TI 2020-082/V
Tinbergen Institute Discussion Paper

Dynamic complementarity in skill production: Evidence from genetic endowments and birth order

Dilnoza Muslimova[1]
Hans van Kippersluis[1]
Cornelius A. Rietveld[1]
Stephanie von Hinke[1]
S. Fleur W. Meddens[1]

[1] Erasmus University Rotterdam

Dynamic complementarity in skill production: Evidence from genetic endowments and birth order

Dilnoza Muslimova[1,4,*], Hans van Kippersluis[1,4], Cornelius A. Rietveld[1,2,4], Stephanie von Hinke[1,3,4], S. Fleur W. Meddens[1,2]

[1]Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
[2] Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus University Rotterdam, The Netherlands
[3] School of Economics, University of Bristol, United Kingdom
[4] Tinbergen Institute, The Netherlands

*Corresponding author: muslimova@ese.eur.nl

Abstract

The birth order literature emphasizes the role of parental investments in explaining why firstborns have higher human capital outcomes than their laterborn siblings. We use birth order as a proxy for investments and interact it with genetic endowments. Exploiting only within-family variation in both ensures they are exogenous as well as orthogonal to each other. As such, our setting is informative about the existence of dynamic complementarity in skill production. Our empirical analysis exploits data from 15,019 full siblings in the UK Biobank. We adopt a family-fixed effects strategy combined with instrumental variables to deal with endogeneity issues arising from omitted variables and measurement error. We find that birth order and genetic endowments interact: those with above-average genetic endowments benefit disproportionally more from being firstborn compared to those with below-average genetic endowments. This finding is a clean example of how genetic endowments (`nature') and the environment (`nurture') interact in producing educational attainment. Moreover, our results are consistent with the existence of dynamic complementarity in skill formation: additional parental investments associated with being firstborn are more 'effective' for those siblings who randomly inherited higher genetic endowments for educational attainment.

Keywords: Birth order, dynamic complementarity, gene-environment interaction, educational attainment, polygenic score

JEL codes: D15, I24, J24

Acknowledgments: The authors gratefully acknowledge funding from NORFACE through the Dynamic of Inequality across the Life Course (DIAL) programme (Grant no. 462-16-100). Research reported in this publication was also supported by the National Institute on Aging of the National Institutes of Health under Award R56AG058726. C.A.R. and S.v.H. gratefully acknowledge funding from the European Research Council (GEPSI 946647; DONNI 851725). We gratefully acknowledge Aysu Okbay and employees and participants of the 23andMe, Inc cohort for sharing GWAS summary statistics for educational attainment. We would like to thank Pietro Biroli, Ben Domingue, Monique de Haan, Gabriella Conti, Kjell Salvanes, and seminar participants at the NORFACE DIAL workshop on intergenerational mobility and human capital accumulation, the Vienna University of Economics and Business, DIAL Mid-term Conference (University of Turku), RSF Summer Institute in Social Science Genomics, Health Economics Symposium (Erasmus University Rotterdam), and the Applied Microeconomics and Gender Workshop (University of Alicante).

1. Introduction

It is increasingly understood that important individual outcomes such as educational attainment are influenced by a complex interplay between 'nature' (i.e., genetic variation) and 'nurture' (i.e., environmental circumstances; see e.g., Rutter, Moffitt, and Caspi, 2006; Heckman, 2007). For example, the formation of skills relevant to educational attainment may result from dynamic complementarity between initial endowments and parental investments (Cunha & Heckman, 2007). Empirically estimating the contribution of genetic endowments and environments, and their possible interaction, is however complicated by the endogenous nature of both. Indeed, environmental characteristics are partially heritable and typically cluster together; e.g., higher educated parents tend to have higher incomes. Hence, it is not straightforward to disentangle what may be driving the 'environment-effect'. Similarly, although several population studies have found specific genetic variants that are associated with human capital outcomes such as educational attainment (Lee et al., 2018; Okbay et al., 2016; Rietveld et al., 2013), genetic variation is only random *conditional on parental genetic variation* (e.g., Lawlor, Harbord, Sterne, Timpson, & Davey Smith, 2008; Smith & Ebrahim, 2003). Not controlling for the latter implies that the 'gene-effect' may in fact reflect 'genetic nurture' – that is, the parental genotype can shape the environment in which children grow up, thereby producing a spurious association between the child's genetic variants and their outcomes (Belsky et al., 2018; Kong et al., 2018).

This paper is the first to exploit *exogenous* variation in *both* genetic endowments for education *and* the family environment to analyse gene-environment interactions for educational attainment. Genetic endowments are measured using a so-called "polygenic score" (PGS) – a highly predictive index constructed as the sum of all measured genetic variants, weighted by the strength of their correlation with educational attainment (Dudbridge, 2013; Lee et al., 2018). Our measure of the environment is an individual's birth order, which is consistently negatively correlated with educational attainment in developed countries (see e.g. Behrman et al., 1986; Black, Devereux, and Salvanes, 2005; Kantarevic et al., 2006; Booth and Kee, 2009; De Haan, 2010; Bagger et al., 2013; De Haan, Plug, and Rosero, 2014).

We overcome endogeneity issues by exploiting *within-family* variation in both genetic and birth order effects. Indeed, siblings' birth order is random within families (e.g., Damian and Roberts, 2015), and genetic variants are randomly assigned across siblings within a family according to

Mendel's Law. Hence, genetic variants are unrelated to birth order by construction.[1] We believe this is the first study in which both genetic variation and environments are truly exogenous, providing a compelling context in which we can fundamentally improve our understanding of the gene-environment interplay in shaping life outcomes. Do advantageous environments complement genetic advantages? Answering this question constitutes our first main contribution to the literature.

We choose birth order as a measure of the environment not merely for being conveniently uncorrelated with genetic endowments. In fact, economic theories of skill production provide a specific hypothesis of the sign of the interaction term, and our specific application is informative about the existence of "dynamic complementarity" in skill formation. Cunha and Heckman (2007) propose a model in which skills at period *t+1* (θ_{t+1}) are produced according to the production function f:

$$\boldsymbol{\theta_{t+1} = f_t(\theta_t, h, I_t)} \qquad (1)$$

where h denotes parental characteristics and I_t reflects parental investments in period *t*. Iteratively substituting equation (1) implies a model in which skills are a function of initial endowments θ_0, family-invariant parental characteristics h, as well as the entire history of parental investments $I_0, \dots, I_t$ (Heckman, 2007). A signature feature of the skill production function is the concept of dynamic complementarity, where skills raise the productivity of later investment $\left(\frac{\partial^2 \theta_{t+1}}{\partial \theta_t \partial I_t} > 0\right)$. In other words, children with higher (genetic) endowments (captured by θ_0) benefit the most from parental investments $I_0, \dots, I_t$.

We use the individual's birth order as an exogenous and highly predictive proxy for parental investments because we do not observe parental investments directly, and because realized parental investments are endogenous to the child's endowments (e.g., Becker and Tomes, 1986; Almond and Mazumder, 2013; Breinholt and Conley, 2019; Sanz-de-Galdeano and Terskaya, 2019). The theoretical literature on the 'quantity-quality trade-off' (Becker, 1960; Becker & Lewis, 1973; Becker & Tomes, 1976; Galor & Weil, 2000) shows that with each additional child, it is more expensive to maintain the same 'quality' children (i.e., with the same level of education or health), implying that parents invest less in laterborn children. Moreover,

[1] A systematic relationship between birth order and genes could arise when parents base their fertility decisions on the observed genetic endowments of their offspring, i.e., a stopping rule depending on the "quality" of children (e.g., Eirnæs & Pörtner, 2004). We find no such evidence in our sample; we discuss this below.

with parental preferences for fairness in investments over equality in outcomes (see e.g., Berry, Dizon-Ross, & Jagnani (2020)), parents tend to distribute their resources equally over their children, leading to natural reductions in time investments in laterborn children compared to their firstborn siblings.[2]

Indeed, firstborns have undivided attention until the arrival of the second child (Breining, Doyle, Figlio, Karbownik, & Roth, 2020), and the empirical literature suggests that a dominant channel through which birth order affects educational attainment is parental time investments (see e.g., Birdsall, 1991; Black, Grönqvist, & Öckert, 2018; Breining et al., 2020; De Haan et al., 2010; Monfardini & See, 2012; Pavan, 2016; Price, 2008). Using the American Time Use Survey (ATUS), Price (2008) shows that firstborns receive 20-30 minutes more daily quality time compared to their younger siblings (see also Black et al., 2018; Monfardini & See, 2012). Lehmann, Nuevo-Chiquero, & Vidal-Fernandez (2018) additionally show that with laterborn children, mothers postpone prenatal care, breastfeed less, and are more likely to smoke when not breastfeeding. Hence, while we do not dismiss other potential channels through which birth order may affect educational attainment,[3] parental investments are a prominent channel through which these effects arise, consistent with the evidence that parental investments are an important input into the child's skill production (e.g., Del Boca, Flinn, & Wiswall, 2014).

The theory of skill production, and in particular the concept of dynamic complementarity therefore provides a clear prediction for the sign of the Gene-by-Environment (G×E) interaction term: in a sample of siblings, being firstborn should show a positive interaction with genetic endowments. Hence, apart from providing a rare context in which economic theory helps formulate hypotheses about fundamental interactions between genetic variation and the environment, estimating the interaction between birth order and genetic endowments provides a unique setting to empirically test for dynamic complementarity. This constitutes our second main contribution.

[2] To illustrate the importance of this environmental determinant, Handy & Shester (2020) estimate that the rise in the fraction of later-born children was responsible for 20-35% of the stagnation in college completion among US baby-boom cohorts born between 1946 and 1974.

[3] Eirnæs & Pörtner (2004) distinguish three additional environmental channels through which birth order effects may arise: (i) younger children may benefit from the interaction with their older siblings; (ii) firstborns benefit from a lower maternal age and better maternal immune system (e.g., Behrman, 1988; Black et al., 2016); and (iii) in some societies the oldest son (or older children more generally) are favoured as they are the first to become economically independent. Hotz & Pantano (2015) additionally argue that parents may demand a higher level of discipline for firstborns to set an example for laterborn children, which may also affect educational attainment.

Our empirical analysis exploits data from 15,019 full siblings from the UK Biobank, a population-based sample from the United Kingdom (Fry et al., 2017). To measure participants' genetic endowments, we construct polygenic scores for educational attainment based on the results from our own tailor-made genome-wide association study (GWAS) that uses the UK Biobank sample but excludes all siblings and their relatives.[4] We adopt a family fixed effects approach to exploit within-family differences in genetic endowments and birth order to study the G, E and G×E effects on educational attainment, and we apply Obviously-Related Instrumental Variables estimation (ORIV, Gillen, Snowberg, & Yariv, 2019) to reduce measurement error in the polygenic score. We focus on firstborns versus laterborns, since the literature suggests birth order effects are particularly salient at this margin (e.g., Breining et al., 2020).

We confirm earlier findings that laterborns have a lower level of education than firstborns, and that one's genetic endowment for education is a strong predictor of years of education. We also confirm the long-held conjecture that genetic endowments do not differ systematically across birth order within a family. This finding corroborates that birth order effects must be due to environmental influences, such as greater parental time investments in firstborns. Our main finding is that birth order and genetic endowments interact: being firstborn and having higher genetic endowments for education exhibit a positive interaction, meaning that those with a high polygenic score benefit disproportionally more from being firstborn compared to those with a low polygenic score. This finding is a clean example of how genetic endowments and the environment interact in producing important life outcomes such as educational attainment. Moreover, our empirical results are consistent with the existence of dynamic complementarity in skill formation: additional parental time investments associated with being firstborn are more 'effective' among those siblings who randomly inherited higher genetic endowments for educational attainment.

Our paper speaks to three main literatures. First, we contribute to an emerging literature on gene-environment interactions (G×E), which addresses how the environment moderates the effect of genetic variants, and vice versa. Previous studies have typically examined interactions between polygenic scores and *endogenous* environments such as socio-economic status (e.g., (Barth, Papageorge, & Thom, 2020; Bierut, Biroli, Galama, & Thom, 2018; Ronda et al., 2020), childhood trauma (e.g., Mullins et al., 2016; Peyrot et al., 2014), or partner's death (e.g.,

[4] See Appendix A for definitions and explanations of the genetic terms used here.

Domingue, Liu, Okbay, & Belsky, 2017). The interpretation of these findings is complicated, because individuals with certain genetic predispositions may self-select into different environments (known as gene-environment correlation, or *rGE*; see Jencks, 1980; Schmitz & Conley, 2017a). In these analyses, therefore, the 'environmental effect' could be reflecting the effect of one's genotype through *rGE*, and the 'genetic effect' could be reflecting the rearing environment shaped by parental genotype ('genetic nurture'). A handful of studies use exogenous variation in environments to study G×E in educational attainment. For example, Conley & Rauscher (2013) analyse how random differences in the prenatal environment alter the genetic effects on education, depression, and delinquency. In a more recent study, Schmitz & Conley (2017b) use a polygenic score for educational attainment jointly with Vietnam War conscription as a natural experiment to study their interaction effect on educational attainment.[5] We push this literature one step further by not only considering exogenous variation in the environment, but *also* in genetic endowments by exploiting within-family variation in polygenic scores.

A second strand of literature that we speak to is the literature on birth order effects. This literature consistently finds that in developed countries, laterborn children have lower educational attainment. Birth order effect have also been found for other outcomes, though sometimes with mixed results, such as intelligence (Black, Devereux, & Salvanesz, 2011), health (Black, Devereux, & Salvanes, 2016; Pruckner et al., 2019), personality and leadership skills (Black et al., 2018), and delinquency (Breining et al., 2020). We contribute to this literature by studying heterogeneity in the birth order effect on educational attainment with respect to genetic endowments. The potential interaction between birth order and genetic endowments is not merely an important source of heterogeneity in the treatment effect, but one that – if present – carries over to the next generation, potentially exacerbating intergenerational inequalities (Havari & Savegnago, 2020).

Finally, we speak to the literature on skill production and dynamic complementarity (Cunha & Heckman, 2007; Cunha et al., 2010; Todd & Wolpin, 2003). Estimating dynamic complementarity in skill formation requires independent variation in initial endowments and subsequent investments, or, alternatively, exogenous variation in sequential investments over

[5] Studies with other outcomes that exploit exogenous environments include, e.g., Barcellos, Carvalho, & Turley (2018), Schmitz & Conley (2016) and Pereira, van Kippersluis, & Rietveld (2020).

time (Almond & Mazumder, 2013; Johnson & Jackson, 2019). Hence, empirically testing for dynamic complementarity is extremely challenging (Almond, Currie, & Duque, 2018). Indeed, the previous literature has almost exclusively focused on early-life outcomes such as birthweight as a measure of endowments (e.g., Datar, Kilburn, & Loughran, 2010; Figlio, Guryan, Karbownik, & Roth, 2014). However, such early life outcomes are affected by prenatal investments (Aizer & Cunha, 2012), meaning they partially capture parental choices and are therefore endogenous. Furthermore, parents respond to children's endowments (e.g. Adhvaryu & Nyshadham, 2016; Aizer & Cunha, 2012; Almond & Mazumder, 2013; G. S. Becker & Tomes, 1986; Bharadwaj, Eberhard, & Neilson, 2018; Datar et al., 2010; Frijters, Johnston, Shah, & Shields, 2013; Giannola, 2020; Hsin & Felfe, 2014), with recent studies suggesting that parental investments also respond to the *genetic* endowments of children (Breinholt & Conley, 2019; Fletcher, Wu, Zhao, & Lu, 2020; Houmark, Ronda, & Rosholm, 2020; Sanz-de-galdeano & Terskaya, 2019). Hence, measures of children's endowments are rarely clean of parental investments, and parental investments are rarely independent of endowments, posing a formidable empirical challenge to accurately identify dynamic complementarities in skill formation.[6]

We contribute to this literature by using exogenous variation in genetic endowments *and* parental investments. Indeed, our measure of endowments is randomly assigned within families and fixed at conception. It is therefore clean from parental investments. Furthermore, we proxy for parental investments using individuals' birth order, which is strongly associated with parental investments, but uncorrelated with genetic endowments. Essentially, employing birth order as a proxy for parental investments exploits the natural reduction in the time and money available with the arrival of a laterborn child, and is thus independent of the child's endowments. Using random within-family variation in genetic endowments and (birth order-induced) parental time investments, provides an innovative setting to empirically test for dynamic complementarity in skill formation. We show that the use of such within-family G×E, exploiting exogenous G as well as E, provides a novel way to test for dynamic complementarity more generally, which is not restricted to birth order effects, but extends to other (exogenous)

[6] A recent set of papers has examined rare cases where exogenous variation exists in both initial endowments as well as later-life investments, with mixed evidence. Some studies find evidence consistent with dynamic complementarity (Adhvaryu et al., 2019; Duque et al., 2018; Gunnsteinsson et al., 2014; Johnson & Jackson, 2019), whereas others find weaker evidence or even substitutability between endowments and investments (Lubotsky & Kaestner, 2016; Malamud et al., 2016; Rossin-Slater & Wüst, 2020). See Appendix C for a detailed overview.

parental investments and policy changes (e.g., on reducing student-teacher ratios). By informing the shape and properties of the production function for skill, our analysis is an important precursor to a structural model where parents face the child's skill production function alongside budget and time constraints to decide between own consumption and investments in their children. Our supportive evidence for dynamic complementarity emphasizes the importance of early-life investments being followed-up by later-life investments to reap the full benefits in terms of human capital outcomes (e.g., Cunha & Heckman, 2007).

The remainder of this paper is organized as follows. In section 2, we outline our empirical strategy. Section 3 discusses the data source. In section 4 we present our main results and a number of robustness checks. Section 5 provides a discussion of our results and concludes.

2. Empirical Strategy

We analyse the gene-environment (G×E) interaction between one's genetic predisposition towards educational attainment and birth order as an important environmental determinant of education. The empirical specification is rooted in the skill production function (1), where we assume that adult educational attainment is an increasing function of acquired skills by the end of childhood $g(\theta_{T+1})$ as in Cunha & Heckman (2008, equation 2) and Cunha et al. (2010, equation 2.2).

Following Todd & Wolpin (2003) and Cunha & Heckman (2008), we specify a linear production function, where years of completed education $\boldsymbol{Y}$ is a function of initial endowments θ_0, the history of parental investments $\boldsymbol{I}$, and unobserved parental characteristics $\boldsymbol{h}$. Empirically, we measure initial endowments by the polygenic score for educational attainment ($\boldsymbol{G}$), the history of parental investments is proxied by an indicator for being firstborn ($\boldsymbol{E}$), and parental characteristics are subsumed into the family fixed effect $\boldsymbol{\delta_j}$. We go beyond Todd & Wolpin (2003) and Cunha & Heckman (2008) by allowing for an interaction term between $\boldsymbol{G}$ and $\boldsymbol{E}$, to allow for dynamic complementarities. This leads to the following specification

$$\boldsymbol{Y_{ij} = \alpha_1 + \alpha_2 G_{ij} + \alpha_3 E_{ij} + \alpha_4 G_{ij} \times E_{ij} + \alpha_5 X_{ij} + \delta_j + \xi_{ij}} \qquad \textbf{(2)}$$

where for each individual *i* in family *j*, Y_{ij} is years of education, E_{ij} is equal to one for those who are firstborn and zero otherwise, and G_{ij} is the standardized polygenic score for education (see section 3 and Appendix A for more information on its construction). $G_{ij} \times E_{ij}$

is the interaction term. $\boldsymbol{X_{ij}}$ is the set of individual level controls, including gender, month and year of birth dummies (Black et al., 2005; Handy & Shester, 2020). It also includes the vector of the first 40 principal components (PCs) of the genetic relatedness matrix.[7] Finally, $\boldsymbol{\delta_j}$ denote family fixed effects and $\boldsymbol{\xi_{ij}}$ is the error term. We employ heteroskedasticity-robust standard errors, clustered at the family level. The parameter α_2 captures the association between the standardized polygenic score for education and the years of schooling, whilst α_3 estimates the average advantage in years of schooling for firstborn children compared with their laterborn siblings. α_4 shows the extent to which the polygenic score and being first born complement each other's effect on education and is therefore informative about the existence of putative G×E effects and dynamic complementarity.[8]

Following Black et al. (2005, 2011), Heiland (2009), Lehmann, Nuevo-Chiquero, & Vidal-Fernandez (2018), we compare within-family and between-family specifications. The inclusion of family fixed effects ensures that variation in the polygenic score and birth order is random, making polygenic score and birth order are orthogonal to each other.[9] Hence, we avoid endogeneity concerns arising from omitted variables by comparing siblings within the same family.

Since the GWAS underlying the construction of a polygenic score is based on a finite sample, our estimated polygenic score is a noisy proxy for the true (latent) polygenic score (e.g., Benjamin, Daniel Cesarini, Laibson, & Turley, 2020; van Kippersluis et al., 2020). Moreover, the GWAS on basis of which the polygenic score was constructed did not control for parental genotypes, again leading to measurement error in the resulting polygenic score (Trejo & Domingue, 2019). Both sources of measurement error lead to an attenuation bias in the coefficient of the polygenic score.[10] While we cannot solve the attenuation bias arising from

[7] Genetic principal components (PCs) can be used to control for subtle forms of population stratification (i.e., correlations between allele frequencies and environmental factors across subpopulations in the sample) in a between-family (population-level) analysis (Price et al., 2006; Rietveld et al., 2014). Although not strictly necessary to include in the within-family analysis due to the inclusion family fixed effects, we keep the PCs in all specifications to facilitate a clean comparison between the between-family and within-family results.

[8] In section 4.5, we study the interaction effect using more flexible approaches than the linear interaction presented here.

[9] In section 4.5, we further test robustness to including a dummy for the lastborn child to assess whether endogenous fertility decisions on basis of genetic endowments influence our results.

[10] The reason why classical measurement error as a result of finite-sample bias leads to attenuation bias is well known. It is more subtle why the exclusion of parental genotype in the discovery GWAS leads to an attenuation bias. The reason is that any polygenic score will reflect both direct genetic effects arising from the individual's genotype as well as indirect genetic effects arising from the omitted parental genotypes. The latter effects are

the omission of parental genotype in the GWAS, we follow DiPrete, Burik, & Koellinger (2018) and van Kippersluis et al. (2020) in applying Instrumental Variables (IV) to tackle the classic measurement error problem. More specifically, we split our discovery GWAS sample into two equal halves and construct two polygenic scores based on the two discovery samples. Even though the two polygenic scores individually have lower predictive power, the measurement error in the two is plausibly orthogonal and so they can be used as instrumental variables for each other. Using these two polygenic scores, we apply Obviously-Related Instrumental Variables (ORIV; Gillen et al., 2019). See Appendix C for more details.

3. Data

We use data from the UK Biobank (2006-2010), a population-based cohort with approximately 500,000 individuals aged between 40-69 at the time of interview and living within a radius of 40 km from one of the 22 assessment centres in England, Wales, and Scotland (Fry et al., 2017). It contains survey data, biomarker and DNA samples, physical measurements, and linkage to inpatient registers and death records (Sudlow et al., 2015). Because participation in the UK Biobank is voluntary, it is not a representative sample of the UK population (see Fry et al. (2017) for a detailed analysis).

We apply the following sample selection criteria. We begin with 502,498 consented individuals. We follow the literature and remove those of non-European descent (92,892 observations), twins and multiple births (9,310 observations), and individuals with missing or conflicting information regarding the number of siblings and/or family size (3,801 observations). In doing so, we arrive at a sample of 396,494 individuals. We further restrict this sample to individuals with at least one sibling[11] in the UK Biobank and without missing values on any of the variables included in our analysis (i.e., years of education, birth order, family size, year and month of birth, principal components, gender, and the polygenic score for education). Since the UK Biobank did not specifically target families, this leads to a final sample size of N = 15,019 siblings.

known as 'genetic nurture' (e.g., Kong et al., 2018). When applied within families, the differences in the polygenic score arising from parental genotype are spurious since parental genotype is the same across siblings. Hence, part of the differences across siblings in the polygenic score is spurious and can be considered measurement error attenuating the resulting within-family estimates (Trejo & Domingue, 2019).

[11] Siblings are identified based on the genetic data; there are no self-reported siblings in the UK Biobank. See Appendix B for the full procedure followed to identify siblings.

We follow the literature (see e.g. Lee et al., 2018; Okbay et al., 2016; Rietveld et al., 2013) and convert individuals' qualifications to equivalent years of education using the International Standard Classification of Education (ISCED).[12] The average years of education for the sibling sample is 13.9 years (see Table 1).

We construct individuals' birth order on the basis of their response to a question of how many older siblings they have. If a respondent reports zero older siblings, the birth order is set to one. For individuals with missing information on the number of older siblings, we determine birth order based on family size and birth year of the individual and his/her siblings if all of them are present in the UK Biobank. This adds information on birth order for 1,752 siblings in our analysis sample. Table 1 shows that we have 5,911 firstborns (39.4%), with an average birth order of 1.91 (where we have censored birth order at 5 for the 245 respondents with birth order beyond 5). Around 37% of our sample is lastborn, and the average family size is 3 (i.e., the average number of siblings is 2).

Table 1: Descriptive statistics analysis sample (N = 15,019)

Variable	Mean	S.D.	Min.	Max.
Years of education	13.899	5.020	7.000	20.000
Firstborn (1/0)	39.4%			
PGS for years of education	0.000	1.000	-3.938	4.166
Birth order	1.913	0.997	1.000	5.000
Secondborn	41.51%			
Thirdborn	11.23%			
Fourthborn	4.31%			
Fifth- or laterborn	3.60%			
Family size	2.987	1.527	2.000	14.000
Last child (1/0)	36.9%			
Male (1/0)	42.5%			

Notes: S.D. = Standard deviation; Min. = Minimum; Max. = Maximum.

Our measure of genetic endowment for education is the polygenic score for education. A polygenic score is a weighted sum of genetic variants called Single Nucleotide Polymorphisms (SNPs, see Appendix A for details). The SNP weights are determined by the association between a SNP and years of education (Dudbridge, 2013) in an independent (discovery) sample:

[12] Years of education ranges from 7 to 20, where College or University degree is equivalent to 20 years, National Vocational Qualification (NVQ), Higher National Diploma (HND), or Higher National Certificate (HNC) to 19 years, other professional qualifications to 15 years, having an A or AS levels or similar to 13 years, O levels, (General) Certificate of Secondary Education ((G)CSE) to 10 years, and if none of the above to the lowest level of 7 years.

$$PGS_i = \sum_{j=1}^{J} \beta_j x_{ij}, \quad (3)$$

where PGS_i is the value for the polygenic score for individual *i*, β_j is the regression coefficient of SNP *j* $(j = 1, \dots, J)$ from the GWAS, and x_{ij} is the genotype of individual *i* for SNP *j* (coded as 0, 1 or 2, indicating the number of "effect" alleles). The polygenic scores are standardized within the sibling sample to have mean 0 and standard deviation 1.

The polygenic score measures the genetic predisposition towards educational attainment within the environmental context and demographic characteristics of the *discovery* GWAS sample (Domingue, Trejo, Armstrong-Carter, & Tucker-Drob, 2020). It is therefore preferable to select discovery and analysis samples from the same environmental context, especially when analysing gene-*environment* interactions. At the same time, the discovery sample should be independent of the analysis sample to avoid overfitting (Dudbridge, 2013). As an optimal balance, we therefore construct the polygenic score by using the weights from our own tailor-made GWAS that uses the UK Biobank sample *without siblings and their relatives*. Siblings' relatives were identified on the basis of their genetic data. The GWAS discovery sample comprises 392,771 individuals from the UK Biobank; we use the summary statistics from these analyses to create the polygenic scores on the sample of 15,019 siblings. This tailor-made polygenic score alleviates the differences between the discovery and the analysis samples in terms of demographics and environmental context, as well as measurement (i.e., the variables of interest are measured in the same way (Elam, Clifford, Shaw, Wilson, & Lemery-Chalfant, 2019; Tropf et al., 2017). Moreover, running our own GWAS enables the construction of two independent polygenic scores, obtained by splitting the discovery GWAS sample into two equal halves, that can be used in ORIV. This approach has been shown to outperform a single polygenic score that is based on meta-analysing multiple cohorts (van Kippersluis et al., 2020).[13]

[13] As a robustness check, we constructed a polygenic score based on the meta-analysed GWAS results of Okbay et al. (2016) including 23andMe summary statistics and our own UK Biobank discovery sample GWAS. As expected, this polygenic score is more predictive for educational attainment than the polygenic score constructed on the basis of the UK Biobank only. However, this polygenic score is based on several discovery cohorts from very different environmental contexts and does not allow us to use ORIV since we do not have access to all underlying samples to allow us to create multiple polygenic scores. The results are qualitatively the same, but as expected given that this polygenic score reflects a different environmental context (see e.g., Domingue et al. 2020), the interaction effect is estimated to be smaller and does not reach statistical significance at conventional levels (results available upon request from the authors).

4. Results

4.1. Predictive power of the polygenic score for educational attainment

Figure 1 shows that our polygenic score for years of education is normally distributed. We divide the polygenic score in 200 bins; the dots represent the average years of education for each bin. The line through the dots is obtained from a local polynomial regression of years of education on our polygenic score. In line with the literature (Lee et al., 2018; Okbay et al., 2016; Rietveld et al., 2013), the polygenic score is positively correlated with years of education ($r = 0.24$, $p < 0.001$). The average difference between those two standard deviations below the mean of the polygenic score, and those two standard deviations above the mean is almost 4 years of completed education, highlighting the substantial predictive power of the polygenic score. Furthermore, Figure 1 suggests the relationship is approximately linear, with little suggestion of any strong non-linearities between years of education and the polygenic score.

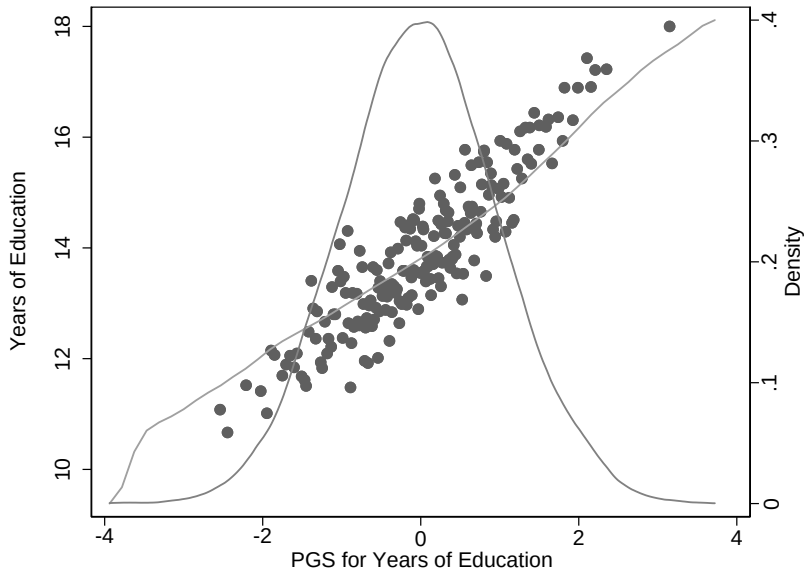

Figure 1. The relationship between the standardized polygenic score and years of education in the analysis sample.

Table 2 shows that the incremental R^2 of the polygenic score (i.e., the additional variance explained by the polygenic score after controlling for gender, month and year of birth, and the first 40 principal components) is 6.0% in the between-family analysis (i.e., 0.095-0.035=0.060; Columns 1 and 2). In the specifications with family fixed effects (Columns 3 and 4), the incremental (within) R^2 for the polygenic score is reduced to 1.0%. This reduction in predictive power when moving from between-family to within-family estimates is well-established in the

literature (see e.g., Koellinger & Harden, 2018; Kong et al., 2018; Lee et al., 2018; Rietveld et al., 2013; Selzam et al., 2019). The reduction reflects the fact that family fixed effects account for the shared family environment and parental genotype, which was not accounted for in the between-family specification. In terms of the effect sizes, we observe that a one standard deviation increase in the polygenic score is associated with an increase of 1.23 years of education. With family fixed effects, the effect size is reduced to 0.633.

Table 2. Results of the regressions of years of education on the polygenic score (PGS).

	Between-family analysis		Within-family analysis	
	(1)	(2)	(3)	(4)
PGS for years of education		1.234***		0.633***
		(0.038)		(0.071)
Constant	15.664***	15.099***	15.383***	15.227***
	(1.696)	(1.766)	(2.336)	(2.433)
R^2	0.035	0.095	0.028	0.038
N	15,019	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family in the within-family analysis; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

4.2. The relationship between birth order and educational attainment

Figure 2 shows the raw differences (without family fixed effects and other control variables) in years of education by birth order (panel A). While the first three children achieve similar levels of education on average, the averages for children with higher birth order are clearly lower (albeit with higher variances). When pooling all laterborns together (panel B), the difference between firstborns and laterborns is relatively small but clearly visible.

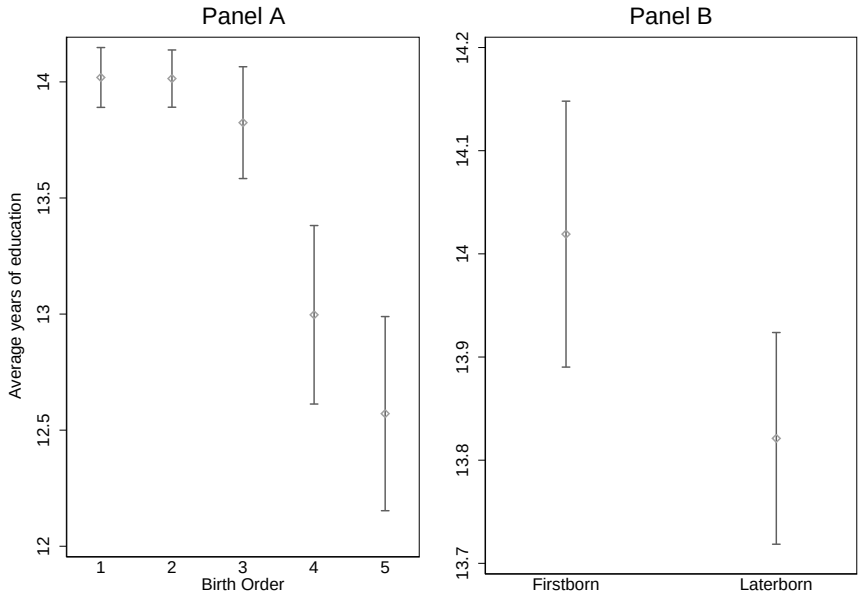

Figure 2. The relationship between birth order and years of education in the analysis sample.

Table 3. Results of the regressions of years of education on different specifications of birth order.

	Between-family analysis		Within-family analysis	
	(1)	(2)	(3)	(4)
Firstborn	0.334***		0.362***	
	(0.089)		(0.111)	
2nd born		-0.302***		-0.348***
		(0.093)		(0.124)
3rd born		-0.317**		-0.391
		(0.148)		(0.243)
4th born		-0.846***		-0.544
		(0.226)		(0.366)
5th born		-0.878***		-0.090
		(0.271)		(0.483)
Constant	16.021***	16.219***	14.476***	14.898***
	(1.752)	(1.743)	(2.333)	(2.333)
R^2	0.044	0.044	0.030	0.030
N	15,019	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family in the within-family analysis; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

Table 3 confirms the birth order effects in the specifications with and without family fixed effects, and shows that conditional on the control variables, the differences become more salient. We observe a consistent gap of 0.3-0.4 years of schooling between first- and laterborn children. The direction and magnitude of the effect is robust to using the binary indicator or the categorical variable for birth order. The within-family effect sizes for birth orders higher than three do not reach statistical significance due to the relatively small number of observations (see Table 1).

4.3. The relationship between birth order and the polygenic score for years of education

To measure causal gene-environment interactions, it is important to show that birth order is orthogonal to the polygenic score. Figure 3 provides a first impression on the raw relationship between the two measures, without controlling for family fixed effects or other control variables. Panel A illustrates that educational attainment polygenic scores for later-born children tend to be slightly lower. The same pattern holds when reducing the comparison to firstborns versus laterborns in Panel B. However, these differences are not statistically significant. Furthermore, the distributions of the polygenic score by birth order are the same (Panels C and D).

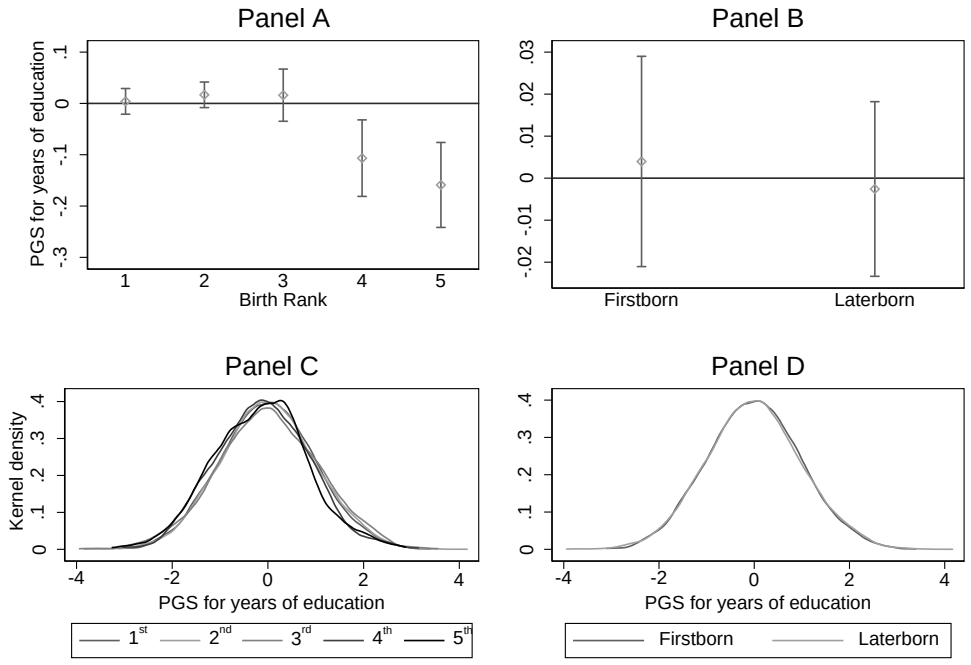

Figure 3. The relationship between birth order and the polygenic score for years of education in the analysis sample.

Consistent with the graphical evidence, Table 4 shows a slight difference of 0.04 standard deviations in the polygenic score between firstborn and laterborn children in the between-family analysis. However, when looking at this relationship *within* families, the difference becomes neither economically nor statistically significant.[14] These results establish the presumption that firstborns on average do not have different genetic endowments compared to their laterborn siblings (c.f. Mendel's law). The results also corroborate the notion that there is no gene-environment correlation (*rGE*) between the polygenic score and birth order.

Table 4. Results of the regressions of polygenic score for educational attainment on birth order.

	Between-family analysis		Within-family analysis	
	(1)	(2)	(3)	(4)
Firstborn	-0.041**		-0.016	
	(0.018)		(0.018)	
2nd born		0.040**		0.018
		(0.019)		(0.020)
3rd born		0.073**		0.022
		(0.031)		(0.038)
4th born		-0.027		0.018
		(0.045)		(0.057)
5th born		-0.037		0.036
		(0.055)		(0.077)
Constant	0.544	0.488	0.288	0.279
	(0.343)	(0.341)	(0.382)	(0.383)
R^2	0.018	0.018	0.012	0.012
N	15,019	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family in the within-family analysis; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

[14] The fact that in the within-family analysis, there is no significant differences in polygenic scores across birth order suggests that the raw relationship depicted in Figure 3 is mostly driven by family size. This is indeed what we confirm, with the coefficient on the control variable family size being negative and statistically significant in the between-family analysis.

4.4. Gene-environment interaction and dynamic complementarity of skill formation

Table 5 presents the gene-environment interaction results in the between-family and within-family analyses. Comparing Columns 1 and 4 in Table 5 to the estimates presented above (Table 2 and Table 3) shows that the addition of the educational attainment polygenic score does not affect the *direct* effect of birth order on years of schooling. This comparison confirms once again that the polygenic score and birth order are independent. A one standard deviation increase in the polygenic score is estimated to raise years of education by 1.23 years (between-family analysis) and 0.64 years (within-family analysis). Furthermore, firstborns enjoy on average 0.37-0.40 extra years of schooling compared with laterborns.

Table 5. Results of the regressions of years of education on the gene-environment interaction.

	Between-family analysis			Within-family analysis		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	ORIV	OLS	OLS	ORIV
Firstborn	0.384***	0.384***	0.404***	0.372***	0.368***	0.371***
	(0.080)	(0.080)	(0.080)	(0.110)	(0.110)	(0.110)
PGS for years of education	1.227***	1.186***	1.528***	0.635***	0.574***	0.823***
	(0.041)	(0.049)	(0.062)	(0.071)	(0.078)	(0.101)
Firstborn × PGS for years of education		0.108	0.184**		0.162**	0.224**
		(0.073)	(0.093)		(0.081)	(0.100)
Constant	15.353**	15.353**	14.683**	14.293***	14.282***	14.144***
	(1.807)	(1.818)	(1.838)	(2.431)	(2.449)	(4.479)
R^2	0.102	0.103		0.040	0.040	
Cragg-Donald *F*-stat.			4887.542			4068.062
N	15,019	15,019	15,019	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family in the within-family analysis; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

The between-family design shows an economically meaningful and positive interaction between the polygenic score and being firstborn of 0.108 (Column 2). Yet, it is not statistically significant at conventional thresholds. When we use Obviously-Related Instrumental Variables (ORIV) regression (Column 3), the interaction term becomes statistically significant, suggesting that measurement error in the polygenic score attenuates the main effect of the polygenic score and the interaction term. In the family fixed effects specification, we find that the interaction effect is significant in both the OLS and the ORIV specifications. The measurement

error correction makes the interaction term stronger, both statistically and in terms of the effect size (0.162 vs. 0.224).

As a complementary approach to test the statistical significance of our main results, we also conducted randomization inference by employing 10,000 permutations of the polygenic score and birth order, each time re-estimating the within-family specification (Fisher, 1935; Rosenbaum, 2002). This approach does not rely on repeated sampling of a hypothetical population and enables gauging how rare the size and significance of our interaction effect is for other permutations of the polygenic score and birth order. Figure 4 provides the distribution of *t*-statistics for all 10,000 permutations with the *t*-statistic of the actual estimate shown as a solid line. The *t*-statistic of the estimated interaction term has an exact *p*-value (i.e., the proportion of *t*-statistics more extreme than our estimate) of 0.04.

The positive and statistically significant interaction term provides strong evidence for gene-environment interactions in education, and is consistent with the existence of dynamic complementarity in skill formation: the effect of being firstborn (associated with more parental investments) is complementary to a higher value for the polygenic score for years of education. In other words: those with a higher polygenic score benefit more from the increased parental investments associated with being firstborn. The magnitude of the coefficients suggests that for those with a below-average polygenic score, there is no advantage of being firstborn. In contrast, for those with a high polygenic score, being firstborn increases one's years of education. For example, firstborns with a polygenic score two standard deviations above the mean on average enjoy 0.8 additional years of education compared to their laterborn siblings.

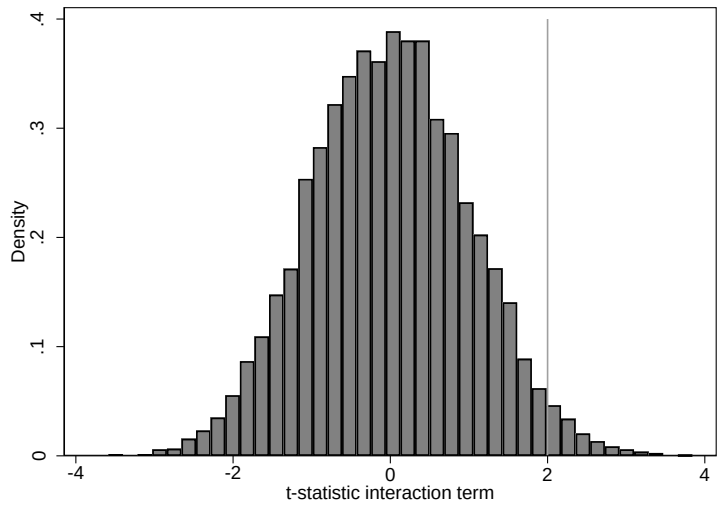

Figure 4. Distribution of the *t*-statistic for the G×E interaction term on basis of randomization inference.

4.5. Robustness checks

In this section we check the robustness of our results against potential non-linearities in the functional form of the polygenic score as well as birth order, and to the addition of further control variables. While the linear form adopted in section 4.4 seems justified by the visual relationship in Figure 1, we explore robustness of our results by allowing for possible non-linearities. Table 6 compares the within-family specification in continuous form (replicated in Column 1 for comparison), with those where we specify the polygenic score in binary form (above and below the mean, Column 2), in quartiles (Column 3) and in squared form (Column 4). We observe positive interaction terms across all specifications. In line with our main findings, the effect of being firstborn is insignificant for those with a lower polygenic score, and the effects are concentrated among those in the upper half of the polygenic score distribution. We also find that the main result in Column 1 is robust to specifying a quadratic in the polygenic score.

Table 6. Results of the regressions of years of education on the gene-environment interaction; Robustness to non-linearities in the polygenic score.

	Within-family analysis			
	(1)	(2)	(3)	(4)
Firstborn	0.368***	0.171	0.198	0.377***
	(0.110)	(0.145)	(0.195)	(0.127)
PGS for years of education	0.574***			0.573***
	(0.078)			(0.078)
Firstborn × PGS for years of education	0.162**			0.164**
	(0.081)			(0.082)
PGS for years of education (>mean)		0.424***		
		(0.137)		
Firstborn × PGS for years of education (>mean)		0.376**		
		(0.175)		
PGS for years of education (2nd quartile)			0.437**	
			(0.185)	
PGS for years of education (3rd quartile)			0.486**	
			(0.192)	
PGS for years of education (4th quartile)			1.133***	
			(0.208)	
PGS for years of education (2nd quartile) × Firstborn			-0.040	
			(0.271)	
PGS for years of education (3rd quartile) × Firstborn			0.457*	
			(0.258)	
PGS for years of education (4th quartile) × Firstborn			0.261	
			(0.240)	
PGS for years of education (squared)				0.032
				(0.047)
PGS for years of education (squared) × Firstborn				-0.008
				(0.062)
Constant	14.282***	14.100***	13.766***	14.227***
	(2.449)	(2.375)	(2.421)	(2.448)
R^2	0.040	0.033	0.046	0.040
N	15,019	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

Table 7 reports the sensitivity of the results to an alternative specification of birth order. Column 1 replicates the results from Table 5 for comparison. In Column 2, we include dummies for each birth rank with firstborns as the reference category. Hence, we expect the effects to be reversed as compared to Column 1. All point estimates of the main effects and the interaction terms are in line with the birth order literature and consistent with dynamic complementarity: on average, those born later have a lower educational attainment compared

to firstborns and benefit less from having a high polygenic score. The estimates for rank 3 and higher are rather imprecisely estimated due to the relatively small sample sizes with a birth rank of 3 or higher, but again in line with our main results.

Table 7. Results of the regressions of years of education on the gene-environment interaction; Robustness to non-linearities in birth order.

	Within-family analysis	
	(1)	(2)
PGS for years of education	0.574***	0.734***
	(0.078)	(0.087)
Firstborn	0.368***	
	(0.110)	
Firstborn × PGS for years of education	0.162**	
	(0.081)	
2nd born		-0.354***
		(0.124)
3rd born		-0.394
		(0.243)
4th born		-0.533
		(0.365)
5th born		-0.097
		(0.485)
2nd born × PGS for years of education		-0.166*
		(0.085)
3rd born × PGS for years of education		-0.188
		(0.142)
4th born × PGS for years of education		-0.010
		(0.247)
5th born × PGS for years of education		-0.118
		(0.294)
Constant	14.282***	14.720***
	(2.449)	(2.454)
R^2	0.040	0.041
N	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

Table 8 presents our final set of robustness checks in which we explore robustness of our results to possible endogenous fertility decisions. A possible correlation between our measure of endowments and birth order could exist when fertility decisions are based on the genetic endowments of the children, known in the literature as the "child stopping rule" (Black et al., 2005; Pavan, 2016). Whereas Section 4.3 shows that such a correlation does not exist, here we

explicitly control for a possible child stopping-rule by including a dummy variable for being lastborn, which is set to one if an individual's birth order is equal to the total number of children in his/her family. Column 1 in Table 8 replicates the results from Table 5 for comparison. As seen in Column 2, the lastborn dummy is not statistically significant, and does not meaningfully affect our results, suggesting that potential endogenous fertility decisions do not change any of our conclusions. In column 3 we employ the correction for missing confounders suggested by (Keller, 2014) for gene-environment interaction analysis. Specifically, we interact both the dummy for being firstborn and the polygenic score for education with year of birth, month of birth, gender, and first 40 principal components and include all these as covariates in the analysis. The direct effect sizes of being firstborn and the polygenic score for years of education are now relative to the reference categories of the control variables (Born in 1937, born in January, being female). Although the standard error of the interaction term increases due to the much larger number of regressors, the magnitude of the interaction effect is robust to this specification.

Table 8. Results of the regressions of years of education on the gene-environment interaction; Robustness to fertility choices and missing confounders.

	Within-family analysis		
	(1)	(2)	(3)
Firstborn	0.368***	0.295**	-7.767*
	(0.110)	(0.129)	(4.401)
PGS for years of education	0.574***	0.577***	-5.981***
	(0.078)	(0.078)	(1.832)
Firstborn × PGS for years of education	0.162**	0.162**	0.166*
	(0.081)	(0.081)	(0.096)
Lastborn		-0.154	
		(0.136)	
Constant	14.282***	14.170***	25.412***
	(2.449)	(2.451)	(3.711)
R^2	0.040	0.040	0.059
N	15,019	15,019	15,019

Notes: Robust standard errors in parentheses, clustered by family; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors.

5. Discussion

A large literature shows consistently higher educational attainments for firstborn children. Using within-family data we move beyond the existing literature by showing that children benefit disproportionally from being firstborn when they have an above-average educational attainment polygenic score. More specifically, firstborns with an average polygenic score enjoy 0.37 years (≈ 4.5 months) of additional schooling compared to their laterborn siblings, on average. However, firstborns with a polygenic score that is one standard deviation above the mean enjoy an *additional* 0.16 years of education, compared to their laterborn siblings with the same genetic endowment. In contrast, for individuals with below-average polygenic scores, being firstborn does not provide an advantage in terms of educational attainment. Since we provide evidence that genetic endowments are orthogonal to birth order, and previous literature suggests that birth order effects on children's education are mainly driven by parental investments, we interpret the positive and significant interaction term as providing support for the existence of the dynamic complementarity in skill formation.

An alternative interpretation of our finding that birth order effects are concentrated among those with higher polygenic scores could be that the additional investments associated with being firstborn are higher for those with higher polygenic scores. That is, if parents would

invest more in the firstborn, or alter fertility decisions, when the child has a higher polygenic score, this could also explain the positive interaction effect. While we cannot fully rule out this explanation, we believe this explanation is less plausible for two reasons. First, Breinholt & Conley (2019) and Houmark et al. (2020) show that parenting during infancy is not driven by genetic make-up because these endowments are not clearly expressed yet, and parental investment responses to polygenic scores do not arise before age 6. This is long after the typical arrival of subsequent children, and so the most precious time of undivided attention for the firstborn is unlikely to be influenced by – at that time unobserved – differences in polygenic scores. Second, for the few early-life parental investments we observe in our data, we do not find evidence of any response to the polygenic score. Appendix D shows that maternal smoking around pregnancy and whether the child was breastfed are all unrelated to the firstborn's polygenic score. If anything, the age gap between first- and secondborns is slightly lower if the firstborn has a higher polygenic score. These findings suggest that the additional investments associated with being firstborn are driven by less restrictive time and budget constraints and are independent of the child's genetic endowment. The appearance of a positive interaction between endowments and being firstborn therefore provides support for the existence of dynamic complementarity.

Finding support for dynamic complementarity is important for understanding the nature of the skill production function. The production function of the child's skills is – next to a parental budget and time constraint – an important input into the broader optimization problem where parents decide between own consumption and investments in their children. By informing the shape and properties of the production function, our analysis is an important precursor to a structural model of parental investment decisions, estimation of which is beyond the scope of this paper (see Houmark et al. 2020 for a recent application that incorporates polygenic scores into a dynamic latent factor model of skill formation). Evidence in support of dynamic complementarity also speaks to whether later-life investments can reduce or eliminate damage originating early in life (Almond et al., 2018), and emphasizes the importance of early-life investments being followed-up by later-life investments to reap the full benefits in terms of human capital outcomes (e.g., Cunha & Heckman, 2007).

More generally, this paper shows how economic theory can inform empirical G×E analyses, and our findings provide one of the first pieces of evidence of how genetic variation (here

measured by the polygenic score for years of education) and environment (here measured by one's birth order) jointly shape and dynamically interact in producing important life outcomes such as years of education. While this finding was long anticipated by numerous scholars (e.g., Heckman, 2007; Rutter et al., 2006), finding credible and independent sources of variation in genes and environments is rare given how tightly genetic and environmental influences are entangled (e.g., Koellinger & Harden, 2018). Showing evidence of an interaction between genetic variation and environments is not just a leap forward in our fundamental understanding of how nature and nurture jointly shape human capital, but also a promising antidote against arguments of genetic (or environmental) determinism.

A number of limitations should be acknowledged. First, our specification may not be the perfect empirical translation of the skill production function. In particular, we do not measure skills directly. Instead, we follow Cunha & Heckman (2008) and Cunha et al. (2010) who specify adult human capital as a combination of skills accumulated by the end of childhood, and employ a commonly used and convenient proxy: years of education. Moreover, we do not measure parental investments directly, and use an environmental variable closely related to parental investments: birth order. The upside of using birth order rather than a direct measure of parental investments is that birth order is randomly assigned within families, whereas parental investments are known to be endogenous to offspring endowments. Moreover, whereas birth order cannot distinguish between early-life and later-life investments, it captures a *persistent* difference across siblings rather than a one-time shock in investments that many other papers rely on (see Almond et al. (2018), and Appendix B). The downside of using birth order is that it is unlikely to capture only parental investments. Other mechanisms through which birth order effects may arise (e.g., interactions with younger siblings) could possibly also interact with genetic endowments. We cannot test this alternative explanation directly because the UK Biobank is very limited in measures of parental and sibling interactions. Still, if we accept that birth order partially captures investment – a premise that should not be controversial given the overwhelming evidence in the literature – then unless these other channels exhibit completely opposite interaction effects, a necessary condition for dynamic complementarity would be a positive interaction between birth order and genetic endowments. This is exactly what we find.

A second limitation is that our measure of genetic endowments is imperfect. In particular, a

polygenic score captures only common genetic variations in the human genome, and even within the realm of common variations the measure is subject to measurement error. While the use of ORIV reduces concerns about classical measurement error, our family fixed effects estimates of the polygenic score are still subject to attenuation bias due to genetic nurture. However, since the sign of the bias arising from genetic nurture is known to be negative, our effect size represents in fact a conservative estimate.

The polygenic score should also not be interpreted narrowly as a measure of immutable biological endowments. While within-family analyses ensure that we can interpret the effect of a polygenic score as a direct (or causal) effect of genetic variation, it is well-established that the environment mediates this effect (e.g., Breinholt & Conley, 2019; Houmark, Ronda, & Rosholm, 2020). Hence, a polygenic score measures education-enhancing endowments, and will reflect how *on average in the discovery sample* environments (including parental investments) respond to differences in genetic endowments. Importantly though, since the measure is fixed at conception and orthogonal to birth order, the measure does not reflect parental investments *of the child's own parents*. As a result, the inclusion of environmental responses to genetic variation into the construction of the polygenic score is not a source of concern for our identification strategy but may affect the interpretation. The definition of dynamic complementarity encompasses both complementarity between investments and initial endowments, as well as complementarity between investments at different ages (e.g., Cunha & Heckman, 2007, 2008). Since a polygenic score reflects both endowments as well as *average* environmental responses to endowments, we cannot distinguish between these channels. However, given the independence of birth-order (and associated investments) from our measure of genetic endowments, we believe our setting does provide a compelling context to test for dynamic complementarity as a putative property of the skill production function.

A final limitation regards the external validity of the empirical findings. As mentioned in the data section, there is sample selection into the UK Biobank, with a bias towards healthier and higher-educated individuals (Fry et al., 2017). On top of this, we focus on European-ancestry individuals and the coincidental sampling of siblings even though these were not specifically targeted, further reducing the representativeness of the sample. Finally, we construct our polygenic score on basis of a tailor-made GWAS, again on basis of the same UK Biobank excluding the siblings and their relatives. While the latter choice helps to maintain the same

environments across discovery and prediction sample, it may further increase the likelihood that our results are specific to the UK Biobank. Future research should replicate our findings, but in light of dynamic complementarity theory undergirding our results, we have good reasons to be positive about the replicability in comparable contexts.

6. References

Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019). Early Life Circumstance and Adult Mental Health. *Journal of Political Economy*, *127*(4), 1516–1549. https://doi.org/10.1086/701606

Adhvaryu, A., & Nyshadham, A. (2016). Endowments at Birth and Parents' Investments in Children. *The Economic Journal*, *126*(593), 781–820. https://doi.org/10.1111/ecoj.12186

Aizer, A., & Cunha, F. (2012). The Production of Human Capital: Endowments, Investments and Fertility. *National Bureau of Economic Research*. https://doi.org/10.3386/w18429

Almond, D., Currie, J., & Duque, V. (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, *56 (4)*, 1360–1446. https://doi.org/10.1257/jel.20171164

Almond, D., & Mazumder, B. (2013). Fetal Origins and Parental Responses. *Annual Review of Economics*, *5*(1), 37–56. https://doi.org/10.1146/annurev-economics-082912-110145

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., … Schloss, J. A. (2015). A Global Feference for Human Genetic Variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Bagger, J., Birchenall, J., Mansour, H., & Urzua, S. (2013). Education, Birth Order, and Family Size. *National Bureau of Economic Research*. https://doi.org/10.3386/w19111

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education Can Reduce Health Differences Related to Genetic Risk of Obesity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(42), E9765–E9772. https://doi.org/10.1073/pnas.1802909115

Barth, D., Papageorge, N. W., & Thom, K. (2020). Genetic Endowments and Wealth Inequality. *Journal of Political Economy*, *128*(4), 1474–1522. https://doi.org/10.1086/705415

Becker, G. S. (1960). Underinvestment in College Education? *The American Economic Review*, *50*(2), 346–354.

Becker, G. S., & Lewis, H. G. (1973). On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy*, *81*(2), S279–S288.

Becker, G. S., & Tomes, N. (1976). Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy*, *84*(4), 143–162.

Becker, G. S., & Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, *4*(3), 1–47. https://doi.org/10.1086/298118

Behrman, J. R. (1988). Nutrition, Health, Birth Order and Seasonality. Intrahousehold Allocation among Children in Rural India. *Journal of Development Economics*, *28*(1), 43–62. https://doi.org/10.1016/0304-3878(88)90013-2

Behrman, J. R., Taubman, P., Behrman, J., & Taubman, P. (1986). Birth Order, Schooling, and

Earnings. *Journal of Labor Economics*, *4*(3), S121-45. Retrieved from https://econpapers.repec.org/RePEc:ucp:jlabec:v:4:y:1986:i:3:p:s121-45

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., … Harris, K. M. (2018). Genetic Analysis of Social-class Mobility in Five Longitudinal Studies. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7275–E7284. https://doi.org/10.1073/pnas.1801238115

Benjamin, D., Cesarini, D., Laibson, D. I., & Turley, P. (n.d.). Social Science Genetics: A Primer and Progress Report. *Journal of Economic Literature*.

Berry, J., Dizon-Ross, R., & Jagnani, M. (2020). Not Playing Favorites: An Experiment on Parental Fairness Preferences. *NBER Working Paper Series*, *26732*. https://doi.org/10.3386/w26732

Bharadwaj, P., Eberhard, J. P., & Neilson, C. A. (2018). Health at Birth, Parental Investments, and Academic Outcomes. *Journal of Labor Economics*, *36*(2), 349–394. https://doi.org/10.1086/695616

Bierut, L., Biroli, P., Galama, T., & Thom, K. (2018). Childhood Socioeconomic Status Moderates Genetic Predisposition for Peak Smoking. *BioRxiv*, 336834. https://doi.org/10.1101/336834

Birdsall, N. (1991). Birth Order Effects and Time Allocation. *Research in Population Economics*, *7*, 191–213. Retrieved from https://europepmc.org/article/med/12317028

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *Quarterly Journal of Economics*, *120*(2), 669–700. https://doi.org/10.1162/0033553053970179

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2016). Healthy(?), Wealthy, and Wise: Birth Order and Adult Health. *Economics and Human Biology*, *23*, 27–45. https://doi.org/10.1016/j.ehb.2016.06.005

Black, S. E., Devereux, P. J., & Salvanesz, K. G. (2011). Older and Wiser? Birth Order and IQ of Young Men. *CESifo Economic Studies*, *57*(1), 103–120. https://doi.org/10.1093/cesifo/ifq022

Black, S. E., Grönqvist, E., & Öckert, B. (2018). Born to Lead? The Effect of Birth Order on Noncognitive Abilities. *Review of Economics and Statistics*, *100*(2), 274–286. https://doi.org/10.1162/rest_a_00690

Booth, A. L., & Kee, H. J. (2009). Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment. *Journal of Population Economics*, *22*(2), 367–397. https://doi.org/10.1007/s00148-007-0181-4

Breinholt, A., & Conley, D. (2019). Child-driven Parenting: Differential Early Childhood Investmnt by Offsrping Genotype. *Working Paper*.

Breining, S., Doyle, J., Figlio, D. N., Karbownik, K., & Roth, J. (2020). Birth Order and

Delinquency: Evidence from Denmark and Florida. *Journal of Labor Economics*, *38*(1), 95–142. https://doi.org/10.1086/704497

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Marchini, J. (2017). Genome-wide Genetic Data on ~500,000 UK Biobank Participants. *BioRxiv*, 166298. https://doi.org/10.1101/166298

Cesarini, D., & Visscher, P. M. (2017). Genetics and Educational Attainment. *Npj Science of Learning*, *2*(1), 1–8. https://doi.org/10.1038/s41539-017-0005-6

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The Fourth Law of Behavior Genetics. *Current Directions in Psychological Science*, *24*(4), 304–312. https://doi.org/10.1177/0963721415580430

Conley, D., & Rauscher, E. (2013). Genetic Interactions with Prenatal Social Environment: Effects on Academic and Behavioral Outcomes. *Journal of Health and Social Behavior*, *54*(1), 109–127. https://doi.org/10.1177/0022146512473758

Correia, S. (2017). Linear Models with High-dimensional Fixed Effects: An Efficient and Feasible Estimator. *Working Paper*.

Correia, S. (2019). REGHDFE: Stata Module to Perform Linear or Instrumental-variable Regression Absorbing Any Number of High-dimensional Fixed Effects. *Statistical Software Components*. Retrieved from https://ideas.repec.org/c/boc/bocode/s457874.html

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

Cunha, F., & Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, *43*(4), 738–782. https://doi.org/10.3368/jhr.43.4.738

Cunha, F., Heckman, J. J., Schennach, S. M., Attanasio, O., Becker, G., Cattan, S., … Schennach, S. M. (2010). Estimationg The Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931. https://doi.org/10.3982/ECTA6551

Damian, R. I., & Roberts, B. W. (2015). The Associations of Birth Order with Personality and Intelligence in a Representative Sample of U.S. High School Students. *Journal of Research in Personality*, *58*, 96–105. https://doi.org/10.1016/j.jrp.2015.05.005

Datar, A., Kilburn, M. R., & Loughran, D. S. (2010). Endowments and Parental Investments in Infancy and Early Childhood. *Demography*, *47*(1), 145–162. https://doi.org/10.1353/dem.0.0092

Davey Smith, G., & Ebrahim, S. (2003). "Mendelian randomization": Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease? *International Journal of Epidemiology*, *32*(1), 1–22. https://doi.org/10.1093/ije/dyg070

De Haan, M. (2010). Birth Order, Family Size and Educational Attainment. *Economics of Education Review*, *29*(4), 576–588. Retrieved from

https://econpapers.repec.org/RePEc:eee:ecoedu:v:29:y:2010:i:4:p:576-588

De Haan, M., Plug, E., & Rosero, J. (2014). Birth Order and Human Capital Development: Evidence from Ecuador. *Journal of Human Resources*, *49*(2), 359–392. https://doi.org/10.3368/jhr.49.2.359

Del Boca, D., Flinn, C., & Wiswall, M. (2014). Household Choices and Child Development. *Review of Economic Studies*, *81*(1), 137–185. https://doi.org/10.1093/restud/rdt026

DiPrete, T. A., Burik, C. A. P., & Koellinger, P. D. (2018). Genetic Instrumental Variable Regression: Explaining Socioeconomic and Health Outcomes in Nonexperimental Data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(22), 4970–4979. https://doi.org/10.1073/pnas.1707388115

Domingue, B., Trejo, S., Armstrong-Carter, E., & Tucker-Drob, E. (2020). *Interactions Between Polygenic Scores and Environments: Methodological and Conceptual Challenges*. https://doi.org/10.31235/osf.io/u7sh4

Domingue, B. W., Liu, H., Okbay, A., & Belsky, D. W. (2017). Genetic Heterogeneity in Depressive Symptoms Following the Death of a Spouse: Polygenic Score Analysis of the U.S. Health and Retirement Study. *American Journal of Psychiatry*, *174*(10), 963–970. https://doi.org/10.1176/appi.ajp.2017.16111209

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, *9*(3), 1003348. https://doi.org/10.1371/journal.pgen.1003348

Duque, V., Rosales-Rueda, M., Sanchez, F., Currie, J., Almond, D., Bailey, M., … Shah, M. (2018). How Do Early-Life Shocks Interact with Subsequent Human-Capital Investments? Evidence from Administrative Data. *IZA World of Labor Conference.*

Eirnæs, M., & Pörtner, C. C. (2004). Birth order and the intrahousehold allocation of time and education. *Review of Economics and Statistics*, *86*(4), 1008–1019. https://doi.org/10.1162/0034653043125176

Elam, K. K., Clifford, S., Shaw, D. S., Wilson, M. N., & Lemery-Chalfant, K. (2019). Gene Set Enrichment Analysis to Create Polygenic Scores: a Developmental Examination of Aggression. *Translational Psychiatry*, *9*(1), 212. https://doi.org/10.1038/s41398-019-0513-7

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., … Tress, M. L. (2014). Multiple Evidence Strands Suggest that There May be as Few as 19 000 Human Protein-coding Genes. *Human Molecular Genetics*, *23*(22), 5866–5878. https://doi.org/10.1093/hmg/ddu309

Figlio, D., Guryan, J., Karbownik, K., & Roth, J. (2014). The Effects of Poor Neonatal Health on Children's Cognitive Development? *American Economic Review*, *104*(12), 4205–4230. https://doi.org/10.1257/aer.104.12.3921

Fisher, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver & Boyd.

Fletcher, J., Wu, Y., Zhao, Z., & Lu, Q. (2020). The Production of Within-family Inequality: Insights and Implications of Integrating Genetic Data. *BioRxiv*, 2020.06.06.137778. https://doi.org/10.1101/2020.06.06.137778

Frijters, P., Johnston, D. W., Shah, M., & Shields, M. A. (2013). Intrahousehold Resource Allocation: Do Parents Reduce or Reinforce Child Ability Gaps? *Demography*, *50*(6), 2187–2208. https://doi.org/10.1007/s13524-013-0224-2

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *American Journal of Epidemiology*, *186*(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Galor, O., & Weil, D. N. (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, *90*(4), 806–828. https://doi.org/10.1257/aer.90.4.806

Giannola, M. (2020). Parental Investment, and Intra-household Inequality in Child Development: Theory, Measurement and Evidence from a Lab-in-the-field Experiment. *Mimeo, University College London.*

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy*, *127*(4), 1826–1863. https://doi.org/10.1086/701681

Gunnsteinsson, S., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Ahmed, A., … Venkataramani, A. (2014). *Resilience to Early Life Shocks Evidence from the Interaction of a Randomized Controlled Trial and a Natural Experiment.* Retrieved from http://sites.bu.edu/neudc/files/2014/10/paper_47.pdf

Handy, C., & Shester, K. (2020). The Effect of Birth Order on Educational Attainment among the Baby Boom Generation. *MPRA Paper No. 102426*.

Havari, E., & Savegnago, M. (2020). The Intergenerational Effects of Birth Order on Education. *Journal of Population Economics*, 1–29. https://doi.org/10.1007/s00148-020-00810-5

Heckman, J. J. (2007). The Economics, Technology, and Neuroscience of Human Capability Formation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13250–13255. https://doi.org/10.1073/pnas.0701362104

Heiland, F. (2009). Does the Birth Order Affect the Cognitive Development of a Child? *Applied Economics*, *41*(14), 1799–1818. https://doi.org/10.1080/00036840601083220

Hotz, V. J., & Pantano, J. (2015). Strategic Parenting, Birth Order, and School Performance. *Journal of Population Economics*, *28*(4), 911–936. https://doi.org/10.1007/s00148-015-0542-3

Houmark, M. A., Ronda, V., & Rosholm, M. (2020). The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills. *IZA Discussion Paper*, (13780).

Hsin, A., & Felfe, C. (2014). When Does Time Matter? Maternal Employment, Children's Time With Parents, and Child Development. *Demography*, *51*(5), 1867–1894. https://doi.org/10.1007/s13524-014-0334-5

Jencks, C. (1980). Heredity, Environment, and Public Policy Reconsidered. *American Sociological Review*, *45*(5), 723–736. https://doi.org/10.2307/2094892

Jiang, L., Zheng, Z., Qi, T., Kemper, K. E., Wray, N. R., Visscher, P. M., & Yang, J. (2019). A Resource-efficient Tool for Mixed Model Association Analysis of Large-scale Data. *Nature Genetics*, *51*(12), 1749–1755. https://doi.org/10.1038/s41588-019-0530-8

Johnson, R. C., & Jackson, C. K. (2019). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy*, *11*(4), 310–349. https://doi.org/10.1257/pol.20180510

Kantarevic, J., Mechoulan, S., Kantarevic, J., & Mechoulan, S. (2006). Birth Order, Educational Attainment, and Earnings: An Investigation Using the PSID. *Journal of Human Resources*, *41*(4). Retrieved from https://econpapers.repec.org/RePEc:uwp:jhriss:v:41:y:2006:i:4:p755-777

Keller, M. C. (2014). Gene × environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biological Psychiatry*, *75*(1), 18–24. https://doi.org/10.1016/j.biopsych.2013.09.006

Koellinger, P. D., & Harden, K. P. (2018). Using Nature to Understand Nurture: Genetic Associations Show How Parenting Matters for Children's Education. *Science*, *359*(6374), 386–387. https://doi.org/10.1126/science.aar6429

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., … Stefansson, K. (2018). The Nature of Nurture: Effects of Parental Genotypes. *Science*, *359*(6374), 424–428. https://doi.org/10.1126/science.aan6877

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Davey Smith, G. (2008). Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology. *Statistics in Medicine*, *27*(8), 1133–1163. https://doi.org/10.1002/sim.3034

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Turley, P. (2018). Gene Discovery and Polygenic Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lehmann, J. Y. K., Nuevo-Chiquero, A., & Vidal-Fernandez, M. (2018). The Early Origins of Birth Order Differences in Children's Outcomes and Parental Behavior. *Journal of Human Resources*, *53*(1), 123–156. https://doi.org/10.3368/jhr.53.1.0816-8177

Lubotsky, D., & Kaestner, R. (2016). Do `Skills Beget Skills'? Evidence on the Effect of Kindergarten Entrance Age on the Evolution of Cognitive and Non-cognitive Skill Gaps in Childhood. *Economics of Education Review*, *53*, 194–206. https://doi.org/10.1016/j.econedurev.2016.04.001

Malamud, O., Pop-Eleches, C., Urquiola, M., Malamud, O., Pop-Eleches, C., & Urquiola, M. (2016). Interactions Between Family and School Environments: Evidence on Dynamic Complementarities? *NBER Working Paper Series*, *22112*. Retrieved from https://econpapers.repec.org/RePEc:nbr:nberwo:22112

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. M. (2010). Robust Relationship Inference in Genome-wide Association Studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Mills, M. C., Barban, N., & Tropf, F. C. (2020). An Introduction to Statistical Genetic Data Analysis. *Cambridge, MA: MIT Press.*

Monfardini, C., & See, S. G. (2012). Birth Order and Child Outcomes: Does Maternal Quality Time Matter? *CHILD Working Papers Series*. Retrieved from https://ideas.repec.org/p/cca/wchild/3.html

Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., … Lewis, C. M. (2016). Polygenic Interactions with Environmental Adversity in the Aetiology of Major Depressive Disorder. *Psychological Medicine*, *46*(4), 759–770. https://doi.org/10.1017/S0033291715002172

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Benjamin, D. J. (2016). Genome-wide Association Study Identifies 74 Loci Associated with Educational Attainment. *Nature*, *533*(7604), 539–542. https://doi.org/10.1038/nature17671

Pavan, R. (2016). On the Production of Skills and the Birth-order Effect. *Journal of Human Resources*, *51*(3), 699–726. https://doi.org/10.3368/jhr.51.3.0913-5920R

Pereira, R. D., van Kippersluis, H., & Rietveld, C. A. (2020). The Interplay between Maternal Smoking and Genes in Offspring Birth Weight. *MedRxiv*.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. J. H. (2014). Effect of Polygenic Risk Scores on Depression in Childhood Trauma. *British Journal of Psychiatry*, *205*(2), 113–119. https://doi.org/10.1192/bjp.bp.113.143081

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal Components Analysis Corrects for Stratification in Genome-wide Association Studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Price, J. (2008). Parent-child Quality Time: Does Birth Order Matter? *Journal of Human Resources*, *43*(1), 240–265. https://doi.org/10.3368/jhr.43.1.240

Pruckner, G. J., Schneeweis, N., Schober, T., Zweimüller, M., Pruckner, G. J., Schneeweis, N., … Zweimüller, M. (2019). Birth Order, Parental Health Investment, and Health in Childhood. *IZA Discussion Paper No. 12774*, (August). Retrieved from www.iza.org

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A Tool Set for Whole-genome Association and Population-based Linkage

Analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A. E., … Koellinger, P. D. (2014). Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits. *Psychological Science*, *25*(11), 1975–1986. https://doi.org/10.1177/0956797614545132

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., … Koellinger, P. D. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*, *340*(6139), 1467–1471. https://doi.org/10.1126/science.1235488

Ronda, V., Agerbo, E., Bleses, D., Mortensen, P. B., Borglum, A., Hougaard, D. M., … Rosholm, M. (2020). Family Disadvantage, Gender and the Returns to Genetic Human Capital. *IZA Discussion Paper*, *13441*.

Rosenbaum, P. R. (2002). *Observational Studies*. https://doi.org/10.1007/978-1-4757-3692-2

Rossin-Slater, M., & Wüst, M. (2020). What is the Added Value of Preschool for Poor Children? Long-term and Intergenerational Impacts and Interactions with an Infant Health Intervention. *American Economic Journal: Applied Economics*, *12*(3), 255–286. https://doi.org/10.1257/app.20180698

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene-environment Interplay and Psychopathology: Multiple Varieties but Real Effects. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(3–4), 226–261. https://doi.org/10.1111/j.1469-7610.2005.01557.x

Sanz-de-galdeano, A., & Terskaya, A. (2019). Sibling Differences in Educational Polygenic Scores : How Do Parents React ? *IZA DISCUSSION PAPER SERIES*, (12375), 1–33.

Schmitz, L., & Conley, D. (2016). The Long-Term Consequences of Vietnam-Era Conscription and Genotype on Smoking Behavior and Health. *Behavior Genetics*, *46*(1), 43–58. https://doi.org/10.1007/s10519-015-9739-1

Schmitz, L., & Conley, D. (2017). Modeling Gene-Environment Interactions With Quasi-Natural Experiments. *Journal of Personality*, *85*(1), 10–21. https://doi.org/10.1111/jopy.12227

Schmitz, L. L., & Conley, D. (2017). The effect of Vietnam-era Conscription and Genetic Potential for Educational Attainment on Schooling Outcomes. *Economics of Education Review*, *61*(October), 85–97. https://doi.org/10.1016/j.econedurev.2017.10.001

Selzam, S., Ritchie, S. J., Pingault, J. B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing Within- and Between-Family Polygenic Score Prediction. *American Journal of Human Genetics*, *105*(2), 351–363. https://doi.org/10.1016/j.ajhg.2019.06.006

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, *113*(485), F3–F33.

Trejo, S., & Domingue, B. (2019). Genetic Nature or Genetic Nurture? Quantifying Bias in Analyses Using Polygenic Scores. *BioRxiv*, 524850. https://doi.org/10.1101/524850

Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., Van Der Most, P. J., De Vlaming, R., … Mills, M. C. (2017). Hidden Heritability due to Heterogeneity Across Seven Populations. *Nature Human Behaviour*, *1*(10), 757–765. https://doi.org/10.1038/s41562-017-0195-1

van Kippersluis, H., Biroli, P., Galama, T., von Hinke, Stephanie Meddens, S. F. W., Muslimova, D., Pereira, R., & Rietveld, C. A. (2020). Stop Meta-Analyzing, Start Instrumenting: Reducing Measurement Error in Polygenic Scores. *Working Paper*.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., … Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Human Genetics*, *97*(4), 576–592. https://doi.org/10.1016/j.ajhg.2015.09.001

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and Efficient Meta-analysis of Genomewide Association Scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Winkler, T. W., Day, F. R., Croteau-Chonka, D. C., Wood, A. R., Locke, A. E., Mägi, R., … Loos, R. J. F. (2014). Quality Control and Conduct of Genome-wide Association Meta-analyses. *Nature Protocols*, *9*(5), 1192–1212. https://doi.org/10.1038/nprot.2014.071

7. Appendix

A. Genetic data, GWAS, and Polygenic scores

Genetic Data. A complete human genome consists of 23 pairs of chromosomes, from which the 23rd pair determines the biological sex of a person. One of each pair of chromosomes is inherited from the father, and the other is inherited from the mother. A chromosome is composed of two intertwined strands of deoxyribonucleic acid (DNA), each made up of a sequence of four possible nucleotide molecules: adenine, cytosine, thymine, and guanine. Adenine (A) on one strand is always paired with thymine (T) on the other strand, and cytosine (C) is always paired with guanine (G). These pairs are called base pairs. Every human genome consists of approximately 3 billion base pairs and stretches of base pairs coding for proteins are called genes. There are approximately 20,000 genes in the human genome, with varying lengths in terms of base pairs (Ezkurdia et al., 2014).

Two unrelated human beings share approximately 99.6% of their DNA, and most genetic differences across humans can be attributed to single nucleotide polymorphisms (SNP) (Auton et al., 2015). A SNP is a locus in the DNA at which two different nucleotides can be observed in the population. Each of the two possible nucleotides is called an allele for that SNP. An individual's genotype is coded as 0, 1, or 2, depending on the number of "effect" alleles present. In the human genome, there are at least 85 million SNPs with a "minor" allele prevalence of at least 1% (Auton et al., 2015).

Genome-Wide Association Studies (GWASs) aim to identify genetic variants that are associated with a particular trait of interest by relating each variant to the trait in a hypothesis-free approach. Stringent significance thresholds are used to identify variants that are robustly associated with the trait, with other independent samples used for replication. Using the GWAS approach, thousands of genetic discoveries have been made (Visscher et al., 2017).

Individual SNPs typically explain less than 0.02% of the variance in a behavioural outcome (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015; Visscher et al., 2017). It is therefore common to combine multiple SNPs into a polygenic score (Dudbridge, 2013), constructed as a weighted sum of SNPs.

Through increases in GWAS sample sizes, the predictive power of the polygenic score for education has increased from 2-3% (Rietveld et al., 2013), to 6-8% (Okbay et al., 2016), to currently 11-13% (Lee et al., 2018). In terms of biological pathways, there is evidence that many

of the identified genes associate with health, cognitive, and central nervous system traits (Rietveld et al., 2013). Likewise, the majority of the significant SNPs in Okbay et al. (2016) and Lee et al. (2018) relate to genomic regions responsible for gene expression in a child's brain during the prenatal period.

Methods. *Relatedness.* As a first step, we identify siblings and their relatives using the kinship matrix provided by the UK Biobank. The kinship matrix is based on genetically identified relatedness and contains relatives of third degree and closer identified using the KING software (Manichaikul et al., 2010). The UK Biobank does not have information about self-reported relatedness (Bycroft et al., 2017). The degree of relatedness between the pairs of individuals is based on the combination of the kinship coefficient and genetic similarity in terms of the identity by state (IBS_0) coefficient. IBS_0 measures the fraction of markers for which the related individuals do not share alleles. We follow the KING manual regarding the thresholds for how to determine family relationship (see Table A.1). The identified number of pairs per relationship type differs slightly from that of Bycroft et al. (2017), because some UK Biobank participants withdrew their consent to analyse their data since then.

Table A.1: Thresholds used to determine relatedness between individuals in the UK Biobank.

	Duplicate / Monozygotic twins	1st degree / Parent-child	1st degree siblings	2nd -3rd degree relatives / cousins	Total
Kinship coefficient	>0.3540	0.1770–0.3540	0.1770–0.3540	0.0442–0.1770	
IBS_0		<0.0012	>0.0012		
N (pairs)	179	6,271	22,659	78,038	107,147

For our analyses, we go one step further by separating those who are related to the siblings up to the 3rd degree (kinship coefficient ≥ 0.025), i.e., siblings, parents of siblings, cousins of siblings (See Table A.2). In this way, our holdout sample for polygenic score construction and prediction (i.e., the sibling subsample) is unrelated to the GWAS discovery sample which is used to calibrate the SNP weights that are used to construct the polygenic score.

Table A.2: Relatedness to the individuals in the siblings' subsample of UK Biobank.

Relationship to siblings	Unrelated to siblings	Full siblings	2nd-3rd relative of siblings	Parent or child of siblings	Total
N (individuals)	91,055	41,498	10,207	4,740	147,500

Notes: Relatedness to siblings is computed based on the relatedness classification as reported in Table A.1.

GWAS. Our tailor-made GWAS is performed using the fastGWA tool for Genome-wide Complex Trait Analysis (GCTA) developed by Jiang et al. (2019). fastGWA applies mixed linear modelling (MLM) to the genetic data of the UK Biobank. fastGWA requires the following steps. First, we generate a sparse genetic relatedness matrix (GRM) using the family relatedness file from the UK Biobank based on the KING software output. Next, we perform an MLM-based GWAS using the SNP data, the sparse GRM, the phenotype file and the minor allele frequency (MAF) filter of 0.001. The phenotype file provides the data on individual years of education residualised with respect to birth year, gender, interaction of birth year and gender, batch, and the first 40 principal components (PCs). We also perform quality control.[15] The eventual GWAS discovery sample includes 392,771 individuals: 181,459 males and 211,312 females.

We further quality control the resulting GWAS summary statistics using EasyQC tool (Winkler et al., 2014) and meta-analyse our tailor-made GWAS weights with the summary statistics from Okbay et al. (2016). We use these for constructing an alternative polygenic score that is used in the robustness analysis (see footnote 13). Meta-analysis is conducted using the software package METAL (Willer, Li, & Abecasis, 2010).

Polygenic scores. The polygenic scores are constructed while accounting for linkage disequilibrium between SNPs using LDpred (Vilhjálmsson et al., 2015), version 1.06, and Python, version 3.6.6. Linkage disequilibrium pertains to the non-random correlations between SNPs at various loci of a single chromosome. LDpred is a software package based on Python that adjusts the GWAS weights for LD using a Bayesian approach. We follow the steps as outlined in Mills, Barban, & Tropf (2020), including the coordination of the base and target files, computing the LD adjusted weights, and then applying them for polygenic score

[15] More specifically, we exclude individuals who withdrew consent, have missing gender or whose self-reported gender does not match the genetically identified, are of other than European ancestry, have bad genotyping quality, putative sex chromosome aneuploidy, whose second chromosome karyotypes are different from XX or XY, with outliers in heterozygosity, or have missing information on any of the former criteria.

construction using PLINK (Purcell et al., 2007). We re-weight the SNP effects on the basis of LD and the supposed fraction of causal SNPs, which we set to 1, as is standard practice for behavioural traits (Cesarini & Visscher, 2017). Our hold-out sample for constructing polygenic scores consists of 49,866 siblings and their relatives, where the final analysis sample with observations for all variables available is 15,019 individual siblings. The polygenic scores include all SNPs, that is 1,065,078 SNPs after filtering for HapMap3 SNPs at the coordination step. For the split sample GWAS, we split the clean discovery sample (*N*=392,771) randomly into two samples of ~196,380 individuals each and use the same FastGWA procedure as for the full UKB GWAS to obtain SNP weights. We proceed by using LDpred to construct two polygenic scores based on the two sets of summary statistics. Likewise, we include all SNPs (1,065,146 after filtering for HapMap SNPs at the coordination step).

B. Dynamic complementarity

Testing the concept of dynamic complementarity is challenging, since it requires independent variation in initial endowments and later-life investments, or alternatively exogenous variation in multiple investments over time (Almond & Mazumder, 2013; Johnson & Jackson, 2019) Cunha & Heckman (2007) and Cunha et al. (2010) adopt a structural approach, modelling both skills as well as parental investments as low-dimensional latent variables, and find evidence consistent with dynamic complementarity. A number of studies have examined whether the effect of specific interventions or investments varies by initial skills. Aizer & Cunha (2012) correct early life health measures for certain prenatal investments, and find that pre-school enrolment is more productive for children with higher levels of this residualised measure of endowments. Lubotsky & Kaestner (2016) use entrance-age in kindergarten as plausibly exogenous variation in initial cognitive skills, and find some evidence for dynamic complementarity, although the effect dies out after the first grade.

A recent set of papers have examined rare cases where there exists exogenous variation in both initial endowments as well as later-life investments. For example, Malamud et al. (2016) study the interaction between exogenous variation in access to better schools and variation in family backgrounds induced by access to abortion in Romania. Their findings do not suggest a meaningful interaction between initial endowments and later-life investments. Rossin-Slater & Wüst (2020) exploit a nurse home visiting program as an exogenous shock to endowments, and staggered access to high quality preschool childcare in Denmark as an exogenous shock to investment, and find that these interventions are substitutes rather than complements. Gunnsteinsson et al. (2014) exploit a unique combination where a tornado struck an area of Bangladesh that was coincidentally involved in a randomized experiment on vitamin A supplementation. Their findings are consistent with dynamic complementarity since children treated with Vitamin A supplements were better protected from the consequences of the earthquake. Adhvaryu et al. (2019) exploit local rainfall in the year of birth as exogenous variation in endowments, and randomized cash incentives from Progresa as an exogenous shock to investment. Their main finding is that children from families who received cash transfers were protected better against adverse endowments, consistent with dynamic complementarity. Similarly, Duque et al. (2018) also use a combination of adverse weather shocks and conditional cash transfers in Colombia to show that children born under normal weather conditions benefit more from the cash transfers. Finally, Johnson & Jackson (2019)

exploit the rollout of Head Start and the implementation of court-ordered school finance reforms (SFRs) that increased spending at public K-12 schools as two exogenous shocks to human capital investment, again finding evidence in favour of dynamic complementarity.

The papers above therefore suggest that an increasing number of papers explore variation in both initial endowments and later-life investments to investigate dynamic complementarity. However, the evidence from this literature is mixed, with some studies supporting the existence of dynamic complementarity, and others finding weaker evidence or even substitutability between endowments and investments.

C. Obviously-Related Instrumental Variable (ORIV) regression

In this section, we explain the technique of Obviously-Related Instrumental Variable (ORIV; Gillen et al., 2019) regression. Suppose we would like to predict an outcome variable of interest, *Y*, using a polygenic score, i.e., estimate the following model:

$$Y = \alpha + \beta PGS^* + \varepsilon, \quad \textit{(C.1)}$$

where α is a constant, β is the effect of a true polygenic score PGS^* and ε is the error term. We have two estimates of the true polygenic score: $PGS_1 = PGS^* + \vartheta_1$ and $PGS_2 = PGS^* + \vartheta_2$. The covariance between the two is zero, $Cov(\vartheta_1, \vartheta_2) = 0$ and they have the same relative variance of the measurement errors ϑ_1, ϑ_2. That is:

$$\frac{\sigma^2_{\vartheta_1}}{\sigma^2_{PGS_1}} = \frac{\sigma^2_{\vartheta_2}}{\sigma^2_{PGS_2}} = \frac{\sigma^2_{\vartheta}}{\sigma^2_{PGS}}, \quad \textit{(C.2)}$$

where $\sigma^2_{\vartheta_1}$ and $\sigma^2_{\vartheta_2}$ are the variances of the measurement errors ϑ_1, ϑ_2 respectively, and $\sigma^2_{PGS_1}$ and $\sigma^2_{PGS_2}$ are the variances of respective polygenic scores. If we use PGS_2 as an instrumental variable for PGS_1, the following applies:

$$\hat{\beta}_{IV} = \frac{Cov(Y, PGS_2)/V(PGS_2)}{Cov(PGS_1, PGS_2)/V(PGS_2)} = \frac{Cov(\alpha + \beta PGS^* + \varepsilon, PGS^* + \vartheta_2)}{Cov(PGS^* + \vartheta_1, PGS^* + \vartheta_2)} = \beta \frac{\sigma^2_{PGS^*}}{\sigma^2_{PGS^*}} = \beta. \quad \textit{(C.3)}$$

ORIV regression as developed by Gillen et al. (2019) estimates a 'stacked' model:

$$\begin{pmatrix} Y \\ Y \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \beta \begin{pmatrix} PGS_{1+} \\ PGS_{2+} \end{pmatrix} + \varepsilon, \quad \textit{(C.4)}$$

where one instruments the stack of estimated polygenic scores $\begin{pmatrix} PGS_{1+} \\ PGS_{2+} \end{pmatrix}$ with $\begin{pmatrix} PGS_{2+} & 0_N \\ 0_N & PGS_{1+} \end{pmatrix}$, where *N* is the sample size and 0_N is a *N*x1 vector with zero's. We include a family-stack fixed effect to conduct the within-family comparisons within a stack of the data. Standard errors are clustered at both the family and individual level following Correia (2017, 2019).

D. Early-life parental investments

The only early-life parental investments that are observed in the UK Biobank are whether the child was breastfed and whether the mother smoked around birth. We also observe the age gap between subsequent siblings. Table D.1 reports the results of regressions explaining the few early-life parental investment as a function of being firstborn, the polygenic score for education, and their interaction. This shows whether mothers change their behaviour depending on whether it is their first- or laterborn child and the polygenic score of their children. The results show that the probability of being breastfed (Column 1) and the likelihood of maternal smoking around pregnancy (Column 2) are similar between first- and laterborns. Furthermore, the polygenic score for (the child's) education is insignificantly different from zero, and we find no evidence of any differences in maternal investments around pregnancy by the firstborn's polygenic score. Finally, Column 3 presents the estimates from a regression of the age gap between the first two siblings on the polygenic score of the first sibling, suggesting that the age gap is slightly lower (0.08 years) if the firstborn has a higher polygenic score for education. Note, however, that this latter column compares families with different age gaps and is therefore a between- rather than within-family analysis. Overall, however, we do not find evidence for any meaningful investment responses to the polygenic score of the firstborn, and if anything the time of undivided attention for firstborns is marginally *smaller* if the child has a higher polygenic score.

Table D.1. Results of regressions of early life parental investments on the gene-environment interaction.

	Within-family analyses		Between-family analysis
	Breastfed	Mother smoked around birth	Age gap
	(1)	(2)	(4)
Firstborn	0.007	0.006	
	(0.010)	(0.006)	
PGS for years of education	-0.004	0.001	-0.082**
	(0.007)	(0.004)	(0.037)
Firstborn × PGS for years of education	-0.010	0.003	
	(0.007)	(0.005)	
Constant	0.695***	0.266***	3.509***
	(0.067)	(0.040)	(0.906)
R^2	0.040	0.019	0.096
N	11,929	13,303	4,003

Notes: Robust standard errors in parentheses, clustered by family in the within-family analysis of columns 1-3; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; Coefficients for the control variables (year and month of birth, gender and the first 40 principal components) are not displayed, but available upon request from the authors. Sample sizes vary depending on the availability of early life parental investments.