Assumptions (18) and (19) maintain independence of the shocks over time, and (20) maintains the independence between shocks to investment and skills.

In addition, we array the measurement errors for skills ν^{θ}_{jt} in a vector ν^{θ}, and the measurement errors for investments ν^{I}_{kt} in a vector $\nu^{\mathbf{I}}$ and assume that,

$$\nu^{\theta}_{jt} \perp\!\!\!\perp \nu^{\theta}_{j't} \ \forall j \neq j', \tag{21}$$

$$\nu^{\theta}_{jt} \perp\!\!\!\perp \nu^{\theta}_{jt'} \ \forall t \neq t', \tag{22}$$

$$\nu^{I}_{kt} \perp\!\!\!\perp \nu^{I}_{k't} \ \forall k \neq k', \tag{23}$$

$$\nu^{I}_{kt} \perp\!\!\!\perp \nu^{I}_{kt'} \ \forall t \neq t', \tag{24}$$

$$(\epsilon, \eta) \perp\!\!\!\perp (\nu^{\theta}\nu^{\mathbf{I}}), \tag{25}$$

$$\nu^{\theta} \perp\!\!\!\perp \nu^{\mathbf{I}}. \tag{26}$$

Assumptions (21) to (26) maintain that the measurement errors are independent from each other, independent across time and independent from the shocks.

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 normalising 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 B6 in Appendix B describes all the measures we use to identify the latent investments at the

different periods. We have three measures that are asked at all periods: "Frequency the mother reads to the child", "Frequency the child goes to a library", and the "Frequency the child goes to places of interest". 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 the other two measures. We make the following normalising assumption on the age-invariant measure of investments:

$$m^I_{i1t} = 0 + 1 \cdot \ln I_{it} + \nu^I_{i1t} \quad \text{for} \quad t \in \{0, 1, 2, 3, 4\} \tag{27}$$

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 B5 in Appendix B). 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 normalising assumption on the two measures:

$$m^\theta_{i1t} = 0 + 1 \cdot \ln \theta_{it} + \nu^\theta_{i1t} \quad \text{for} \quad t \in \{0, 1, 2\} \tag{28}$$

$$m^\theta_{i2t} = \mu_{21} + \lambda_{21} \cdot \ln \theta_{it} + \nu^\theta_{i2t} \quad \text{for} \quad t \in \{1, 2, 3, 4, 5\} \tag{29}$$

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

5.2 Estimation and Inference

The model is estimated in a series of steps. The estimation procedure is a slightly modified version of the algorithm proposed in Agostinelli and Wiswall (2020), and has several appealing properties. First, the estimation procedure does not impose any assumptions on the distribution of the latent variables. Also, it does not impose any additional assumptions on the distribution of the measurement errors. The algorithm is computationally simple and

does not require simulating the full model. Lastly, it has the advantage that the model can be estimated in a sequence of steps, helping with estimation and computational tractability. We describe the estimation procedure in detail in Appendix E.

We rely on a bootstrap procedure for inference. We re-sample the individuals from our initial sample at random with replacement and re-do 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 90% confidence intervals that are reported in the paper. The procedure takes into account the estimation error at all the steps.

6 Estimation Results

Here we present the key empirical findings from the estimation model described in the previous section. We first discuss the estimates for the initial skills function (Section 6.1). This is followed by the estimates for the technology of skill formation (Section 6.2). Next, we discuss how parental investments are determined (Section 6.3). 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 how the growth of skills is different across the genetic spectrum (Section 6.4). We then use the simulations to decompose the variation of simulated skills at different ages explained by the various model components: the child's genes, parental genes, observable child characteristics and shocks to investments and skill formation (Section 6.5).

6.1 Genetic Influences on Initial Skills

We document small but significant effects of the child's genes on its initial skills and a negligible influence of parental genes. This means that the impact on initial skills only explains a small part of overall direct effects. Also, the nature of nurture effect does not operate via initial skills. These effects are documented in the first column of Table 3, which presents parameter estimates from the empirical specification in equation 15. Note that these initial skills are realized at ages 0-2 years, and they are therefore affected by conditions in utero as well as by very early investments after birth. We find no effect of parental genes and only a small effect of the child's own genes on initial skills.[25] A one-standard-deviation

[25]In Appendix D.1, we distinguish the influence of maternal and paternal genes. There, we document a positive effect of maternal genes and a negative effect of paternal genes on early skills. The positive effect of maternal genes is intuitive, given the possibility of maternal genes influencing in-utero effects. However, it is unclear why paternal genes should harm the child's in-utero development.

increase in the child' EA PGS increases initial child skills by 2.2%.

6.2 Genetic Influences on the Technology of Skill Formation

The estimated parameters of the technology of skill formation (equation 13) are presented in the second column of Table 3. 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 estimate is 0.469 with 90% confidence intervals between 0.419 and 0.538. Its order of magnitude is similar to that found by Cunha, Heckman, and Schennach (2010) for their first stage (until age 5-6 years) of development. Both Cunha and Heckman (2008) and Agostinelli and Wiswall (2020) find considerably larger parameter estimates, close to unity, but their models begin at the age where ours end, so their estimates are not directly comparable to ours. The parameter on investments is considerably smaller than the self-productivity parameter but once again in line with the estimates for the first stage of childhood in Cunha, Heckman, and Schennach (2010). It is larger than that found by Cunha and Heckman (2008) but in line with that found by Agostinelli and Wiswall (2020), keeping in mind that those two latter studies are focused on children aged 5-6 years onwards.

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 effects of the child's own EA PGS and the parents' EA PGS are similar in magnitude, and both are statistically significant. The effect of the child's EA PGS captures the ability to acquire new skills, given the current stock of skills and parental investments (the direct effect in Section 2.4). It is interesting that, even after allowing parental genes to affect investments, parental genes still have a non-negligible influence on skill formation that is not captured by investments. This reflects the nature of nurture effect through the quality of parenting (as opposed to investments, which capture more quantitative elements of the nature of nurture effect).[26] The parameters suggest that, everything else equal, a one-standard-deviation increase in the child's EA PGS increases the current stock of skills by 1.6%, and a similar increase in the parental score increases skills by 2%. These effects accumulate over time. In Sections 6.4 and 6.5, we provide more insights into the importance of genes by analysing the importance of the cumulative genetic effects in explaining skill formation over time as well as variation in skill formation across individuals.

[26]In Appendix D.3, we re-estimate the model under the assumption that the technology of skill formation follows a translog production function, which allows for the interaction between the different inputs. There, we show that parental genes seem to be a substitute for parental investments. While statistically significant, the confidence intervals on the interaction are very large for any meaningful conclusion to be made. For this reason, and since our main conclusions do not change under the translog specification, we decided to go with the Cobb-Douglas as our main specification.

6.3 Genetic Influences on Parental Investments

The estimated parameters for the investment policy function (equation 14) are presented in the third column of Table 3. 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. While not statistically significant, we find a positive effect of children's genes on parental investments once we condition on parental genes and the child's current stock of skills. We interpret this as suggestive evidence that children with different genetic makeup can elicit different responses from their parents even after conditioning on the effect of genes on skills.

However, we also find that the child's genes indirectly influence parental investment through the stock of skills. We find that parents reinforce initial skill differences and invest more in high-skilled children. This 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 propensity for education will tend to have a higher stock of skills, leading parents to invest more in their skill development. This makes sense; parents do not observe their child's genes directly; hence, only to the extent that they manifest in the stock of skills can they infer something about their genetic propensity for education. We will return to the implications of this important result in Section 7.

We also document a large effect of parental genes on parental investments. This corresponds to the nature of nurture effect, where parental genes influence the quantity of investments in their children. Parents with a higher genetic propensity for education (EA PGS) invest more in their children conditional on the children's current stock of skills. This parameter is much larger than that of the child's own genes in the investment equation and is also highly statistically significant. The child and parental genetic effects on parental investments are also similar in direction and significance to those found by Wertz et al. (2020).[27]

6.4 Genetic Influences Across Early Ages

To make better sense of the magnitude and importance of the genetic effects, we graphically depict the relationship between the child's propensity for education 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 standardise the stock of

[27] In Appendix D.1, we distinguish the influence of maternal and paternal genes. There, we document a larger influence of maternal genes on parental investments than that of paternal genes.

skills at each period for ease of comparison. The results are shown in Figure 3. Figure 3(a) shows the association between the child's own EA PGS in percentiles and the standardized stock of skills at different ages.[28] Two important results are evident; first, the effects are relatively large; a child at the 80^{th} percentile in the distribution of polygenic scores has a 0.6 standard deviation higher stock of skills at age 6-7 years than a child at the 20^{th} percentile of the PGS distribution. Second, the effect of genes accumulate over time. At age 0-2 years, the difference in the stock of skills between the same percentiles is less than 0.2 standard deviations. This is consistent with Belsky et al. (2016), who find that the association between EA PGS and cognitive ability increases from age 3 to 11, and with earlier twin and adoption studies showing an increasing heritability of IQ with age (Bouchard, 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.

It is not only child genes which influence skill formation; parental genes are also associated with skill formation. Figure 3(b) shows the relation between the parental EA PGS and the child's stock of skills by age. Once again, the associations are relatively large. Thus, a child whose parents genetically endowed with a PGS at the 80^{th} percentile has about a 0.6 standard deviation higher stock of skills at age 6-7 years than a child whose parents are at the 20^{th} percentile of the PGS distribution. As was the case for the child's own genes, we observe an increasing importance of parental genes by age. At age 0-2 years, the difference from the 80^{th} to the 20^{th} percentile in parental polygenic scores correspond only to a 0.2 standard deviation increase in the child's stock of skills.

6.5 Variance Decomposition

Lastly, we attempt to further illustrate the implications of our results using a variance decomposition exercise that is common in the genetics literature. To do so, we compute the variance of child skills at various ages that can be attributed to the child's genetic score, to the parental genetic score, to the two observable variables (gender and birth order), and to exogenous shocks. We report the results from this variance decomposition exercise in Table 4.

The proportion of variation in an outcome (in our case, the stock of skills) that can be attributed to genetic factors is usually referred to as the heritability of the outcome (Visscher, Hill, and Wray, 2008). It is important to keep a few facts in mind when comparing our results

[28] We call these associations instead of effects since we do not impose independence between children and parental genes when doing the simulations to construct these figures.

to other heritability estimates in the literature. Our estimates capture the variation in skills that can be attributed to the child's and its parents' EA PGS. This is usually referred to in the literature as the PGS heritability of an outcome. The PGS heritability is often much smaller than the total variation in the outcome that can be attributed to the child's or its parents' genes, often referred to as the broad-sense heritability, (see, e.g., Plomin and von Stumm, 2018).

There are three reasons for that. First, the PGS heritability captures only the genetic variation captured by the set of SNPs included in the PGS. It does not capture genetic variation in rare variants and structural variants that might be important for some outcomes.[29] Second, the polygenic score only captures the additive variation in the SNPs since it is a linear combination of the SNP counts (see equation 12). Lastly, the GWAS weights used to compute the PGS ($\{\beta_j^w\}$) are measured with error, and this estimation error leads to measurement error in the PGS. This measurement error will lead to attenuation bias in both the PGS's estimated effects and in the heritability estimates. However, we should note that, as long as the measurement error is classical, we should expect all estimated genetic effects to be attenuated by the same amount. So while measurement error in the polygenic scores is problematic for understanding the magnitude of genetic effects and the proportion of the variance of an outcome that can be explained by genes, it might not be a problem for understanding the mechanisms through which genetic effects operate and their relative effects over time, which is our main interest in this paper. With that in mind, we proceed with the variance decomposition analysis.

We decompose the variance in skills at different ages into four components; variance attributable to child genetics, parental genetics, observable factors unrelated to the genetic potential for education (gender and birth order), and the unobserved shocks. Most such decompositions conflate the effects of child and parental genes, as demonstrated by, e.g., Young et al. (2018), such that the estimate of SNP heritability is upward biased in such studies. This is taken into account in the decomposition conducted here by decomposing the variance into separate contributions from child and parent genes. To do so, we simulate skills and investments using the estimated model parameters under different conditions and compare the variance in skills across models. We simulate the model under incremental changes, first by setting children's PGS to its mean level for all children, then by setting parental PGS to its mean level, and lastly, by setting children's observable characteristics to their mean level.

[29]Rare variants correspond to genetic variation in a single nucleotide that is rare in the population and thus not captured by SNPs. Structural variants correspond to genetic variation that affects multiple bases. See Consortium et al. (2015) for an overview of the variation in the human genome.

The variance decomposition at different ages is shown in Table 4. The table reveals several important results. First, the relative importance of genes, both child and parent genes, increases over time. Heritability increases from 1.55% to 7.4% going from age 0-2 years to age 6-7 years. Similarly, the fraction of variance explained by parental genes increases from 0% to 5.0%. The relative importance of parents' versus child's genes that we find is completely in line with studies by Young et al. (2018) and Kong et al. (2018), where the former find a reduction of heritability of around 23% (SNP heritability reduced from 22% to 17%) and the latter find that about a third of the heritability can be attributed to the nature of nurture effect.

We can perform a similar decomposition of the variance in parental investment; we call this investment heritability. Results are shown in Table 5. First, note that this decomposition is much more stable over time. The fraction of variance explained by the child's PGS increases from 4.1% to 6.1%, while that of the parents increases from 4.8% to 7.9%. Note also the relatively larger importance of parental genes in this decomposition, where the relative contribution of the child's and the parents' genes is reversed as compared to the skill decomposition, which reflects an important element of the nature of nurture effect operating through the investment channel.

7 Policy Simulations and Implications

In this section, we perform a couple of counterfactual simulations using the estimation model. First, to illustrate the importance of the nature of nurture and nurture of nature mechanisms, we simulate what happens to the link between child genes, investments and the distribution of skills when we shut down either or both of these channels. Second, we perform simulations that are meant to mimic real-world policy changes that help equalize investments across children, and argue that our findings also have implications for the allocation of public investments.

7.1 Mechanism Decomposition

We perform the first simulations to illustrate the relative contribution of the nature of nurture and nurture of nature effects. We assess how the relationship 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 high genetic potential (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., parental investments are independent of child endowments.

Our main findings also show a significant influence of parental genes on skill formation (the nature of nurture effect). 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 nature of nurture, i.e., where parental genes only matter by being transmitted to the child.

Figure 4(a) plots the standardized simulated latent investments at age 5-6 years under the different specifications as a function of the child's EA PGS in percentiles. The blue line corresponds to the predicted relationship between the child's genetic endowments and latent investments from our empirical model. It shows that children with a higher genetic endowment receive higher levels of parental investment. The predicted effects are large; a change from the 20th to the 80th percentile in the genetic score predicts about a 0.6 standard deviation increase in parental investments. This association reflects both nature of nurture and nurture of nature effects. In comparison, the yellow line shows the relationship between the child's EA PGS and investments when we close the nature of nurture channel. When we thus eliminate the effect of parents' genes, the association between child EA PGS and investments goes down by around 53 percent, as shown in Panel B of Table 6. The remaining association is due to the nurture of nature channel. Conversely, the red line illustrates that the counterfactual that eliminates this mechanism reduces the association between child EA PGS and investments by around 47 percent. This shows that both mechanisms contribute with a comparable magnitude to the observed inequality in investments. Finally, eliminating both channels simultaneously is equivalent to completely eliminating the link between child EA PGS and investments, as illustrated by the purple line.

Figure 4(b) presents the corresponding relationship between the child's genetic score and standardized latent skills at ages 6-7 years under the different scenarios. Again, the blue line depicts the predicted relationship from our empirical model between the child's genetic endowments and latent skills, and we see that the predicted skill gradient is substantial. Next, we illustrate to what extent the inequality in skills is caused by the inequality in investments. The yellow illustrates the relationship between the child's EA PGS and skills in the counterfactual world with no nature of nurture. Thus eliminating the effect of parental genes on investments and skills reduces the association between child EA PGS and skills by

49 percent, as shown in Panel A of Table 6. Conversely, the red line depicts the predicted relationship between child EA PGS and skills when we eliminate the nurture of nature effect. Doing so, we see that this reinforcing parental behavior accounts for 11 percent of the total association between child EA PGS and skills. Finally, even after eliminating both the nature of nurture and nurture of nature channels, the purple line shows that about 40 percent of the association between child EA PGS and skills remains; this is the direct effect of genes on skill formation.

This exercise serves two purposes. First, it gives a sense of the relative importance of the different channels of genetic influence. We see that the nature of nurture and the nurture of nature contribute about equally to the gradient between child genes and parental investments. On the other hand, the nature of nurture effect is relatively more important for the distribution of skills. The reason is that this mechanism is twofold: higher parental EA PGS makes parents invest more in the children, but it also increases child skill accumulation conditional on investments, showing that the child benefits more from interacting with its parents if the parents have a high EA PGS. 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. This implies that, in opposition to the commonly held view, there is no trade-off between the importance of genes and family environment. In the next section, we argue that the same is true if we broaden the concept of investments to include public policies. Thus, genetic effects do not entail a lesser role for public investments in shaping the distribution of skills.

7.2 Equal Investments Policy Simulation

While our focus has been on parental investments, one can view the allocation of public investments through the same framework. This distinction is not sharp in practice: while some children are home-schooled, others grow up in institutions. And like parental investments, public policies can affect the distribution of children's skills. This is sometimes even the purpose of such policies. Specific examples include Head Start, the Abecedarian Project, and various home-visiting programs, all aimed at levelling the investments that children receive at a high level.

In this next counterfactual experiment, we simulate the formation of skills, forcing parental investments to be at the 95^{th} percentile for all children. This equalization could come about in different ways, but is meant to mimic real-life policies that aim at providing high-quality investments for all children (whether directly or through an increase in parental

investments). The purpose of this experiment is to demonstrate the value of social polices aimed at reducing disparity in skills, even when genetics play an important role in skill formation.

Figure 5(a) plots the standardized simulated latent investments at ages 5-6 years under the two specifications as a function of the child's EA PGS in percentiles. As before, the blue line corresponds to the predicted relationship between the child's genetic endowments and latent investments from our empirical model, while the red line now corresponds to the counterfactual where investments are equalized at the 95^{th} percentile level.[30]

Figure 5(b) presents the corresponding relationship between the child's genetic score and standardized latent skills at ages 6-7 years. As before, the 'Baseline' scenario is depicted via the blue line and the 'Equal Investments' counterfactual by the red line. We see that equalizing investments at a high level also makes the distribution of skills much more equal. This illustrates how the genetic influences on skill formation are attenuated when investments are equalized across children. However, a significant association remains between the child's EA PGS and skills. There are two reasons for this: The child's own genetic endowments has a direct effect on skill formation, and, at the same time, parental endowments also influence skill formation conditional on the level of investments. In other words, in the 'Equal Investments' counterfactual, part of the nature of nurture effect is still present, as only the part that works through investments is eliminated. Thus, this corresponds to a world where all children receive the same quantity of investments, e.g., all families are provided with books and the parents are effectively encouraged to read to the children. Yet, parental endowments still play a role because they also affect how much the child learns from interacting with the parent in this way (the quality component of the nature of nurture effect).

In a sense, this counterfactual is thus a more realistic illustration of the role that public policy can and sometimes aim to play. The simulations show that policies aimed at levelling the playing field and reducing inequalities in child investments will not only reduce disparities in skills but also decrease the importance of genes for skill formation. And while this will

[30]Because investments are latent, we cannot directly translate what this high level investments corresponds to. However, we can get a sense of it by looking at our crude measure of investments from the reduced form analysis. There, the individual at the 95^{th} percentile of investments is someone for whom, at age 0-2, the parents 'almost daily' read, sing and play with. This child also owns 10+ books, and the mother often shows the child picture books (the partner hardly ever). The mother also teaches songs and shapes and sizes, and the child is taken to the park, a library, and other places of interest once per week. Similarly, at age 5-6, the individual at the 95^{th} percentile of investments is someone to whom the parents read and sing often. This child plays a musical instrument daily, goes to special classes/clubs once a week, goes swimming and goes to places of interest once a month, and visits the library a few times a year. Many of these types of investments would presumably also be available in high-quality daycare institutions.

naturally be a consequence of policies aimed specifically at more disadvantaged children, public institutions may play a similar role. For example, the provision of universal high quality child care may be equivalent to a partial equalization of parental investments. And some public school systems even have a reduction of the influence of social background on academic achievement as a stated goal. Our simulations show that genetic effects do not obstruct the pursuit of such a goal. Rather, reducing the influence of social background and reducing the influence of genetic background are complementary goals.

To reinforce this point, we re-do the variance decomposition exercise discussed in Section 6.5 and presented in Table 4 under the 'Equal Investments' counterfactual. The new heritability estimates are presented in Table 7. We show that heritability decreases in the 'Equal Investments' scenario compared with the 'Baseline' scenario. The variation in latent skills at ages 6-7 years explained by child and parental genes decrease by 38% and 54%, respectively. This is a direct consequence of the mechanisms illustrated by the previous simulations, namely that the interaction between genes and investments entails that making the distribution of investments more equal also decreases genetic influences on skill formation.

Is it possible to level the playing field completely, such that skills at the age of school entry (6-7 years) are equalized across all children? In Appendix F, we perform a second policy counterfactual to assess how investments should be reallocated to achieve such an outcome. The counterfactual mimics a world where parents engage in a compensatory instead of a reinforcing behavior (a reversal of the nurture of nature effect), or a world where such compensating investments are publicly provided. Figure F10(a) shows that this is indeed possible, although it would require more than a reversal of the estimated relationship between child EA PGS and investments (because it also has to compensate for the direct effect of genes). While such extreme compensating behavior may be unrealistic, this simulation reemphasizes the point that the relative importance of genes in a society is tightly related to how investments and resources are allocated across children. Thus, the existence of genetic effects is not at odds with the value of social policies in reducing inequality. The existence of genetic effects makes these policies even more relevant.

8 Discussion and Conclusion

To better understand the interplay between genetics and family resources for skill formation and its relevance for policy, we incorporated genetic endowments measured by an EA PGS into a dynamic model of skill formation. We modelled and estimated the joint evolution of skills and parental investments throughout early childhood (ages 0 to 7 years). We observed both child and parental genetic endowments, allowing us to estimate the independent effect

of the child's genes on skill formation and to estimate the importance of parental genes for 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.

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 parental genes matter, as parents with different genetic makeups invest differently in their children (the nature of nurture effect). 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 highlight the critical implications of our findings using two simulation counterfactuals. In one counterfactual, we illustrate the relative importance of the three channels of genetic influence by simulating the distribution of investments and skills while closing down the nature of nurture and/or the nurture of nature channel. This shows that both of these interactions contribute significantly to the inequality in investments, while the nature of nurture dominates with respect to the inequality in skills. More generally, this illustrates that, in contrast to what is often told, there is no real trade-off between the importance of genes and family environment. In another counterfactual, which is meant to mimic real-life policies that equalize investments at a high level for all children, we show that the association between genetics and skills is smaller in a world where investments are equalized across families. This finding shows that the existence of genetic effects is not at odds with the value of social policies in reducing inequality. The presence of genetic effects makes these policies even more relevant since genetic inequality increases inequality in parental investments.

A limitation of our work is that genetic endowments are measured using polygenic scores. It is possible for genes unrelated to educational attainment also to influence children's 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 a critical genetic component by using a PGS for educational

attainment. Another limitation of using polygenic scores is measurement error. Since polygenic scores are measured with error, our estimates are lower bounds of the true genetic effects. An interesting extension of our work would be to use different methods, such as genome-based restricted maximum likelihood (GREML), to sidestep the measurement problem and document whether different genetic variants are related to the various mechanisms we outline in Section 2.

Lastly, it is important to recognise that we only include individuals of European ancestry in our analysis. This opens the question whether our findings would extend to other ancestry groups. Unfortunately, this is not something we can test. This is a major issue in the literature since the majority of polygenic scores are constructed from GWAS' performed in Europeans, and their transferability to other populations is dependent 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 PGS' 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

Agostinelli, Francesco and Matthew Wiswall. 2020. "Estimating the technology of children's skill formation." *National Bureau of Economic Research, Working Paper* .

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.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2015. "Human capital development and parental investment in india." *National Bureau of Economic Research, Working Paper* .

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–E9772.

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, Part 2):S143–S162.

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

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

Belsky, Daniel W, Benjamin W Domingue, Robbee Wedow, Louise Arseneault, Jason D Boardman, Avshalom Caspi, Dalton Conley, Jason M Fletcher, Jeremy Freese, Pamela Herd et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31):E7275–E7284.

Belsky, Daniel W, Terrie E Moffitt, David L Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S Williams et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-course Development." *Psychological Science* 27 (7):957–972.

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

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):923–930.

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):111–127. URL `https://doi.org/10.1093/ije/dys064`.

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

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.

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

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

Conley, Dalton and Jason Fletcher. 2017. *The Genome Factor: What the Social Genomics Revolution Reveals About Ourselves, Our History, and the Future*. Princeton University Press.

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

Cunha, Flávio, Irma Elo, and Jennifer Culhane. 2013. "Eliciting maternal expectations about the technology of cognitive skill formation." *National Bureau of Economic Research, Working Paper* .

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

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

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

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

Domingue, Benjamin W, Daniel W Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D Boardman. 2015. "Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health." *AERA Open* 1 (3):2332858415599972.

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

Fraser, Abigail, Corrie Macdonald-Wallis, Kate Tilling, Andy Boyd, Jean Golding, George Davey Smith, John Henderson, John Macleod, Lynn Molloy, Andy Ness, Susan Ring, Scott M Nelson, and Debbie A Lawlor. 2012. "Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort." *International Journal of Epidemiology* 42 (1):97–110. URL `https://doi.org/10.1093/ije/dys066`.

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

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.

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

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, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374):424–428.

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

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

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

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):584–591.

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:e48376.

Nicoletti, Cheti and Valentina Tonei. 2020. "Do parental time investments react to changes in child's skills and health?" *European Economic Review* :103491.

Okbay, Aysu, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens et al. 2016. "Genome-wide Association study identifies 74 loci associated with educational attainment." *Nature* 533 (7604):539.

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

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

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

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

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

Rietveld, Cornelius A, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal et al. 2013. "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment." *Science* :1235488.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm. 2020. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." *IZA Discussion Paper 13441* .

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

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

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

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

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

Visscher, Peter M, William G Hill, and Naomi R Wray. 2008. "Heritability in the genomics era—concepts and misconceptions." *Nature Reviews Genetics* 9 (4):255–266.

Ward, Mary E, George McMahon, Beate St Pourcain, David M Evans, Cornelius A Rietveld, Daniel J Benjamin, Philipp D Koellinger, David Cesarini, George Davey Smith, Nicholas J Timpson et al. 2014. "Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children." *PloS One* 9 (7):e100248.

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

Young, Alexander I. 2019. "Solving the missing heritability problem." *PLoS Genetics* 15 (6):e1008222.

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.

9 Tables and Figures

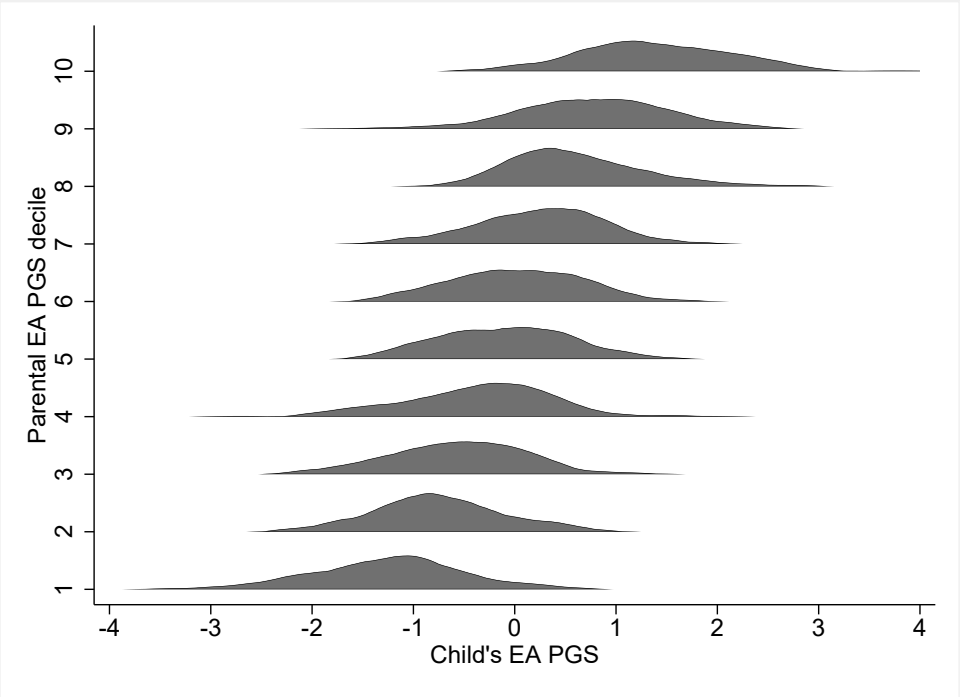


Figure 1: Distribution of Children's EA PGS by Parental EA PGS Decile: This figure plots the density of the standardized EA PGS of the child; separately for each decile of the parental EA PGS. The figure highlights the dependence between the child's and its parents' genetic potential for education. It also highlights the variation in the child's potential for education even after conditioning in the parents' potential. There is an overlap in the distribution of genetic potential for education across all parental PGS deciles.

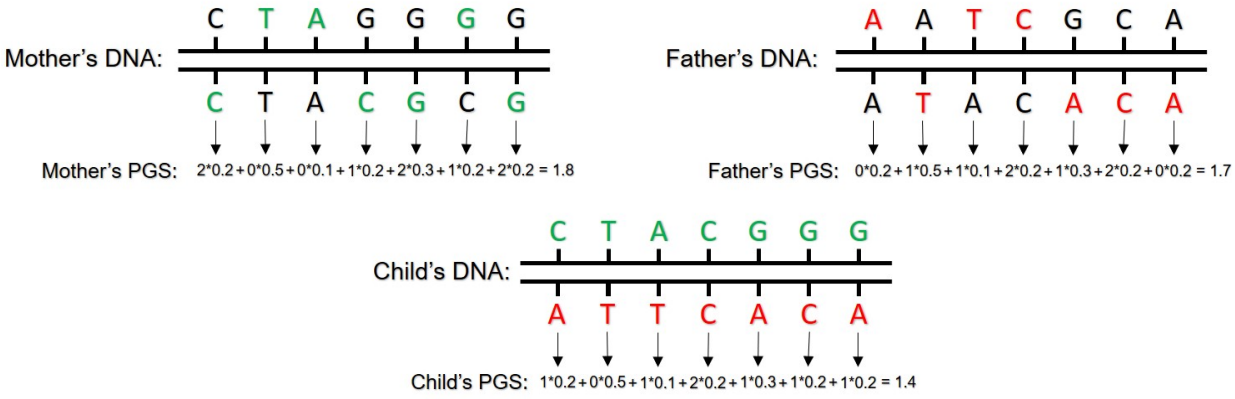


Figure 2: Family Genetic Data: Using a stylized example, this figure depicts how maternal and paternal DNA is transmitted to the child and how the child's polygenic score might differ from that of its parents. The child's DNA is composed of variants inherited at random from the mother (in green) and 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 score.

Table 1: EA PGS and Skills by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.047*	0.047*	0.097***	0.158***	0.169***	0.101***	0.103***
	(0.029)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	(0.021)
R_2	0.002	0.003	0.009	0.027	0.028	0.010	0.012
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS	0.045	0.009	0.024	0.076*	0.099**	0.039	0.049
	(0.044)	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)	(0.032)
Parental PGS	0.003	0.051	0.097**	0.108**	0.092**	0.082*	0.072**
	(0.043)	(0.042)	(0.042)	(0.042)	(0.043)	(0.043)	(0.031)
N	1267	1267	1267	1267	1267	1267	7602

Notes: This table reports parameter estimates from regressions used to link the polygenic score for educational attainment to children's skills across childhood. To test the effect of the EA PGS, we regress at each age the skill measure on the polygenic score, controlling for gender and the first 15 principal components of the genetic matrix. In Panel B, we add the parental polygenic score 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 score. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Table 2: EA PGS and Investments by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.110***	0.180***	0.140***	0.117***	0.146***	0.194***	0.148***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.021)
R_2	0.015	0.031	0.020	0.014	0.022	0.035	0.023
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS	0.034	0.052	0.035	-0.059	-0.008	0.073*	0.021
	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.033)
Parental PGS	0.101**	0.169***	0.139***	0.231***	0.204***	0.160***	0.167***
	(0.043)	(0.043)	(0.043)	(0.042)	(0.042)	(0.043)	(0.034)
N	1267	1267	1267	1267	1267	1267	7602

Notes: This table reports parameter estimates from regressions used to link the polygenic score for educational attainment to family investments across childhood. To test the effect of the EA PGS, we regress at each age the investments measure on the polygenic score, controlling for gender and the first 15 principal components of the genetic matrix. In Panel B, we add the parental polygenic score to the regressions. The investments 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 score. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Table 3: Main Parameter Estimates

	$\ln \theta_{i0}$	$\ln \theta_{it+1}$	$\ln I_{it}$
pgs_i	0.022	0.016	0.013
	[0.002 , 0.039]	[0.005 , 0.032]	[-0.001 , 0.027]
pgs_i^p	-0.001	0.020	0.041
	[-0.018 , 0.018]	[0.010 , 0.037]	[0.023 , 0.056]
$\ln \theta_{it}$	.	0.469	0.265
	.	[0.419 , 0.538]	[0.180 , 0.303]
$\ln I_{it}$	.	0.205	.
	.	[0.120 , 0.293]	.
Constant	1.463	1.151	3.076
	[1.434 , 1.494]	[0.672 , 1.567]	[2.985 , 3.295]

Notes: The parameter estimates for the initial skill equation (Equation 15) are reported in the first column, for the technology of skill formation (Equation 13) in the second column, and for the investment policy function (Equation 14) in the third column. 90% bootstrap confidence intervals in brackets.

Table 4: Skill Heritability by Age

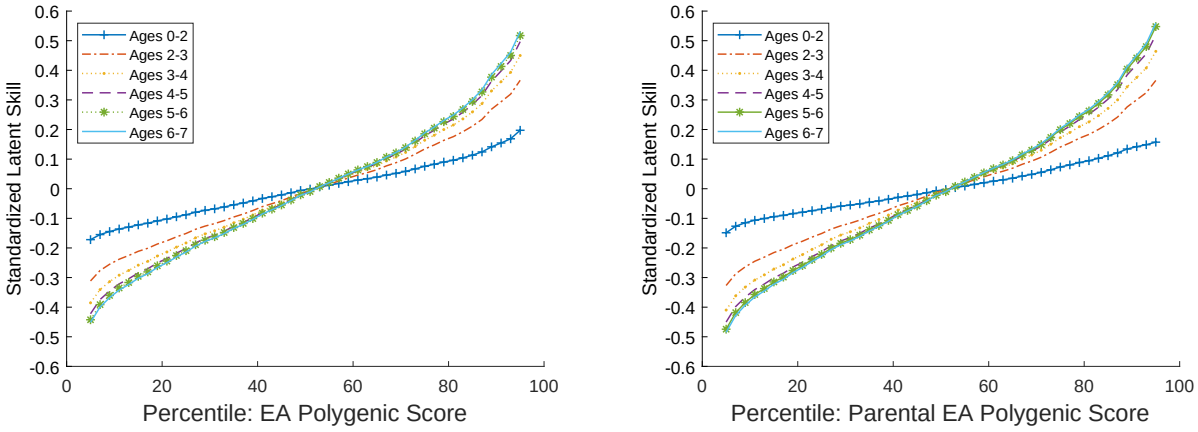
Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
pgs_i	1.55%	4.07%	5.65%	6.63%	7.11%	7.39%
pgs_i^p	-0.00%	1.49%	3.08%	4.15%	4.70%	4.96%
Xs	6.81%	1.79%	0.99%	0.71%	0.58%	0.51%
ϵ and η	91.64%	92.66%	90.29%	88.51%	87.61%	87.14%

Notes: This table presents the proportion of the variance of latent skills at different periods that is explained by the model's four main components: i) the child's polygenic score for educational attainment, ii) the parental polygenic score for educational attainment, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table 5: Investment Heritability by Age

Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6
pgs_i	4.06%	5.01%	5.61%	5.87%	6.06%
pgs_i^p	4.85%	6.30%	7.21%	7.61%	7.95%
Xs	6.97%	5.79%	5.26%	5.07%	4.93%
ϵ and η	84.12%	82.90%	81.92%	81.45%	81.06%

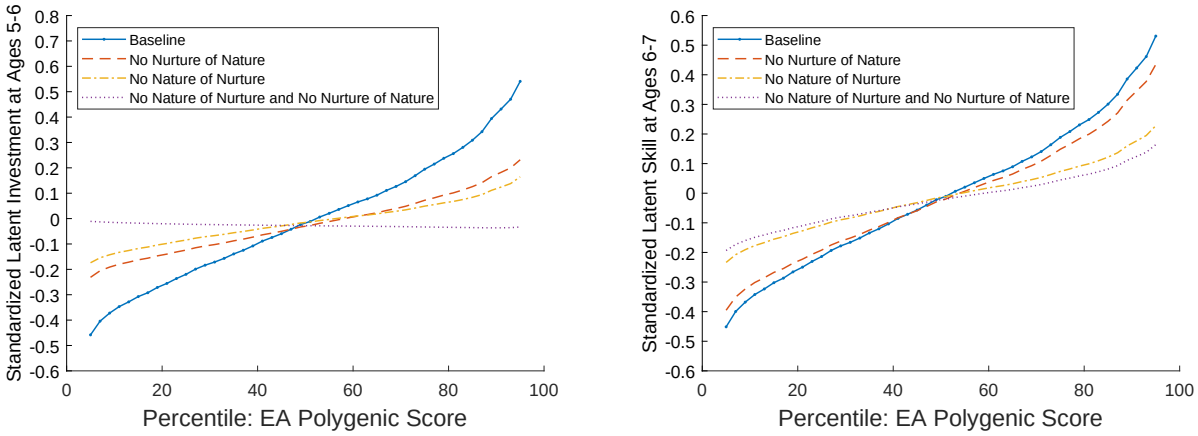
Notes: This table presents the proportion of the variance of latent parental investments at different periods that is explained by the model's four main components: i) the child's EA PGS, ii) the parental EA PGS, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.



(a) Child's EA PGS

(b) Parental EA PGS

Figure 3: Effects of EA PGS on Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' EA PGS 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 and the parental genetic score while keeping the other constant. This figure highlights how the effect of both parental and child genes increases over time.



(a) Latent Investment

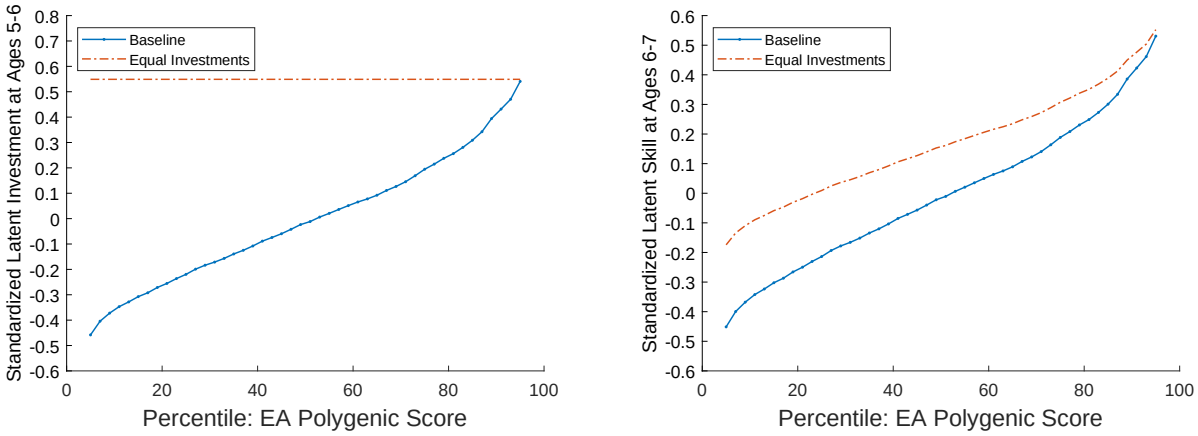
(b) Latent Skills

Figure 4: Mechanisms Decomposition: These figures compare baseline and simulated skills and investments when we shut down the nature of nurture and nurture of nature mechanisms.

Table 6: 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			
Nurture of Nature	0.00%	5.91%	7.74%	9.12%	10.20%	10.90%
Nature of Nurture	-3.75%	41.61%	47.34%	48.73%	49.29%	49.43%
Direct Effect	103.75%	52.48%	44.92%	42.15%	40.51%	39.67%
Panel B:			Parental Investments			
Nurture of Nature	33.99%	41.12%	44.18%	46.19%	46.74%	
Nature of Nurture	66.01%	58.88%	55.82%	53.81%	53.26%	

Notes: This table decomposes the association between the child's polygenic score 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.



(a) Latent Investment (b) Latent Skills

Figure 5: Equalising Investments: These figures compare baseline and simulated skills and investments when investments are equalized at the 95th percentile. We demonstrate graphically how a decrease in social inequality, via equalising parental investments, leads to a decrease in genetic inequality.

Table 7: Equal Investments: Skill Heritability by Age

Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
pgs_i	1.55%	2.93%	3.78%	4.26%	4.44%	4.54%
pgs_i^p	-0.00%	0.76%	1.54%	2.03%	2.24%	2.29%
Xs	6.81%	0.85%	0.18%	0.04%	0.01%	0.00%
ϵ and η	91.64%	95.46%	94.49%	93.68%	93.31%	93.18%

Notes: This table presents the proportion of the variance of latent skills at different periods that is explained by the four main components of the model under the scenario where all investments are equalized at the 95th percentile. The four components are i) the child's EA PGS, ii) the parental EA PGS, iii) observed characteristics of the child unrelated to genes (gender and birth order) and iv) unobserved factors unrelated to genetics.

Online Appendix

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

Mikkel Aagaard Houmark, Victor Ronda and Michael Rosholm

Appendix A Additional Information on the Estimation Sample

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

Because we rely on specific information about the child and the parents, we select a subsample of individuals for whom this information is available. The sample selection process is illustrated in Table A1. Our ALSPAC data set includes 14,062 children. We have genetic information on 8,804 of them. Among these, there are 6,638 individuals for whom we can compute the PGS of the mother as well. We drop individuals with missing information on gender. To ensure that the polygenic scores are valid, we further exclude non-European individuals and individuals who are outliers because they have extreme values (above 3.5 or below -3.5) on the first three principal components. This leaves us with 6,260 individuals who may be analysed in terms of the genetics of the mother and the child. Additionally, because we need the genetic information from both parents to be able to identify the effect of the child's genes, our main analysis includes only children for whom this information was collected. This reduces the sample size to 1,471 individuals.

Finally, we exclude individuals with many missing responses for the measures of skills and investments. We allow up to ten such missing values to be imputed for each summary measure, producing a final sample size of 1,267 for the main analysis. The results are robust to varying the maximum number of imputations, which is shown in Tables A2 and A3. Indeed, the parameter estimates are very similar in Panels A through E for both skills and investments.

Because we need to select our sample on the above criteria to make valid inference, our results should be interpreted for this subsample of ALSPAC individuals, i.e. those of European descent for whom both parents are present.[31] Excluding non-Europeans and principal component outliers is necessary for the polygenic scores to be valid. Furthermore, genetic information on both parents is necessary for identifying the effect of the child's genes that is independent of parental genes. The sample is somewhat positively selected from this

[31] The parents do not necessarily live together, but since information on the father's genetic endowments was collected at a later stage, this was only possible if the fathers were still available for the study.

selection procedure, as shown in Table A4. It is not surprising that especially the presence of the father correlates with higher PGS of the mother and child. However, while the averages are significantly different, the distributions are largely overlapping with the median child PGS in the estimation sample corresponding to the 59^{th} percentile in the full sample. Our estimates are therefore valid for a population that is genetically similar to the full population. Furthermore, the selected sample is not significantly different from the full sample in terms of gender and birth order, which are birth characteristics unrelated to the polygenic score.

Table A1: Sample Selection Procedure

Sample criteria	Observations
Full ALSPAC sample of children	14,062
With information on child genetic endowments	8,804
With information on maternal genetic endowments	6,638
Excluding individuals with missing information on gender	6,635
Excluding individuals of non-European descent	6,320
Excluding outliers on first three principal components	6,260
With information on paternal genetic endowments	1,471
With no more than ten missing measures of skills or investments	1,267

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

Table A2: EA PGS and Skills (missing values imputed)

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A (max 2 imputed):							
Child's PGS	0.041	0.016	0.023	0.060	0.091**	0.043	0.046
	(0.048)	(0.046)	(0.045)	(0.045)	(0.046)	(0.049)	(0.034)
Parental PGS	0.008	0.026	0.100**	0.097**	0.091*	0.095*	0.070**
	(0.049)	(0.047)	(0.046)	(0.047)	(0.047)	(0.050)	(0.033)
N	1051	1051	1051	1051	1051	1051	6306
Panel B (max 5 imputed):							
Child's PGS	0.039	0.009	0.024	0.066	0.099**	0.049	0.048
	(0.045)	(0.044)	(0.043)	(0.044)	(0.045)	(0.046)	(0.033)
Parental PGS	0.011	0.054	0.093**	0.115**	0.092**	0.082*	0.075**
	(0.046)	(0.045)	(0.044)	(0.045)	(0.045)	(0.047)	(0.032)
N	1178	1178	1178	1178	1178	1178	7068
Panel C (max 10 imputed):							
Child's PGS	0.046	0.008	0.024	0.078*	0.100**	0.041	0.050
	(0.044)	(0.043)	(0.042)	(0.044)	(0.044)	(0.045)	(0.033)
Parental PGS	0.002	0.053	0.097**	0.111**	0.094**	0.085*	0.074**
	(0.044)	(0.043)	(0.042)	(0.044)	(0.044)	(0.045)	(0.031)
N	1267	1267	1267	1267	1267	1267	7602
Panel D (max 20 imputed):							
Child's PGS	0.050	0.009	0.029	0.084**	0.108**	0.055	0.056*
	(0.043)	(0.042)	(0.042)	(0.043)	(0.043)	(0.044)	(0.032)
Parental PGS	-0.006	0.060	0.094**	0.102**	0.090**	0.073*	0.069**
	(0.043)	(0.042)	(0.042)	(0.043)	(0.043)	(0.044)	(0.031)
N	1336	1336	1336	1336	1336	1336	8016
Panel E (all imputed):							
Child's PGS	0.044	0.011	0.022	0.076*	0.104***	0.051	0.051*
	(0.041)	(0.039)	(0.039)	(0.040)	(0.040)	(0.040)	(0.030)
Parental PGS	-0.005	0.054	0.094**	0.093**	0.072*	0.066	0.062**
	(0.041)	(0.039)	(0.039)	(0.040)	(0.040)	(0.040)	(0.029)
N	1471	1471	1471	1471	1471	1471	8826

Notes: This table reports parameter estimates from regressions used to link the polygenic score for educational attainment to children's skills across childhood. For each of the skill measures, missing values are imputed with the median value for that measure. In Panel A, individuals with more than two such missing values for either skills or investments are excluded. This restriction is relaxed gradually in Panels B through E. Standard errors are reported in parenthesis.

Table A3: EA PGS and Investments (missing values imputed)

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A (max 2 imputed):							
Child's PGS	0.043	0.053	0.043	-0.061	-0.001	0.051	0.021
	(0.046)	(0.047)	(0.046)	(0.046)	(0.045)	(0.047)	(0.035)
Parental PGS	0.100**	0.175***	0.128***	0.226***	0.191***	0.167***	0.165***
	(0.047)	(0.048)	(0.047)	(0.047)	(0.046)	(0.048)	(0.035)
N	1051	1051	1051	1051	1051	1051	6306
Panel B (max 5 imputed):							
Child's PGS	0.034	0.058	0.040	-0.047	0.003	0.067	0.026
	(0.044)	(0.045)	(0.044)	(0.045)	(0.043)	(0.045)	(0.034)
Parental PGS	0.106**	0.177***	0.137***	0.227***	0.179***	0.168***	0.166***
	(0.045)	(0.046)	(0.045)	(0.045)	(0.044)	(0.046)	(0.036)
N	1178	1178	1178	1178	1178	1178	7068
Panel C (max 10 imputed):							
Child's PGS	0.034	0.054	0.035	-0.060	-0.008	0.075*	0.022
	(0.044)	(0.044)	(0.044)	(0.044)	(0.042)	(0.044)	(0.034)
Parental PGS	0.103**	0.174***	0.141***	0.237***	0.204***	0.166***	0.171***
	(0.044)	(0.044)	(0.044)	(0.044)	(0.042)	(0.044)	(0.035)
N	1267	1267	1267	1267	1267	1267	7602
Panel D (max 20 imputed):							
Child's PGS	0.027	0.054	0.035	-0.046	-0.004	0.077*	0.024
	(0.043)	(0.043)	(0.043)	(0.042)	(0.041)	(0.043)	(0.033)
Parental PGS	0.099**	0.172***	0.133***	0.218***	0.198***	0.158***	0.163***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.035)
N	1336	1336	1336	1336	1336	1336	8016
Panel E (all imputed):							
Child's PGS	0.035	0.047	0.036	-0.024	-0.014	0.076*	0.026
	(0.040)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)	(0.031)
Parental PGS	0.096**	0.171***	0.121***	0.191***	0.203***	0.147***	0.155***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)	(0.032)
N	1471	1471	1471	1471	1471	1471	8826

Notes: This table reports parameter estimates from regressions used to link the polygenic score for educational attainment to family investments across childhood. For each of the investments measures, missing values are imputed with the median value for that measure. In Panel A, individuals with more than two such missing values for either skills or investments are excluded. This restriction is relaxed gradually in Panels B through E. Standard errors are reported in parenthesis.

Table A4: Summary statistics

	Full sample	Reduced sample	Dropped individuals	Diff
Child's PGS	-0.287	3.01e-09	-0.336	***
	(1.006)	(1.000)	(0.999)	
N	8804	1267	7537	8804
Maternal PGS	-0.266	7.01e-09	-0.308	***
	(0.998)	(1.000)	(0.991)	
N	9265	1267	7998	9265
Female	0.488	0.468	0.490	
	(0.500)	(0.499)	(0.500)	
N	15052	1267	13785	15052
Birth order	1.708	1.676	1.711	
	(0.942)	(0.849)	(0.950)	
N	15645	1267	14378	15645

Notes: This table reports means and standard deviations for the full sample and for the estimation sample for whom we observe the polygenic scores of both parents. The last column reports the significance of a t-test for different means between the estimation sample and the individuals who are dropped from the full sample.

Appendix B Skill and Investment Measures

In this section, we describe the measurements used to identify skills and investments in more detail. The ALSPAC interviewed families in multiple periods during the child's development. Interviews started being conducted before the child's birth, at 8-42 gestational weeks. Interviews were conducted frequently, around every 6 months, and are still ongoing. The latest survey was conducted in 2019, when most children were 25 years old. The measurements used in this paper come from questionnaires completed by the primary caregiver, usually the child's mother.

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. These are displayed in Table B5.

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, either relating to objects present in the child's home or behavior involving the child and her parents. For such activities, we strived 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 the questions about belongings, the primary caregiver indicates the number of belongings to which the child has access (e.g., "number of books"). For the remaining questions, the primary caregiver indicates whether the parent does certain activities with the child (e.g., "Frequency child goes to the library") and at a which frequency: (i) "Nearly every day", (ii) "2-5 times per week", (iii) "Once per week" or (iv) "Once per month", (v) "A few times per year" or (vi) "Never" . The selected measures are displayed in Table B6.

To compute the summary score for skills and investments in each period, we first standardise each measure, then average across them and standardise again. The results are not sensitive to different standardisation procedures, e.g. not standardising each item before averaging.

Table B5: Measures of Child Skills

	Period	0	1	2	3	4	5
Measure	Age	0-2	2-3	3-4	4-5	5-6	6-7
1	Can build tower of 8 bricks	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	Can fit shapes in a board		x				
8	Can copy vertical line with pencil		x	x			
9	Can copy and draw a circle		x	x			
10	Combines two different words		x	x			
11	Uses plurals		x	x			
12	Uses possessives		x	x			
13	Adds -ing to words		x	x			
14	Adds -ed to words		x	x			
15	Can understand instructions			x			
16	Can copy and draw a plus sign / cross			x			
17	Can copy and draw a square			x	x		
18	Number of words child can string together			x	x		
19	Can write their name				x		
20	Can write any numbers				x		
21	Knows at least 10 letters				x		
22	Can read simple words				x		
23	Can read a story with <10 words per page				x		
24	Can count up to 20				x		
25	Can read a story with >10 words per page				x	x	x
26	Can count up to 100				x	x	x
27	Can play any board games				x	x	x
28	Plays chess					x	x

Notes: This table reports the individual measures of child skills. The x's indicate the specific periods in which each measure is used to construct the summary measure.

Table B6: Measures of Investments

	Period	0	1	2	3	4	5
Measure	Age	0-2	2-3	3-4	4-5	5-6	6-7
1	Freq. goes to places of interest	x	x	x	x	x	x
2	Freq. goes to library	x	x	x	x	x	x
3	Freq. mum reads to child	x	x	x	x	x	x
4	Freq. partner sings to child	x	x	x	x	x	
5	No. of books owned	x	x				x
6	Mum teaches shapes and sizes	x	x	x			
7	Mum teaches other things	x	x	x			
8	No. of times going to park	x	x	x			
9	Freq. partner reads to child	x		x	x	x	x
10	Mum teaches songs	x		x			
11	Freq. mum shows child picture books	x		x			
12	Freq. partner shows child picture books	x		x			
13	Freq. partner plays with toys with child	x		x			
14	No. of jigsaw puzzles		x				
15	No. of interlocking toys		x				
16	Freq. goes swimming				x	x	x
17	Freq. goes to special classes or clubs				x	x	x
18	Freq. plays a musical instrument				x	x	x
19	Freq. goes to special groups					x	x
20	Freq. mum has conversation with child						x
21	Freq. dad has conversation with child						x
22	Has board games						x
23	Makes collection of things, e.g. stamps						x

Notes: This table reports the individual measures of family investments. The x's indicate the specific periods in which each measure is used to construct the summary measure.

Appendix C Additional Reduced Form Results

In this section, we present additional reduced form results to show two points. First, we show that our use of a combined parental PGS does not affect the main results, as adding maternal and paternal PGS separately gives very similar estimates. Second, we provide additional insight into the effect of parental PGS by showing how this effect is mediated by socioeconomic controls.

The main question relating to the aggregation of parental genes is whether using the combined parental PGS affects the estimates of the effect of the child's genes. In Table C7, we show that this is not the case. In Panel C, where maternal and paternal PGS is added separately, the point estimates of the child's PGS are very similar to those in Table 1.

The table makes a couple of other points. Clearly, it is insufficient to just control for maternal PGS. This shows why it is essential to use information on both parents' genes to identify the importance of the child's genes. Second, although the coefficients are imprecisely estimated, they suggest that the genes of the mother may be more important than those of the father. However, recall that we cannot treat parental genes as exogenous. Doing so would require additional genetic information about the parents' parents. The question of relative importance of parents' genes is thus left for future research.

Table C8 shows that the same points apply to the investments measure. Here, it is worth noting that the coefficients on maternal and paternal PGS do differ significantly. This may be because maternal genes are more important for investments, but it may also be that our measure of investments is more closely related to the mother (who is the main respondent).

Next, we show additional results explaining why parental genes matter above their influence on the child's genes. We do this by introducing an extended set of family controls (specifically, education, income, occupational group and age of birth). As seen in Table C9 Panel B, this removes the influence of parental PGS. This means that the influence of the parental PGS is strongly mediated through these controls. Similarly, in Table C10, we show that the same is true for investments.

It is not surprising that such socio-economic variables are mediators in the relationship between parental genes and child development — especially so since we are using a polygenic score meant to capture genetic potential for educational attainment. The reason that it has an effect is thus that it influences the educational attainment and similar outcomes for the parents, which in turn affects the child. Note that this does not mean that all effects of parental genes are captured by these controls. Genes matter for a large number of skills and traits, some of which are not related to EA, but nevertheless may have an effect on the

family environment and child development.

Table C7: EA PGS and Skills by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.047*	0.047*	0.097***	0.158***	0.169***	0.101***	0.103***
	(0.029)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	(0.021)
Maternal PGS	( )	( )	( )	( )	( )	( )	( )
Paternal PGS	( )	( )	( )	( )	( )	( )	( )
Panel B:							
Child's PGS	0.017	0.037	0.084**	0.118***	0.132***	0.062*	0.075***
	(0.036)	(0.034)	(0.035)	(0.035)	(0.035)	(0.036)	(0.026)
Maternal PGS	0.051	0.017	0.022	0.066*	0.061*	0.065*	0.047*
	(0.036)	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)	(0.025)
Maternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Paternal PGS	( )	( )	( )	( )	( )	( )	( )
Panel C:							
Child's PGS	0.043	0.009	0.025	0.076*	0.099**	0.038	0.048
	(0.044)	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)	(0.032)
Maternal PGS	0.043	0.026	0.040	0.079**	0.071**	0.072**	0.055**
	(0.036)	(0.035)	(0.035)	(0.036)	(0.036)	(0.036)	(0.026)
Paternal PGS	-0.038	0.040	0.083**	0.059*	0.047	0.033	0.037
	(0.035)	(0.034)	(0.034)	(0.034)	(0.035)	(0.035)	(0.024)
Maternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Paternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)

Notes: This table reports parameter estimates from regressions used to link the EA PGS to children's skills across childhood. To test the effect of the EA PGS, we regress at each age the skill measure on the PGS, controlling for gender and the first 15 principal components of the genetic matrix. In Panel B, we add the maternal PGS, and in Panel C, we additionally add the paternal PGS. The outcomes have been standardized as described in the data section. Standard errors are reported in parenthesis.

Table C8: EA PGS and Investments by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.110***	0.180***	0.140***	0.117***	0.146***	0.194***	0.148***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.021)
Maternal PGS	( )	( )	( )	( )	( )	( )	( )
Paternal PGS	( )	( )	( )	( )	( )	( )	( )
Panel B:							
Child's PGS	0.071**	0.087**	0.083**	0.021	0.086**	0.133***	0.080***
	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.027)
Maternal PGS	0.065*	0.154***	0.095***	0.157***	0.099***	0.101***	0.112***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.028)
Maternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Paternal PGS	( )	( )	( )	( )	( )	( )	( )
Panel C:							
Child's PGS	0.033	0.050	0.034	-0.060	-0.008	0.072*	0.020
	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.033)
Maternal PGS	0.077**	0.165***	0.110***	0.182***	0.128***	0.120***	0.130***
	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.036)	(0.029)
Paternal PGS	0.053	0.053	0.070**	0.115***	0.133***	0.085**	0.085***
	(0.035)	(0.035)	(0.035)	(0.034)	(0.034)	(0.035)	(0.026)
Maternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Paternal PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)

Notes: This table reports parameter estimates from regressions used to link the EA PGS to family investments across childhood. To test the effect of the EA PGS, we regress at each age the investments measure on the PGS, controlling for gender and the first 15 principal components of the genetic matrix. In Panel B, we add the maternal PGS, and in Panel C, we additionally add the paternal PGS. The investment outcomes have been standardized as described in the data section. Standard errors are reported in parenthesis.

Table C9: EA PGS and Skills by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.045	0.009	0.024	0.076*	0.099**	0.039	0.049
	(0.044)	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)	(0.032)
Parental PGS	0.003	0.051	0.097**	0.108**	0.092**	0.082*	0.072**
	(0.043)	(0.042)	(0.042)	(0.042)	(0.043)	(0.043)	(0.031)
Family Controls	( )	( )	( )	( )	( )	( )	( )
Parental PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS	0.053	0.007	0.022	0.069*	0.085**	0.034	0.045
	(0.044)	(0.042)	(0.042)	(0.042)	(0.043)	(0.044)	(0.031)
Parental PGS	-0.048	-0.009	0.003	0.024	0.018	-0.000	-0.002
	(0.047)	(0.045)	(0.044)	(0.044)	(0.045)	(0.046)	(0.032)
Family Controls	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Parental PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
N	1267	1267	1267	1267	1267	1267	7602

Notes: This table reports parameter estimates from regressions used to link the EA PGS to children's skills across childhood. To test the effect of the EA PGS, we regress at each age the skill measure on the PGS, controlling for the parental PGS, gender and the first 15 principal components of the genetic matrix. In Panel B, we also control for the average family income between ages 1 to 5 years, maternal and paternal years of education, maternal and paternal occupational groups, mother's age at birth and birth order. 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 score. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Table C10: EA PGS and Investments by Age

Ages:	[0-2[	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[Pooled]
Panel A:							
Child's PGS	0.034	0.052	0.035	-0.059	-0.008	0.073*	0.021
	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.033)
Parental PGS	0.101**	0.169***	0.139***	0.231***	0.204***	0.160***	0.167***
	(0.043)	(0.043)	(0.043)	(0.042)	(0.042)	(0.043)	(0.034)
Family Controls	( )	( )	( )	( )	( )	( )	( )
Parental PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
N	1267	1267	1267	1267	1267	1267	7602
Panel B:							
Child's PGS	0.034	0.027	0.022	-0.075*	-0.022	0.049	0.006
	(0.041)	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)	(0.030)
Parental PGS	-0.045	-0.006	-0.006	0.085**	0.080*	0.043	0.025
	(0.043)	(0.042)	(0.044)	(0.043)	(0.043)	(0.043)	(0.033)
Family Controls	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Parental PGS	(X)	(X)	(X)	(X)	(X)	(X)	(X)
N	1267	1267	1267	1267	1267	1267	7602

Notes: This table reports parameter estimates from regressions used to link the EA PGS to family investments across childhood. To test the effect of the EA PGS, we regress at each age the investments measure on the PGS, controlling for the parental PGS, gender and the first 15 principal components of the genetic matrix. In Panel B, we also control for the average family income between ages 1 to 5 years, maternal and paternal years of education, maternal and paternal occupational groups, mother's age at birth and birth order. The investments 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 score. Standard errors are reported in parenthesis. In the pooled specification, standard errors are clustered at the individual level.

Appendix D Alternative Specification of Empirical Model

Here we present additional estimates for the estimation model parameters. We consider three different alternative specifications. In our benchmark specification, we aggregate paternal and maternal genetic effects into a single "parental PGS". In Appendix D.1, we consider paternal and maternal genetic influences separately. We also investigate what happens if we only control for maternal genes, which is theoretically problematic but allows a substantially larger sample size. These estimation results are reported in Appendix D.2. Finally, in our benchmark specification, we assume that the technology of skill formation is Cobb-Douglas. In Appendix D.3, we relax this assumption and allow for interactions between all parameters by estimation a translog function.

Appendix D.1 Mother and Father results

We first present the estimates where we decompose the combined effect of the parental EA PGS (pgs_i^p) into the effect due to mother's (pgs_i^m) and father's (pgs_i^f) genetic propensity for education. We report the parameter estimates, the heritabilities by age and the mechanisms decomposition in Tables D11, D12, D13 and D14 (corresponding to Tables 3, 4, 5 and 6 for the baseline specification). In addition, we illustrate the association between PGS and skills over time, the mechanisms decomposition counterfactual and the equalizing investments counterfactual in Figures D1, D2 and D3 (corresponding to Figures 3, 4 and 5 for the baseline specification).

The first thing to notice is that all other parameter estimates are virtually unchanged from our benchmark specification. The coefficient on the child's EA PGS is now 0.021 (compared to 0.022), 0.016 (0.016) and 0.012 (0.013) for initial skills, technology of skill formation and investments, respectively. The same is true for the effect of skills on skill formation and investments, where the parameters are completely unchanged, and for the effect of investments on skills, where the coefficient is 0.204 instead of 0.205. In other words, the results are completely robust to whether or not the mother's and the father's polygenic scores are included separately or in combination. Because of this, the mechanism decomposition in Table D14 shows that the relative contributions from the nurture of nature, the nature of nurture, and from the direct effect to skills at ages 6-7 are 11 percent, 50 percent, and 39 percent, which is within one percentage point of the corresponding numbers from the baseline model. And, as illustrated in Figure D2(a) and D2(b), the corresponding graphical illustrations of the contributions of each mechanism to investments and skills also look very similar to those from the baseline.

The second thing to notice is that, in all equations (Table D11), we have that the parameter estimate of the maternal PGS is larger than that of the paternal PGS. For both the technology of skill formation and for the investment policy function, the coefficient on the mother's PGS is approximately twice as large as that of the father's. These results suggest that, although the nature of nurture effect works through both parents, it mainly works through the mother. The consequence of this is illustrated in Figure D1(b) and D1(c), which shows that the association between parent EA PGS and child skills is stronger for the mother than the father at all ages.

However, as previously explained, parental genes are not exogenous in the model, and the nature of nurture effect may also include the nurturing influence of previous generations, in so far as this is associated with the parents' genes. Thus, while this apparent gap may reflect underlying biological differences, it may also reflect that mothers are generally more

involved in bringing up the child, or that the mother's family exerts a larger influence than the father's. If we had access to grandparents' genetic endowments, we could investigate this further. Given this limitation, we note that the difference between parents is interesting, but for our purpose, the important takeaway is that all the main findings are the same regardless of whether or not we combine the parental PGS.

Table D11: Mother and Father: Main Parameter Estimates

	$\ln \theta_{i0}$	$\ln \theta_{it+1}$	$\ln I_{it}$
pgs_i	0.021	0.016	0.012
	[0.001 , 0.038]	[0.005 , 0.031]	[-0.001 , 0.026]
pgs_i^m	0.013	0.017	0.037
	[-0.002 , 0.028]	[0.008 , 0.032]	[0.022 , 0.050]
pgs_i^f	-0.013	0.009	0.017
	[-0.027 , 0.002]	[0.001 , 0.020]	[0.003 , 0.028]
$\ln \theta_{it}$	.	0.469	0.265
	.	[0.418 , 0.538]	[0.180 , 0.302]
$\ln I_{it}$	.	0.204	.
	.	[0.118 , 0.290]	.
Constant	1.463	1.153	3.076
	[1.435 , 1.494]	[0.678 , 1.567]	[2.985 , 3.296]

Notes: The parameter estimates for the initial skill equation (Equation 15) are reported in the first column, for the technology of skill formation (Equation 13) in the second column, and for the investment policy function (Equation 14) in the third column. 90% bootstrap confidence intervals in brackets.

Table D12: Skill Heritability by Age

Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
pgs_i	1.56%	3.99%	5.54%	6.50%	6.97%	7.25%
pgs_i^m	0.50%	2.06%	3.23%	4.00%	4.38%	4.55%
pgs_i^f	0.65%	0.05%	0.38%	0.66%	0.82%	0.89%
Xs	6.64%	1.73%	0.96%	0.69%	0.56%	0.50%
ϵ and η	90.65%	92.18%	89.90%	88.15%	87.26%	86.81%

Notes: This table presents the proportion of the variance of latent skills at different periods that is explained by the model's four main components: i) the child's polygenic score for educational attainment, ii) the parental polygenic score for educational attainment, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table D13: Investment Heritability by Age

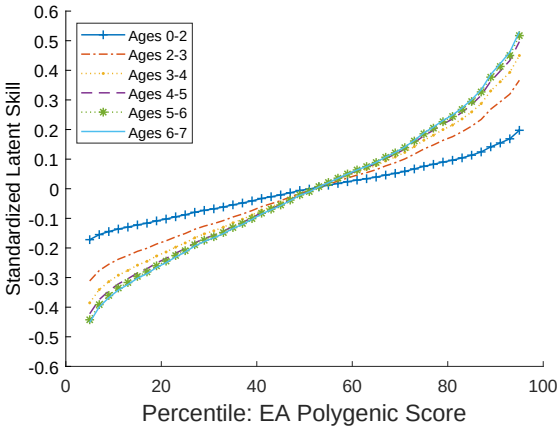
Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6
pgs_i	3.85%	4.77%	5.35%	5.61%	5.79%
pgs_i^m	5.39%	6.42%	7.00%	7.27%	7.52%
pgs_i^f	0.48%	0.84%	1.10%	1.21%	1.30%
Xs	6.73%	5.59%	5.08%	4.90%	4.77%
ϵ and η	83.56%	82.39%	81.47%	81.00%	80.62%

Notes: This table presents the proportion of the variance of latent parental investments at different periods that is explained by the model's four main components: i) the child's EA PGS, ii) the parental EA PGS, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

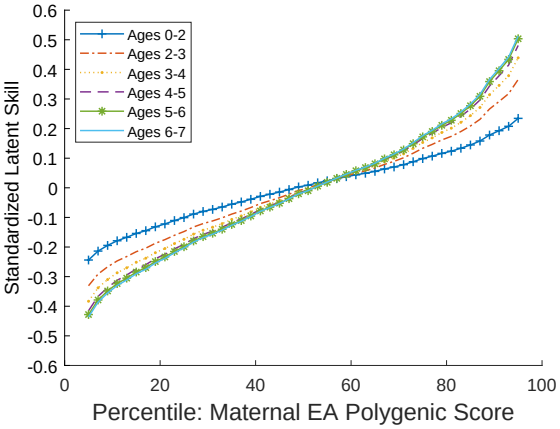
Table D14: Mechanisms Decomposition by Age

	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
Panel A:			Child's Skills			
Nurture of Nature	0.00%	5.59%	7.38%	8.75%	9.82%	10.51%
Nature of Nurture	0.32%	43.25%	48.56%	49.81%	50.31%	50.43%
Direct Effect	99.68%	51.15%	44.06%	41.44%	39.87%	39.06%
Panel B:			Parental Investments			
Nurture of Nature	32.42%	39.73%	42.89%	44.92%	45.50%	
Nature of Nurture	67.58%	60.27%	57.11%	55.08%	54.50%	

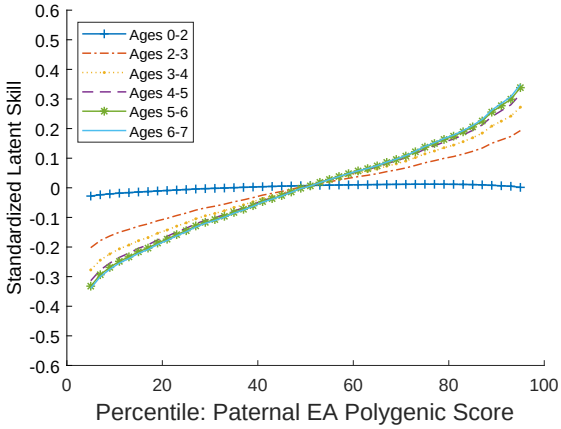
Notes: This table decomposes the association between the child's polygenic score for educational attainment and child's skills (in Panel A) and parental investments (in Panel B) by the three mechanisms for the different developmental periods.



(a) Child's EA PGS

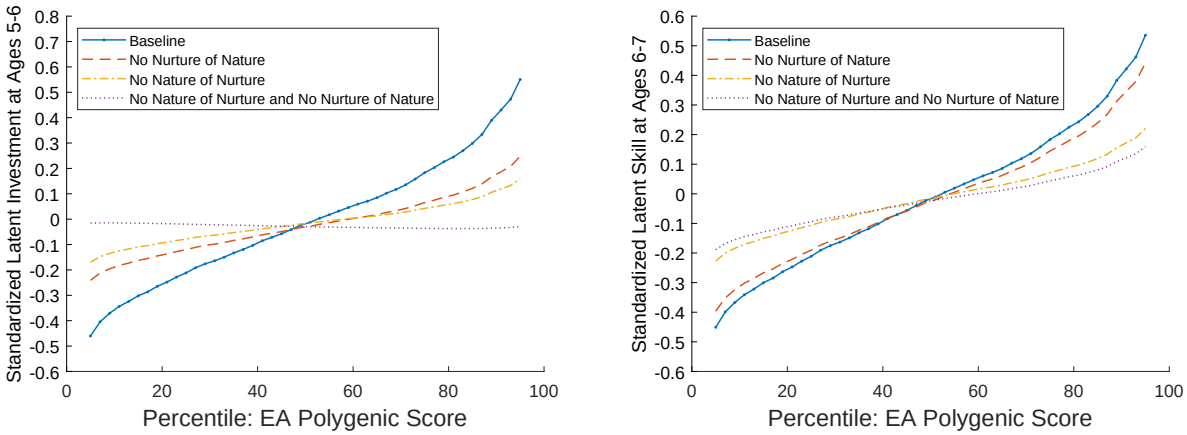


(b) Maternal EA PGS



(c) Paternal EA PGS

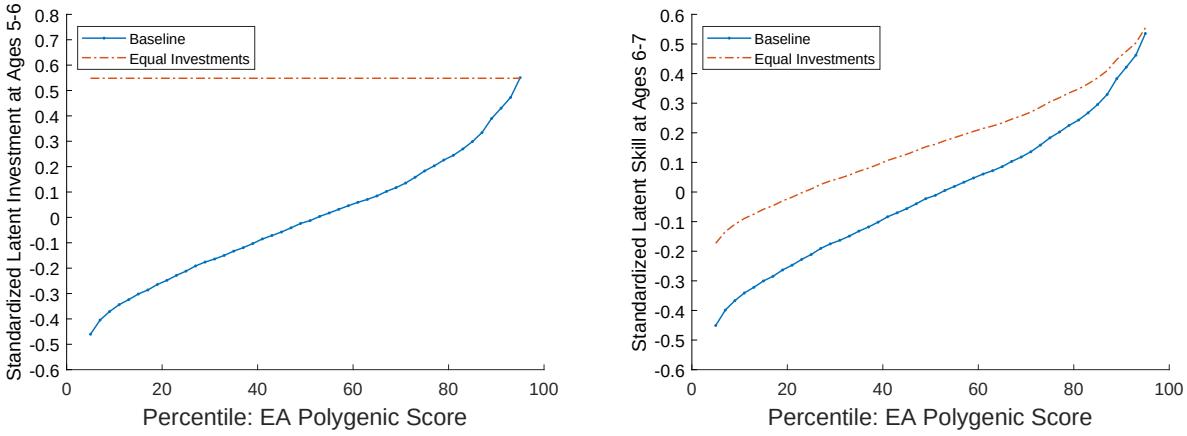
Figure D1: Effects of EA PGS on Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' EA PGS and the child's latent skill at different ages. Using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's and the parental genetic score while keeping the other constant. This figure highlights how the effect of both parental and child genes increases over time.



(a) Latent Investment

(b) Latent Skills

Figure D2: Mechanisms Decomposition: These figures compare baseline and simulated skills and investments when we shut down the nature of nurture and nurture of nature mechanisms.



(a) Latent Investment

(b) Latent Skills

Figure D3: Equalising Investments: These figures compare baseline and simulated skills and investments when investments are equalized at the 95th percentile. We demonstrate graphically how a decrease in social inequality, via equalising parental investments, leads to a decrease in genetic inequality.

Appendix D.2 Only Maternal Genes

Next, we present the estimates where we do not include the father's EA PGS in the model (parental PGS is thus simply the mother's PGS). This allows us to increase the sample size from 1267 to 4327. While not controlling for the father's genes is theoretically problematic, we thus consider the practical implication of this omission. This is especially relevant given the resulting sacrifice in sample size from including fathers in the analysis. We report the parameter estimates, the heritabilities by age and the mechanisms decomposition in Tables D15, D16, D17 and D18 (corresponding to Tables 3, 4, 5 and 6 for the baseline specification). In addition, we illustrate the association between PGS and skills over time, the mechanisms decomposition counterfactual and the equalizing investments counterfactual in Figures D4, D5 and D6 (corresponding to Figures 3, 4 and 5 for the baseline specification).

We see that, for both skills and investments, the coefficient on the child's genes is larger than in the main specification, while the coefficient on parental (now equivalent to maternal) genes is smaller than in the main specification. As previously explained, not controlling for parental genes induces an upwards bias in the estimate of the effect of the child's genes. We see that this is what happens: because the nature of nurture effect also works through the father, not including the father's polygenic score in the model means that the nature of nurture effect is underestimated. Furthermore, because the child's PGS is correlated with the father's PGS, this also means that the direct effect of the child's genes is overestimated.

Comparing the effect sizes, we see that this is especially critical for the investment policy function. Comparing Table D15 and Table 3, we see that the coefficient on parental PGS goes down from 0.041 to 0.037 when we exclude fathers. Despite this relatively modest change, we at the same time observe that the coefficient on the child's own PGS increases substantially from 0.013 to 0.034. This shows that not controlling for the father's genes causes us to not only underestimate the nature of nurture effect; it also causes us to greatly overestimate the nurture of nature effect.

In Table D18, we decompose the association between the child's EA PGS and skills at various ages into the parts explained by each mechanism. Compared to Table 6, we see that, at ages 6-7, the part explained by the nurture of nature is now much larger - 23 rather than 11 percent. Conversely, the part explained by the nature of nurture effect is down to 31 percent (from 49). This is also illustrated in Figure D5(a) and D5(b), which again suggests too large a role for the nurture of nature effect and too small a role for the nature of nurture effect. These additional results show that including genetic information on both parents in the analysis is important both theoretically and practically.

Table D15: Main Parameter Estimates

	$\ln \theta_{i0}$	$\ln \theta_{it+1}$	$\ln I_{it}$
pgs_i	0.008	0.017	0.034
	[0.000 , 0.016]	[0.013 , 0.022]	[0.028 , 0.041]
pgs_i^p	0.019	0.013	0.037
	[0.010 , 0.028]	[0.009 , 0.018]	[0.030 , 0.044]
$\ln \theta_{it}$	.	0.462	0.302
	.	[0.439 , 0.485]	[0.273 , 0.328]
$\ln I_{it}$	.	0.212	.
	.	[0.152 , 0.246]	.
Constant	1.448	1.164	2.950
	[1.430 , 1.465]	[1.001 , 1.381]	[2.870 , 3.032]

Notes: The parameter estimates for the initial skill equation (Equation 15) are reported in the first column, for the technology of skill formation (Equation 13) in the second column, and for the investment policy function (Equation 14) in the third column. 90% bootstrap confidence intervals in brackets.

Table D16: Skill Heritability by Age

Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
pgs_i	0.68%	3.60%	5.76%	7.11%	7.86%	8.29%
pgs_i^p	1.13%	1.96%	2.56%	2.91%	3.09%	3.20%
Xs	13.37%	6.78%	4.51%	3.50%	2.98%	2.75%
ϵ and η	84.82%	87.66%	87.17%	86.48%	86.07%	85.76%

Notes: This table presents the proportion of the variance of latent skills at different periods that is explained by the model's four main components: i) the child's polygenic score for educational attainment, ii) the parental polygenic score for educational attainment, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table D17: Investment Heritability by Age

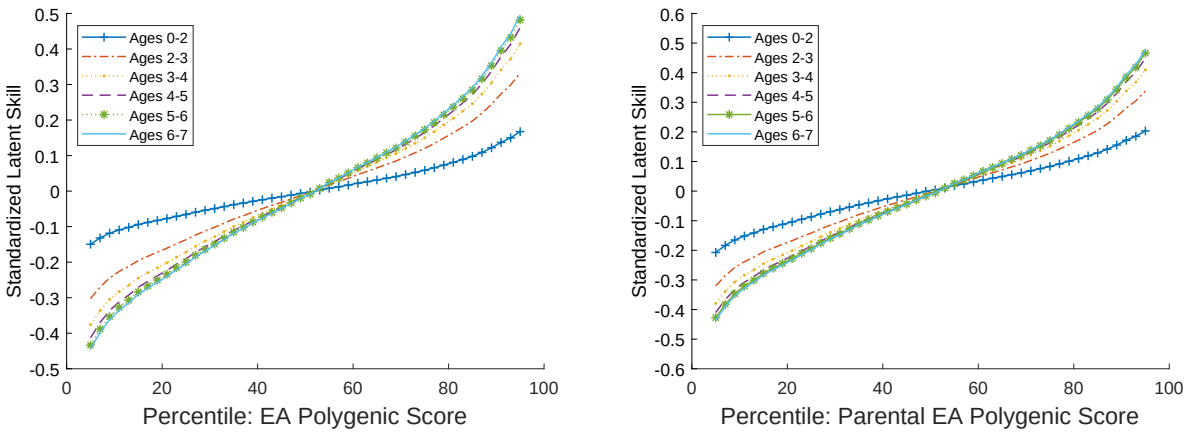
Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6
pgs_i	5.58%	7.09%	7.89%	8.38%	8.65%
pgs_i^p	3.43%	3.88%	4.11%	4.22%	4.32%
Xs	12.24%	10.59%	9.83%	9.46%	9.23%
ϵ and η	78.76%	78.44%	78.16%	77.94%	77.81%

Notes: This table presents the proportion of the variance of latent parental investments at different periods that is explained by the model's four main components: i) the child's EA PGS, ii) the parental EA PGS, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table D18: 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			
Nurture of Nature	0.00%	15.68%	19.22%	21.18%	22.46%	23.23%
Nature of Nurture	57.52%	37.77%	34.07%	32.48%	31.71%	31.25%
Direct Effect	42.48%	46.55%	46.70%	46.34%	45.83%	45.52%
Panel B:			Parental Investments			
Nurture of Nature	61.69%	65.23%	67.37%	68.24%	68.46%	
Nature of Nurture	38.31%	34.77%	32.63%	31.76%	31.54%	

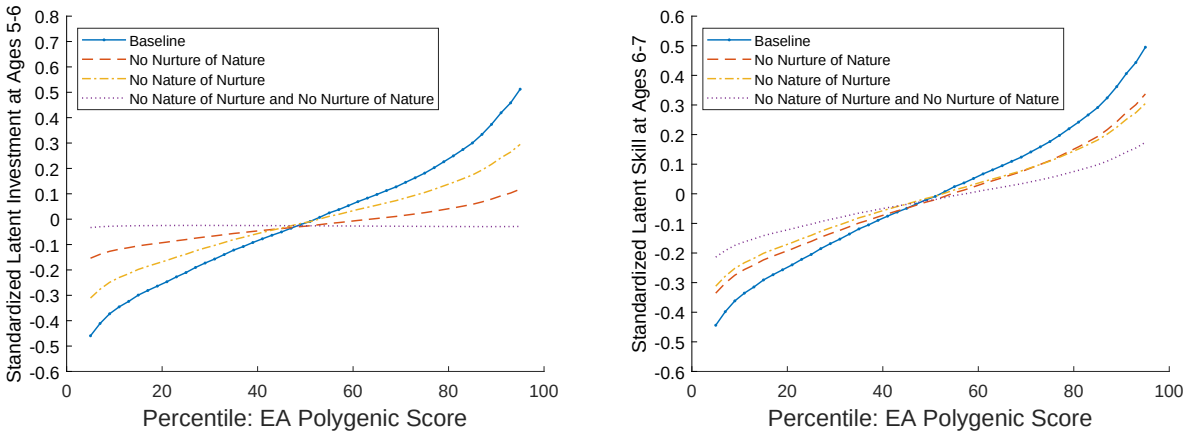
Notes: This table decomposes the association between the child's polygenic score for educational attainment and child's skills (in Panel A) and parental investments (in Panel B) by the three mechanisms for the different developmental periods.



(a) Child's EA PGS

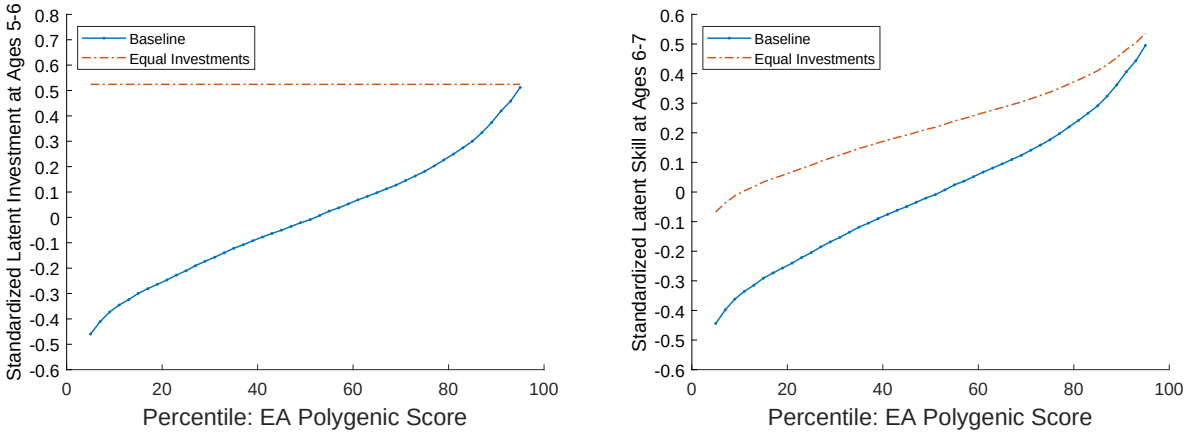
(b) Parental EA PGS

Figure D4: Effects of EA PGS on Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' EA PGS 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 and the parental genetic score while keeping the other constant. This figure highlights how the effect of both parental and child genes increases over time.



(a) Latent Investment (b) Latent Skills

Figure D5: Mechanisms Decomposition: These figures compare baseline and simulated skills and investments when we shut down the nature of nurture and nurture of nature mechanisms.



(a) Latent Investment (b) Latent Skills

Figure D6: Equalising Investments: These figures compare baseline and simulated skills and investments when investments are equalized at the 95th percentile. We demonstrate graphically how a decrease in social inequality, via equalising parental investments, leads to a decrease in genetic inequality.

Appendix D.3 Translog

Finally, we also present translog estimates, where the technology of skill formation allows for interactions between the four inputs; the child's genes, parental genes, the child's stock of skills, and family investments. We report the parameter estimates, the heritabilities by age and the mechanisms decomposition in Tables D19, D20, D21 and D22 (corresponding to Tables 3, 4, 5 and 6 for the baseline specification). In addition, we illustrate the association between PGS and skills over time, the mechanisms decomposition counterfactual and the equalizing investments counterfactual in Figures D7, D8 and D9 (corresponding to Figures 3, 4 and 5 for the baseline specification).

First, we note that the estimates are quite imprecisely estimated with respect to all the direct effects. For example, the confidence interval for the direct effect of the child's genes on skill formation ranges from -0.244 to 0.125, with the point estimate even being negative. However, this does not mean that the direct effect of genes is estimated to be negative, as there is also a large positive effect of previous skills as well as a positive effect of the child's PGS on initial skills.

Second, while four of the six interaction terms are insignificant, the remaining two are negative. These estimates suggest a negative complementarity between parental genes and investments, and another negative complementarity between skills and investments. The former can be interpreted as implying that investments are more important in families where the parents have a lower genetic propensity for education (or that parental genes are more important in families with lower levels of investments). In other words, the quantity and quality of investments are substitutes. The latter negative interaction can be interpreted as the returns to investments being lower for higher ability children.

What is not directly evident from the table is that the average effect of the four inputs are quite similar to the standard Cobb-Douglas model. The main difference is that the negative complementarities reduce the role of the mechanisms working through investments somewhat, as shown in Table D22. The nurture of nature and nature of nurture mechanisms now only account for 6 and 30 percent of the association between child EA PGS and skills at ages 6-7. Nevertheless, the overall implications for the relationship between child EA PGS and investments and are similar to the baseline model, as illustrated in Figure D8(a), with both the nature of nurture and the nurture of nature contributing the inequality in investments. Similarly, as illustrated in Figure D8(b), it is still the case that the inequality in skills is mostly due to the nature of nurture (in addition to the direct genetic effect). Interestingly, because of the negative complementarities outlined above, it now appears that these interactions between nature and nurture are mostly important for the lower half of the

skills distribution.

While these potential complementarities are interesting and relevant, the estimates are too noisy, and the confidence intervals on the interactions are too large for us to take these estimates seriously. For that reason and the fact that the implications discussed in Sections 6.5 and 7 still apply under this less-restrictive model, we chose the more restrictive Cobb-Douglas specification as our benchmark specification. We believe that our benchmark specification is a good compromise between the conceptual ideas outlined in our conceptual model and the identification that is possible with our data.

Table D19: Main Parameter Estimates

	$\ln \theta_{i0}$	$\ln \theta_{it+1}$	$\ln I_{it}$
pgs_i	0.022	-0.052	0.007
	[0.002 , 0.039]	[-0.244 , 0.125]	[-0.011 , 0.025]
pgs_i^p	-0.001	0.172	0.053
	[-0.018 , 0.018]	[0.009 , 0.348]	[0.027 , 0.074]
$\ln \theta_{it}$	.	0.771	0.268
	.	[0.525 , 0.957]	[0.172 , 0.308]
$\ln I_{it}$	.	0.445	.
	.	[0.151 , 0.633]	.
$\text{pgs}_i \times \text{pgs}_i^p$	.	-0.004	.
	.	[-0.010, 0.002]	.
$\text{pgs}_i \times \ln \theta_{it}$	.	0.001	.
	.	[-0.048, 0.057]	.
$\text{pgs}_i \times \ln I_{it}$	.	0.023	.
	.	[-0.017, 0.061]	.
$\text{pgs}_i^p \times \ln \theta_{it}$	.	0.007	.
	.	[-0.040, 0.046]	.
$\text{pgs}_i^p \times \ln I_{it}$	.	-0.066	.
	.	[-0.110, -0.017]	.
$\ln \theta_{it} \times \ln I_{it}$	.	-0.108	.
	.	[-0.166, -0.025]	.
Constant	1.463	0.612	2.028
	[1.434 , 1.494]	[-0.026 , 1.463]	[1.908 , 2.298]

Notes: The parameter estimates for the initial skill equation (Equation 15) are reported in the first column, for the technology of skill formation (Equation 13) in the second column, and for the investment policy function (Equation 14) in the third column. 90% bootstrap confidence intervals in brackets.

Table D20: Skill Heritability by Age

Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6	Ages 6-7
pgs_i	1.55%	2.35%	3.00%	3.54%	3.67%	3.62%
pgs_i^p	-0.00%	2.30%	2.65%	1.77%	0.86%	0.22%
Xs	6.81%	2.58%	1.30%	0.63%	0.29%	0.14%
ϵ and η	91.64%	92.78%	93.05%	94.06%	95.17%	96.02%

Notes: This table presents the proportion of the variance of latent skills at different periods that is explained by the model's four main components: i) the child's polygenic score for educational attainment, ii) the parental polygenic score for educational attainment, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table D21: Investment Heritability by Age

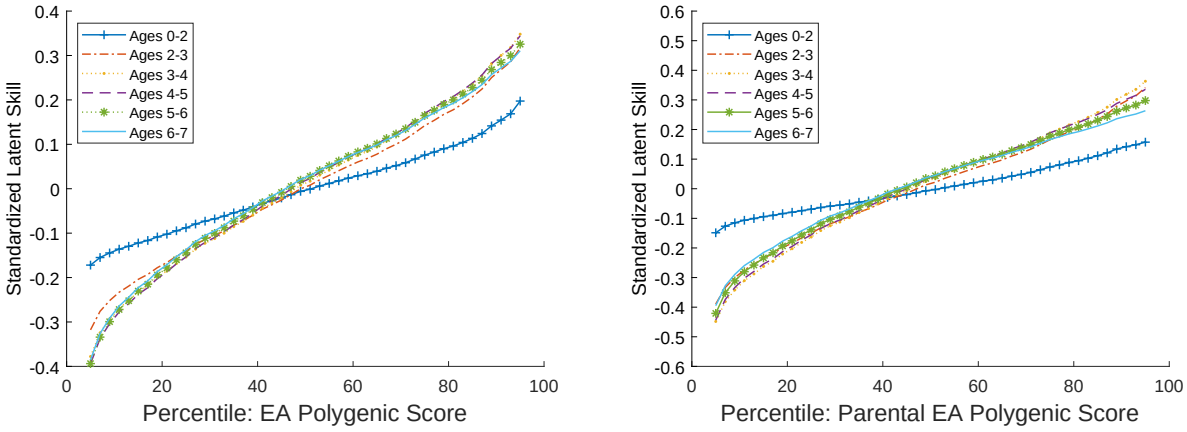
Mechanism	Ages 0-2	Ages 2-3	Ages 3-4	Ages 4-5	Ages 5-6
pgs_i	3.17%	3.19%	3.50%	3.77%	3.98%
pgs_i^p	7.68%	10.05%	10.41%	9.78%	9.27%
Xs	7.44%	6.56%	5.86%	5.46%	5.13%
ϵ and η	81.71%	80.20%	80.23%	80.99%	81.62%

Notes: This table presents the proportion of the variance of latent parental investments at different periods that is explained by the model's four main components: i) the child's EA PGS, ii) the parental EA PGS, iii) observed characteristics of the child unrelated to genes (gender and birth order), and iv) unobserved factors unrelated to genetics.

Table D22: 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			
Nurture of Nature	0.00%	4.80%	5.91%	6.34%	6.18%	5.70%
Nature of Nurture	-3.75%	60.72%	56.67%	46.37%	36.96%	29.88%
Direct Effect	103.75%	34.48%	37.42%	47.29%	56.86%	64.42%
Panel B:			Parental Investments			
Nurture of Nature	20.14%	27.36%	29.48%	29.99%	29.36%	
Nature of Nurture	79.86%	72.64%	70.52%	70.01%	70.64%	

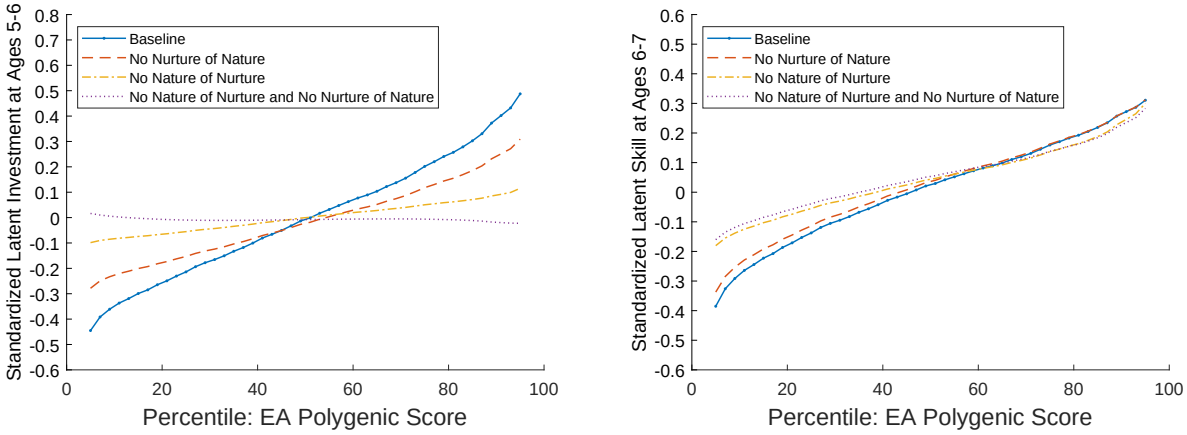
Notes: This table decomposes the association between the child's polygenic score for educational attainment and child's skills (in Panel A) and parental investments (in Panel B) by the three mechanisms for the different developmental periods.



(a) Child's EA PGS

(b) Parental EA PGS

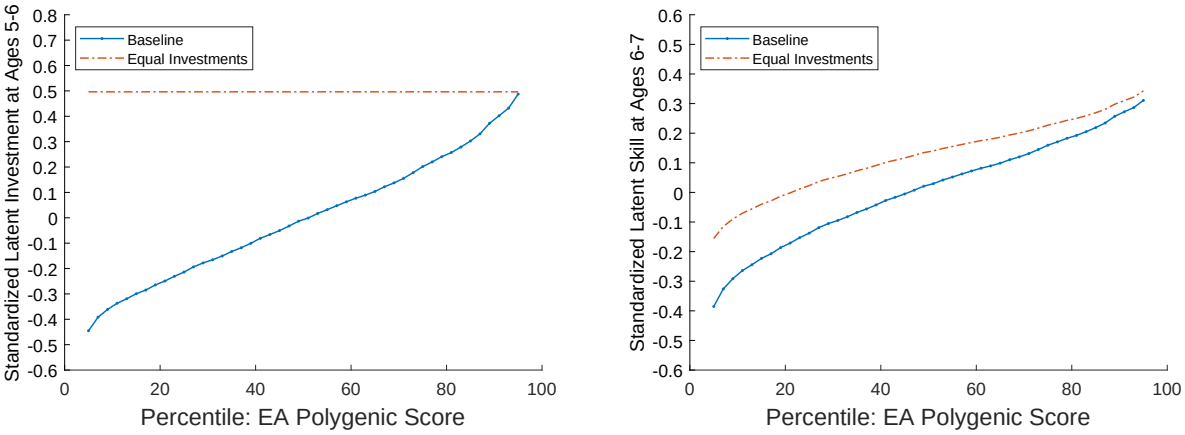
Figure D7: Effects of EA PGS on Latent Skills Across Child Development: These figures plot the relationship between the child's and its parents' EA PGS and the child's latent skill at different ages. Using the estimated model parameters, we simulate the expected latent skill at different ages when we separately increase the child's and the parental genetic score while keeping the other constant. This figure highlights how the effect of both parental and child genes increases over time.



(a) Latent Investment

(b) Latent Skills

Figure D8: Mechanisms Decomposition: These figures compare baseline and simulated skills and investments when we shut down the nature of nurture and nurture of nature mechanisms.



(a) Latent Investment

(b) Latent Skills

Figure D9: Equalising Investments: These figures compare baseline and simulated skills and investments when investments are equalized at the 95th percentile. We demonstrate graphically how a decrease in social inequality, via equalising parental investments, leads to a decrease in genetic inequality.

Appendix E Estimation Procedure

As described in Section 5.2, our estimation procedure is a slightly modified version of the algorithm proposed in Agostinelli and Wiswall (2020). The model is estimated in a series of steps. Here, we describe these steps in some detail. We first describe how we estimate the parameters in the measurement system. We then explain how these can be used to estimate the structural equations.

Appendix E.0.1 Estimating the Measurement System

First, we 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^{θ}_{jt} and λ^{I}_{kt}), the measurement means (μ^{θ}_{jt} and μ^{I}_{kt}) and the variance of the measurement errors ($\sigma^2_{jt,\theta}$ and $\sigma^2_{kt,I}$). 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^I_{11}, m^I_{21}, and m^I_{31}). Recall that we assume $\lambda^I_{11} = 1$ and that the measurement errors are independent (Assumption 23), so we can write the covariance between each pair of measurements as:

$$\begin{aligned}
Cov(m^I_{11}, m^I_{21}) &= 1 \ \cdot \lambda^I_{21} \cdot Var(\ln I_1) \\
Cov(m^I_{11}, m^I_{31}) &= 1 \ \cdot \lambda^I_{31} \cdot Var(\ln I_1) \\
Cov(m^I_{21}, m^I_{31}) &= \lambda^I_{21} \cdot \lambda^I_{31} \cdot Var(\ln I_1)
\end{aligned}$$

As first shown in Carneiro, Hansen, and Heckman (2003), we can use these three identities to identify the three unknowns (λ^I_{21}, λ^I_{31} and $Var(\ln I_1)$). To see this, note that:

$$\begin{aligned}
Var(\ln I_1) &= \frac{Cov(m^I_{11}, m^I_{21}) \cdot Cov(m^I_{11}, m^I_{31})}{Cov(m^I_{21}, m^I_{31})} \\
\lambda^I_{21} &= \frac{Cov(m^I_{21}, m^I_{31})}{Cov(m^I_{11}, m^I_{31})} \\
\lambda^I_{31} &= \frac{Cov(m^I_{21}, m^I_{31})}{Cov(m^I_{11}, m^I_{21})}
\end{aligned}$$

We can extend this procedure to include additional measurements beyond the first three. Also, the procedure can be applied to all periods to identify all factor loadings (λ^I_{kt}). 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 ($Var(\ln I_t)$ and $Var(\ln \theta_t)$) and the factor loadings are identified, we can also identify the mean of the latent variables ($E[\ln I_t]$ and $E[\ln \theta_t]$), and then the measurement means (μ^I_{kt} and μ^θ_{jt}). To see this, note that:

$$E[\ln I_t] = E[m^I_{1t}]$$

since we assume that $\mu^I_{1t} = 0$. Also, note that:

$$\mu^I_{kt} = E[m^I_{kt}] - \lambda^I_{kt} \cdot E[\ln I_t]$$

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

Lastly, once all other parameters are identified, we can identify the variance of the measurement errors ($\sigma^2_{jt,\theta}$ and $\sigma^2_{kt,I}$) from each measurement variance. To see this, note that:

$$\begin{aligned}\sigma^2_{kt,I} =& Var(m^I_{kt}) - (\lambda^I_{kt})^2 \cdot Var(\ln I_t)\\ \sigma^2_{jt,\theta} =& Var(m^\theta_{jt}) - (\lambda^\theta_{jt})^2 \cdot Var(\ln \theta_t)\end{aligned}$$

Appendix E.0.2 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 (eq. 13), the investment function (eq. 14), and in the early skills function (eq. 15). 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 and investments, we construct "residual measures" by subtracting the estimated measurement mean and the estimated factor loading, such that:

$$\widetilde{m}^\theta_{ijt} = \frac{m^\theta_{ijt} - \mu^\theta_{jt}}{\lambda^\theta_{jt}} = \ln \theta_{it} + \frac{\nu^\theta_{ijt}}{\lambda^\theta_{jt}} \tag{30}$$

$$\widetilde{m}^I_{ikt} = \frac{m^I_{ikt} - \mu^I_{kt}}{\lambda^I_{kt}} = \ln I_{it} + \frac{\nu^I_{ikt}}{\lambda^I_{kt}} \tag{31}$$

We use these residual measures to estimate the remaining parameters. For example, to estimate the investment policy function (eq. 14), we can use the kth residual measurement for the latent investment and the jth residual measurement for the latent skill instead of the

true unobserved latent variables.[32]

$$\widetilde{m}^I_{ikt} = \gamma_1 \widetilde{m}^\theta_{ijt} + \gamma_2 \, \text{pgs}_i + \gamma_3 \, \text{pgs}^p_i + \gamma_x \, X^I_{it} + \widetilde{\eta_{it}} \tag{32}$$

where $\widetilde{\eta_{it}} = \eta_{it} + \frac{\nu^I_{ikt}}{\lambda^I_{kt}} - \gamma_1 \frac{\nu^\theta_{ijt}}{\lambda^\theta_{jt}}$.

Estimation of equation 32 by OLS would yield inconsistent estimates of the γ coefficients because the two residual measures are correlated with their measurement errors which are included in the residual term $\widetilde{\eta_{it}}$. A common solution in the literature, which we follow here, is to use an instrumental variables estimator with the vector of excluded measurements $[m^\theta_{ij't}]$ as instruments for $\widetilde{m}^\theta_{ijt}$. 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 (eq. 13) and early skills function (eq. 15).

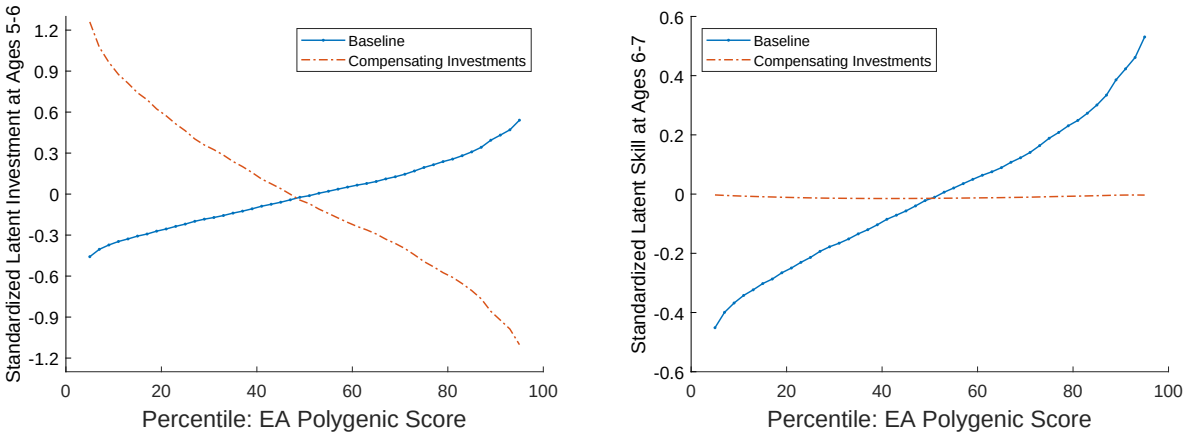
[32] In practice, we can use all possible combinations of investments and skill measurements to estimate the model parameters. There are many possible ways to use this large amount of measures. In our preferred specification, our parameters are averages of all possible combinations of measures for each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

Appendix F Compensating Investments

In this section, we present one additional counterfactual policy simulation. Similar to the 'Equal Investments' scenario, this counterfactual illustrates how the distribution of skills is affected by changing the distribution of investments. However, now we go one step further and ask: Is it possible to allocate investments such that skills at age 6-7 are completely equalized? Figure F10(a) shows that this is indeed possible. The blue line depicts the predicted relationship from our empirical model between the child's genetic endowments and latent investments, while the red line depicts the relationship between child EA PGS and investments that would result in an equalization of skills, as depicted by the corresponding lines in Figure F10(b).

We see that a complete elimination of the inequality in skills would require more than a reversal of the relationship between the child's EA PGS and investments (because there is also a direct effect of genes). A change from the 20th to the 80th percentile in the EA PGS corresponds to a 0.6 standard deviation increase in predicted investments. In the counterfactual scenario, this change conversely corresponds to a 1.2 standard deviation decrease in investments. In other words, this counterfactual simulation predicts that a complete equalization of skills at age 6-7 could come about if the relationship between child genes and investments was reversed and doubled in magnitude.

This 'Compensating Investments' scenario thus corresponds to either a world where parents engage in a strongly compensatory rather than reinforcing behavior, or a world where public investments serve this function. This simulation thus reemphasizes the point that there is no rivalry between the role of genes and the role of the environment. This is especially important for the potential role of public policy. Far from rendering these policies obsolete, the existence of genetic effects make them even more relevant.



(a) Latent Investment

(b) Latent Skills

Figure F10: Compensating Investments: These figures compare baseline and simulated skills and investments when investments are targeted at children with the lowest genetic propensity for education. We demonstrate graphically that genetic inequality can be eliminated under a complete redistribution of investments. Parental investments would need to be reversed in order to eliminate completely the link between genes and skills.