I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17899

The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students' Genetic Differences

Rosa Cheesman
Nicolai T. Borgen
Astrid M. J. Sandsør
Paul Hufe

MAY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17899

The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students' Genetic Differences

Rosa Cheesman
University of Oslo and Norwegian Institute of Public Health

Nicolai T. Borgen
University of Oslo

Astrid M. J. Sandsør
University of Oslo and IZA

Paul Hufe
University of Bristol and IZA

MAY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students' Genetic Differences*

We investigate whether better schools can compensate for the effects of children's genetic differences. To this end, we combine data from the Norwegian Mother, Father, and Child Cohort Study (MoBa) with Norwegian register data to estimate the interaction between genetic endowments and school quality. We use MoBa's genetic data to compute polygenic indices for educational attainment (PGIEA). Importantly, MoBa includes information on the genetic endowments of father-mother-child trios, allowing us to identify causal genetic effects using within-family variation. We calculate school value-added measures from Norwegian register data, allowing us to causally estimate school quality effects. Leveraging the advantages of both data sources, we provide the first causally identified study of gene-environment interactions in the school context. We find evidence for substitutability of PGIEA and school quality in reading but not numeracy: a 1 SD increase of school quality decreases the impact of the PGIEA on reading test scores by 6%. The substitutability arises through gains of students at the lower end of the PGIEA distribution. This shows that investments in school quality may help students to overcome their draw in the genetic lottery.

Keywords: education, gene-environment interaction, polygenic index, school value-added

Corresponding author:
Rosa Cheesman
The PROMENTA Center
Department of Psychology
University of Oslo
P.O. box 1094, Blindern
0317 Oslo
Norway
E-mail: r.c.g.cheesman@psykologi.uio.no

* N.T.B., R.C., P.H., and A.M.J.S. designed research; N.T.B., R.C., P.H., and A.M.J.S. performed research; N.T.B., R.C., and A.M.J.S. analyzed data; and N.T.B., R.C., P.H., and A.M.J.S. wrote the paper. The authors declare no competing interests.

Introduction

A core topic in the social sciences is the challenge articulated in the Coleman report (1): "the question of how well schools reduce the inequity of birth" (p.36). A substantial body of literature has explored the extent to which schools fulfill their purpose of compensating for the effects of background differences among children that lie beyond their control. Traditionally, this research has focused on inequalities in school performance by socioeconomic status (SES), race, and gender (2–4). Here, we consider whether schools compensate or reinforce the effects of children's genetic differences.

Genetic differences between children play a significant role in skills development, with the twin study heritability for childhood school performance estimated at approximately 50% (5, 6). Importantly, these genetic influences interact with social influences on academic skills, such as family socioeconomic status (7) and the broader socio-political environment (8). Evidence of gene-environment interactions is highly relevant to social policy; it underscores the fact that environments do not affect all individuals equally. Interventions that counteract genetic influences could serve as vital policy levers for fostering educational equity.

Despite this, empirical evidence on gene-environment interactions within schools remains sparse. Traditional behavioral genetic study samples limit research on genetic interactions with school environments because twins usually attend the same school. The recent availability of polygenic indices (PGIs) has catalyzed research into how genetic factors interplay with school environments. Prior studies have examined between-school variation in PGI effects using multilevel models (9–11) as well as measured school quality indicators, such as teacher experience and class size (12). While one set of results was inconclusive (9), the later studies converged on the finding that higher-quality schools compensate for differences in genetic endowment in the production of educational attainment (10–12).

However, a primary challenge in existing research is the difficulty of identifying exogenous variation in both genetic factors and school quality measures. Estimates of gene-environment interactions (GxE) are biased when genetic factors (G) and/or the environment of interest (E) correlate with other variables that influence outcomes. An effective approach to address the first source of bias is by using within-family PGIs. Within-family variation in PGIs is a random "genetic lottery," meaning that effects cannot be confounded by other family characteristics. This approach was first adopted in the context of gene-environment interactions within schools (10). However, the study lacked a method for capturing exogenous variation in school quality and was limited in capturing only the school environments of children participating in the cohort study. Without identifying causal effects, we cannot conclude whether social advantages increase the expression of genetic differences or instead support compensation and diathesis stress models of child development, whereby social advantages buffer genetic influences.

Here, we incorporate exogenous variation in both genetics and environments. We leverage within-family PGIs through parent-offspring genetic data in the Norwegian Mother Father Child Cohort study (MoBa) and use established school value-added (VA) indicators, integrating data from

Norwegian registers, thus encompassing all Norwegian children in the cohort (13, 14). Focusing on national tests provides methodological advantages due to their standardized nature and real-world relevance to students' life outcomes. The Norwegian context and rich educational data provide the ideal setting for validating our novel causal gene-environment interaction approach.

Results

Analysis sample

Our analysis is based on the Norwegian Mother, Father, and Child Cohort Study (MoBa) (15), linked to population-wide administrative records containing students' national standardized test scores and school identifiers. The outcomes of interest are students' reading and numeracy test scores in grade 9, which capture students' skills in these domains one year after starting lower secondary school (grades 8-10, see SI Appendix, sections A and B, for details on the schooling system and national test scores in Norway). Students' test scores are standardized within each year in the full population of Norwegian students to have a mean of zero and a standard deviation of one. We measure students' and their parents' genetic endowments with a polygenic index for educational attainment PGI^{EA} and quantify school quality through school value-added measures (VA^{d} with $d \in$ {reading, numeracy}). VA^{d} is constructed using the full population of Norwegian students. We standardize PGI^{EA} and VA^{d} in our estimation sample to have a mean of zero and a standard deviation of one.

Table 1 shows descriptive statistics for our analysis sample comprising 30,939 children with complete data on all relevant outcomes, treatment, and control variables. While our analysis sample of genotyped parent-child trios is comparable to the overall MoBa sample, MoBa participants are positively selected on socioeconomic background characteristics relative to the general population. Importantly, while our study sample is positively selected and therefore not representative of the overall Norwegian population, this selection does not compromise the internal validity of our estimates or their causal interpretation..

Validation of identification assumptions

I. Exogenous variation in polygenic indices (PGI^{EA})

Causal identification of gene-environment interactions requires exogenous variation in both genetic factors and environmental exposures (16). In the absence of exogenous variation in children's genetic endowments (PGI^{EA}), estimates of genetic effects and the corresponding gene-environment interaction will be confounded by indirect genetic effects from parents and population stratification (17). We achieve exogenous variation in PGI^{EA} by leveraging the availability of genetic trios in MoBa. By analyzing children's PGI^{EA} while controlling for maternal and paternal PGI^{EA}, we isolate the component of children's genetic variation that is randomly allocated during meiosis. This within-family genetic variation approach enables causal identification of genetic effects (18).

Figure 1 provides evidence supporting the exogeneity of the within-family PGI^{EA} variation used in our study. Specifically, it shows correlations between students' PGI^{EA} and family characteristics that may influence the educational outcomes of children, such as parents' education and income. The dark-gray dots show correlations without controls for parental PGI^{EA}. Many of the correlations are positive and significantly different from zero — a pattern consistent with indirect parental genetic effects. The light-gray dots show the same correlations after controlling for parental PGI^{EA}. Notably, all correlations with family characteristics converge to zero and become statistically indistinguishable from zero. In line with our expectations, the residual within-family genetic variation of PGI^{EA} is not confounded with other family characteristics that may influence children's educational outcomes, suggesting a causal interpretation of our estimated genetic effects. Figure 1 also demonstrates that PGI^{EA} and VA^{d} are not correlated with each other after conditioning on parental PGI^{EA}. The absence of such gene-environment correlations suggests that we have sufficient independent variation in PGI^{EA} and VA^{d} to separately identify genetic effects, school effects, and the gene-environment interaction of interest.

II. Exogenous variation in school value-added (VA)

Value-added (VA) models estimate the causal effect of schools on student outcomes by comparing students' academic progress relative to comparable peers at different schools (see (13) for a recent overview article). The core identification challenge is to isolate VA^{d} from other factors that contribute to student outcomes. For example, it is well-documented that school enrollment is not random but stratified by factors such as student ability, parental socioeconomic status, and ethnicity (4, 19, 20). As these factors contribute to student outcomes, uncontrolled comparisons of educational outcomes across schools will yield biased estimates of VA^{d}. In some settings, researchers can exploit random student assignments based on lotteries to estimate VA^{d} net of confounding factors (4, 21). In the absence of random assignment, however, we can mimic such experimental variation using observational data following the protocols suggested by (22) and (3). In particular, we calculate VA^{d} while conditioning on observable differences across students, including differences in family socioeconomic status and prior student test scores. Therefore, the identification of school effects relies on the assumption that the predetermined characteristics are sufficient to control for selection into schools. Existing literature has documented that the inclusion of prior test scores usually satisfies this assumption (3, 22). Figure 2 provides evidence supporting the predictive validity and the exogeneity of VA^{d} in our study. In Panel (a), we assess whether our measure of school VA^{d} captures relevant variation in student outcomes. To this end, we regress the outcome of interest, i.e., reading test scores in grade 9, on the corresponding measure of VA^{d}. The slope is precisely estimated and cannot be statistically distinguished from 1: a 1 SD increase in VA^{d} increases reading test scores by 1 SD as well. This property is often called "forecast unbiasedness" and establishes the high predictive validity of VA^{d} for the corresponding student outcome.

In Panel (b), we assess whether this relationship is potentially confounded by selection based on unobserved characteristics. To that end, we predict student test scores from a variety of variables that we do not control for in the construction of VA^{d} and regress these predicted test scores on VA^{d}. A coefficient different from zero would suggest that unobserved variables determine selection into schools, and we would have to reject the exogeneity of VA^{d}. However, the slope is

flat and very close to zero: a 1 SD increase in school VA^d decreases predicted reading test scores by 0.015 SD. Similar to findings in other studies (3, 14, 22), this suggests that there is negligible bias in our value-added estimates after conditioning on a set of controls, including students' prior test scores. In Panels (c) and (d), we repeat the same exercise for numeracy test scores. The results are almost identical, further supporting our identification assumption.

Gene-environment interaction

Reading. Table 2 documents that more effective schools have higher relative impacts on the reading outcomes of children with a low PGI^{EA}.

In our base model without any controls (column 1), a 1 SD increase in children's PGI^{EA} is associated with 0.304 SD higher reading scores, whilst a 1 SD increase in VA^d is associated with 0.091 SD higher reading scores. The interaction between PGI^{EA} and VA^d is negative and significant at the 5% level, suggesting that the effect of children's genetic endowments on their reading skills is moderated by school quality. However, these estimates lack a causal interpretation due to potential confounding by indirect genetic effects and non-random selection into schools.

In column (2), we incorporate controls for parental PGI^{EA} and genotyping protocols. In this model, a 1 SD increase in PGI^{EA} increases reading test scores by 0.230 SD. The drop in the effect of PGI^{EA} in comparison to column (1) is consistent with established literature that suggests that 40-50% of the raw PGI^{EA} associations with academic skills reflect indirect genetic effects and population stratification (17, 23, 24). After controlling for PGI^{EA} of parents, our estimates rely on random within-family variation only and are not confounded by other family characteristics that may correlate with educational outcomes. The effect of PGI^{EA} has a causal interpretation.

In column (3), we incorporate the full set of controls used in VA^d construction, including lagged test scores and school-cohort characteristics (13). The effect of PGI^{EA} remains stable, highlighting its causal interpretation after accounting for the PGI^{EA} of parents. However, the effect of VA^d drops by almost 50% from 0.091 to 0.052. This drop is expected since controlling for parental PGI^{EA} is insufficient to control for school selection. After accounting for the expanded set of covariates, our estimates of school effects account for selection into schools and are not confounded by other family characteristics that may correlate with educational outcomes. The effect of VA^d has a causal interpretation.

In column (3), we can give the base effects of both PGI^{EA} and VA^d a causal interpretation. However, following the arguments of (25) and (26), it is still an open question whether we can give the gene-environment interaction a causal interpretation as well. Since our treatment variables are considered exogenous conditional on a set of covariates, we need to include the full set of two-way interactions of these covariates with the variables of interest (PGI^{EA}, VA^d) to ensure that the gene-environment interaction is not picking up spurious correlations. Importantly, however, when including these two-way interactions, the researcher faces a bias-variance tradeoff. On the one hand, the integration of two-way interactions is necessary for the unbiased estimation of the gene-environment interaction if these two-way interactions are correlated with

the outcome and the gene-environment interaction of interest (25, 26). On the other hand, the two-way interactions may lead to a loss of statistical power and inflate standard errors, particularly when degrees of freedom decrease substantially, R^2 increases minimally, or collinearity exists between interaction terms (26, 27).

In column (4), we augment our regression model by adding all 2-way interactions of PGI^{EA} and VA^d with the vector of covariates X. If our gene-environment interaction of interest was confounded by other interactions between our variables of interest and the controls, we would expect the point estimate of $PGI^{EA} \times VA^d$ in column (4) to diverge from the corresponding estimate in column (3). However, this is not the case. The estimate of the gene-environment interaction remains stable, but the standard errors increase from 0.005 to 0.007. This suggests that the magnitude of the interaction between the PGI^{EA} and VA^d identified in column (3) is unbiased but becomes non-significant in column (4) due to the increased variance of the estimates. In view of the stability of the point estimates and in line with the arguments put forward in (26), we focus on the estimates in column (3) as our preferred estimates.

Our preferred estimate for the gene-environment interaction in column (3) suggests that a 1 SD increase in PGI^{EA} increases the reading test scores of students in the average school in Norway by 0.231 SD. For students attending a school 1 SD above the country average, the impact of a 1 SD increase in PGI^{EA} decreases by approximately 6% [1-(0.231-0.013)/0.231]. This estimate is statistically significant at the 5% level.

Theoretically, PGI^{EA} and VA^d could be complements or substitutes for student learning. If they were complements, school quality (VA^d) would magnify advantages based on PGI^{EA}; if they were substitutes, VA^d would compensate for disadvantages based on PGI^{EA}. Our results point to the substitutability of PGI^{EA} and VA^d as input factors for students' reading test scores. Figure 3 illustrates genetic gradients across Norwegian schools of varying quality (±2 SD from the mean), revealing whether this substitutability stems from gains at the bottom or losses at the top of the PGI^{EA} distribution. The genetic gradients are flatter in higher-quality schools. This pattern suggests that in higher-quality schools, genetic differences between children matter less because schools compensate children with lower PGI^{EA}. Reversely, the impact of genetic differences among students on their test scores is more pronounced in lower-quality schools. This finding suggests that investments in school quality can address educational inequalities based on genetic differences between children.

Numeracy. We repeat the previous analysis with numeracy test scores as the outcome of interest. SI Appendix, Tab. S.1 and SI Appendix, Fig. S.1 suggest that there is no gene-environment interaction for numeracy scores. In our preferred specification, a 1 SD increase in PGI^{EA} is associated with 0.239 SD higher numeracy scores, whilst a 1 SD increase in VA^d is associated with 0.039 SD higher numeracy scores. The point estimate for the gene-environment interaction is 0.000, with an associated 95% confidence band of [-0.0078, 0.0078]. Therefore, this null finding is precise enough to exclude magnitudes that are approximately half the size of the point estimate for the gene-environment interaction in the reading domain (0.013).

Discussion

Summary. We investigated whether schools can mitigate birth-related educational inequalities by integrating exogenous school value-added measures with the natural lottery of within-family genetic variation. Using this stringent causal inference design, we found compelling evidence of a gene-environment interaction influencing reading skill development (though not numeracy) even within the narrow time window of one school year (grade 8, ages 13-14). Our results suggest that investments in school quality can promote equitable skills development by effectively narrowing gaps in reading test scores between students with different genetic predispositions. Notably, these findings also reveal a double disadvantage: the skill development gap between children in low versus high-quality schools is even larger for those with less genetic predisposition to education.

Reading vs. numeracy. We estimate that increases in school quality reduce the impact of PGI^{EA} on reading test scores; however, we do not find a gene-environment interaction for numeracy test scores. This result is likely related to the higher persistence of numeracy skills during this developmental period. Examining our preferred specification (column 3 of Tables 2 and S.1) gives an indication of the persistence of skills. In this specification, we control for subject-specific lagged test scores and list the coefficient of this control at the bottom of the table (ρ). If ρ=0, past achievement does not impact current performance, giving room for new inputs to shape outcomes. Reversely, if ρ=1, skills are highly persistent, suggesting that new inputs have less scope to shape children's skills. The corresponding coefficients are 0.462 (*SE* 0.006) for reading and 0.702 (*SE* 0.004) for numeracy test scores. These estimates suggest that numeracy test scores of adolescents in Norway are significantly more persistent than reading test scores, giving high-quality schools less scope to level up the numeracy skills of children with lower PGI^{EA} relative to their high-PGI^{EA} peers.

Magnitude of effects. We estimate that a 1 SD increase in school quality reduces the impact of PGI^{EA} on student outcomes in reading by approximately 6%. To gauge the magnitude of this effect, it is essential to emphasize that this treatment effect captures students' exposure to high- or low-quality schools for only one school year (grade 8). Lower-secondary education in Norway lasts for three years (grades 8-10), with students usually staying in their initial neighborhood school throughout this period (see SI Appendix, section A, and (14)). Therefore, assuming linear and additive treatment effects across grades 8-10, a 1 SD increase in school quality would reduce the impact of PGI^{EA} on student outcomes in reading by approximately 18% over the total duration of lower secondary school. While the assumption of linear and additive treatment effects is arguably strong, our projected effect size is consistent with results from (12), who find that a 1 SD increase in high school quality in the US (grades 9-12) decreases the impact of PGI^{EA} on educational attainment by approximately 19%.

Potential mechanisms. The gene-environment interaction identified here can be further understood in the context of theoretical frameworks from economics and developmental psychology. The economics literature on skill formation often conceptualizes student outcomes as a function of students' initial skills, school inputs, and family inputs, where families adjust their

behavior depending on students' initial skills and school quality (28, 29). Similarly, developmental psychology frameworks propose that development and learning are a product of dynamic interplay between individual bio-psychological and social processes (30), where an existing genetic diathesis/vulnerability can be compensated for, controlled, or triggered by proximal social processes (31). Our finding that genetic factors matter less in high-quality schools is consistent with the concept of substitutability from economics, as well as the compensation and diathesis-stress models (11, 31, 32) from developmental psychology. Notably, this also goes against the influential bioecological model of child development, whereby social advantage increases the expression of genetic differences (30).

The finding that differences in PGI^{EA} have less impact on the development of reading skills in higher VA^{d} schools could be explained by both direct and indirect mechanisms. First, students with lower PGI^{EA} may gain *directly* from attending higher-quality schools. Emerging evidence shows that schools and teachers in industrialized countries focus on the lower parts of the achievement distribution, suggesting that they attach a higher weight to the learning of disadvantaged students (33). Therefore, students with lower PGI^{EA} who attend better schools receive relatively more and/or higher-quality investments than their peers with higher PGI^{EA}, which could explain the negative gene-environment interaction in this study. This mechanism assumes that Norwegian educators distribute instructional resources unequally across students within the same school. However, even without this assumption, the negative gene-environment interaction can be explained by diminishing returns to educational inputs. Students with lower PGI^{EA} may have more room for improvement and, consequently, may gain more from attending a better school. Notably, compensation and triggering are at the ends of a continuum: just as enriched learning environments may compensate for genetic disadvantage, lower-quality schools could be stressful environments that "trigger" genetic predispositions linked to low educational attainment and hinder the accumulation of reading skills.[*]

Second, students with lower PGI^{EA} may gain *indirectly* from attending high-quality schools through family adjustments to school quality and children's PGI^{EA}. These indirect mechanisms are more complex as they combine the effects of different inputs on student learning with the behavioral responses of parents. For example, if families prioritize supporting children with higher PGI^{EA} and family and school inputs act as substitutes in fostering learning, then the effect of additional family inputs received by high-PGI^{EA} students becomes weaker in high-VA^{d} schools. This could contribute to the negative gene-environment interaction we observed. Alternatively, if families decrease their investment with increases in school quality, and family inputs and PGI^{EA} work as complements in learning, then the effect of decreased family inputs received by students in high-VA^{d} schools will be less pronounced for low-PGI^{EA} students. This could also contribute to the negative gene-environment interaction observed in this study.

Future research. This study forges several interesting avenues for future research. Ideally, investigations aiming to distinguish between different mechanisms involved in the gene-environment interaction that we identified should combine the data prerequisites for causal gene-environment interplay studies with detailed data on school practices and parental inputs. The former will allow us to understand the characteristics of high-quality schools and to study which

features of these schools make them particularly beneficial to students with lower PGI^{EA} (see also our discussion on *direct* gene-environment effects). The latter will allow us to understand how mothers and fathers adapt their parenting strategies in response to their children's PGI^{EA} and the quality of their schools, and whether particular parental inputs are especially beneficial to students with lower PGI^{EA} (see also our discussion on *indirect* gene-environment effects).

The relevant school characteristics and parental inputs are highly multifaceted and are unlikely to be captured in a single dataset. School quality is likely to consist of diverse pedagogical, organizational, cultural, relational, financial, and physical aspects. Similarly, parental inputs may consist of different time and monetary investments as well as parenting styles. However, the increased availability of molecular genetic data and the integration of these data with linked register, survey, and cohort study datasets paves the way for researchers to address these interesting questions convincingly in the future.

Materials and Methods

Data

The Norwegian Mother, Father, and Child Cohort Study (MoBa). MoBa is a prospective population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health (15). Pregnant women were recruited from across Norway from 1999 to 2009. The women consented to initial participation in 41% of the pregnancies. Of the fathers invited to participate, 83% consented. The total cohort includes approximately 114,500 children, 95,200 mothers, and 75,200 fathers. MoBa participants were linked to administrative register data through the Norwegian national ID number system. Analyses are conducted on MoBa children born 2002-2008 with grade 9 national test scores in reading and numeracy, complete data for genome-wide genotyping (see SI Appendix, section C and (34) for details on genotyping and genetic quality control in MoBa), information on VA^d in their school-cohort cell, and non-missing control variables (N=30,939).

Norwegian register data. We estimate VA^d for standardized test scores in reading and numeracy in grade 8. Since standardized tests are conducted at the beginning of the academic year, we can use test scores in grades 8 and 9 to measure student progress in grade 8, i.e., the first year of lower secondary school (see SI Appendix, sections A and B). We construct VA^d using register data on the entire Norwegian student population in birth cohorts 1997-2008 (approximately 60,000 per cohort). The earliest birth cohort that completed comparable standardized tests in grades 8 and 9 in reading and numeracy is 1997. The latest birth cohort for whom standardized test outcomes are available in the register data is 2008.

Treatment variables

Polygenic index for educational attainment (PGI^{EA}). We used beta weights from the largest genome-wide association study of educational attainment to date ("EA4"), excluding MoBa (35). Polygenic indices were calculated using LDPred v.1 (36), a Bayesian approach that uses a prior

on the expected polygenicity of a trait (assumed fraction of non-zero effect markers) and adjusts for linkage disequilibrium (LD) based on a reference panel to compute SNPs weights. LD adjustment was performed using the MoBa genotype data as LD reference panel. The weights were estimated based on the heritability explained by the markers in the GWAS summary statistics and the assumed fraction of markers with non-zero effects. PGI^{EA} were computed based on these weights with the –score command in plink2 (37).

School Value-Added (VA^d). Consider educational outcome *Y* in subject *d* of student *i* attending school *j* in cohort *c*. We model this educational outcome as a function of individual student characteristics *Z*, cohort fixed effects τ, and school quality VA^d:

$$Y_{ijc}^{d} = \beta^{d} Z_{ijc} + \tau_{c}^{d} + VA_{jc}^{d} + \varepsilon_{ijc}^{d} \quad . \qquad (1)$$

In our setting, *Z* comprises lagged grade 8 test scores in numeracy, reading, English, maternal and paternal years of education, second-generation migration status, gender, birth cohort, birth order, number of siblings, and school-cohort averages of all previous controls.

Note that VA^d is a latent variable captured in the composite error term $\mu_{ijc}^{d} = VA_{jc}^{d} + \varepsilon_{ijc}^{d}$ of equation (1). We can construct an estimate of VA^d of school *j* in cohort *c* by estimating equation (1) and calculating the cohort-school average in the resulting residuals:

$$\widehat{VA}_{jc}^{d} = \sum (\mu_{ijc}^{d})/N_{jc}^{d}\,, \qquad (2)$$

where N_{jc}^{d} captures the number of students of cohort *c* in school *j*.

We want to use estimates of VA^d in regression models to explain student outcomes. However, we cannot explain student outcomes of school *j* in cohort *c* using VA^d measures for the same school cohort because of the mechanical relationship between the dependent variable *Y* and the independent VA^d variable (see equation [1]). For example, a student with a high reading test score will simultaneously push up the corresponding measure of VA^d in their school-cohort cell. This mechanical link is particularly pronounced if school-cohort cells are small. To break this mechanical relationship, we predict VA^d in school *j* of cohort *c* from all neighboring cohorts *l=1,...,c-1,c+1,…, L* using an empirical Bayes procedure (see for (38) a recent overview article):

$$\underline{VA}_{jc}^{d} = \sum_{l=1}^{L} \zeta_{jl}^{d} (\widehat{VA}_{jl}^{d}), \qquad (3)$$

where ζ are weights selected to minimize forecast errors. Similar to (22), we use all neighboring cohorts and not just preceding cohorts to increase the precision of the VA^d estimates. Therefore, our final measure of VA^d is the best linear predictor of VA^d for cohort *c* in school *j* from all preceding and subsequent cohorts who attended this school while excluding the cohort itself to avoid biased estimates through reversed causality. This procedure yields a noisy measure of VA^d. Furthermore, it is well-known that measurement error in the independent variables leads to attenuation bias in the relevant coefficients in downstream analyses. The empirical Bayes procedure takes care of

this concern. Specifically, it chooses weights ζ such that noisy estimates of VA^d are shrunk to the mean in proportion to their signal-to-noise ratio. It can be shown analytically that this weighting is the exact inverse of attenuation bias in error-in-variables regressions – see (38) for an outline of the formal argument. Therefore, while we measure VA^d with error, our regressions recover estimates of school effects that are not afflicted by attenuation bias.†

This procedure yields an unbiased estimate of VA^d if there is no selection into schools based on factors not captured in observable characteristics *Z*. Following (22), we can evaluate the plausibility of the exogeneity assumption using "as-if-unobservable" variables. Specifically, we treat students' 5th-grade literacy and numeracy scores, fathers' and mothers' earnings rank at age 6, and fathers' and mothers' age at birth as unobserved variables that we do not include in the control vector *Z*. In turn, we can test whether they confound the relationship between VA^d and student outcomes.

The validation exercise consists of three steps. First, we separately regress each of our outcomes of interest and the as-if-unobservables on *Z* and store the residuals. This step ensures that we only exploit variation that is not captured by Z. Second, we regress each (residualized) outcome on all (residualized) "as-if-unobservables" and store the predicted outcomes. This step creates a summary statistic for variation in the (residualized) outcomes that is accounted for by our (residualized) "as-if-unobservables." It captures variation in the outcomes of interest that is not accounted for by *Z* and which, therefore, is a potential source for omitted variable bias. Finally, we regress this summary statistic on VA^d. If VA^d is substantially associated with the predicted outcomes, then this indicates that there is selection into schools based on "as-if-unobservables" (i.e., 5th-grade test scores, parental earnings rank, and parental age at birth). The results are shown in Figure 2.

VA^d can be interpreted as a summary statistic for all school factors contributing to students' academic progress in skill dimension *d*. However, since VA^d is constructed controlling for an extensive set of school-cohort characteristics, including the prior attainment and socio-economic status of peers, it is unlikely to capture peer effects. Moreover, while VA^d captures persistent differences in quality across schools, it does not capture within-school differences in quality due to, for instance, teacher quality.

Analysis

We estimate the following model through ordinary least-squares and cluster standard errors at the level of schools *j*:

$$Y_{ijc}^{d} = \alpha^{d} PGI_{ijc}^{EA} + \beta^{d} VA_{jc}^{d} + \kappa^{d}(PGI_{ijc}^{EA} \times VA_{jc}^{d}) + \delta^{d} X_{ijc} + \varepsilon_{ijc}^{d}. \quad (4)$$

PGI^{EA} and VA^d are the variables of interest, *X* is a vector of control variables, and ε^d is the error term. α^d, β^d, and κ^d are the parameters of interest, identifying the causal effects of PGI^{EA}, VA^d, and the corresponding gene-environment interaction.

Controls *X* include genetic controls, i.e., paternal and maternal PGI^{EA}, genotyping batch, and the vector of covariates *Z* used for the construction of VA^d, i.e., lagged grade 8 test scores in numeracy, literacy, English, maternal and paternal years of education, second generation migration status, gender, birth cohort, birth order, number of siblings, and school-cohort averages of all previous controls. Note that lagged test scores are a function of PGI^{EA}. Therefore, they are "bad controls" for estimating genetic effects (27). To address this concern, we regress grade 8 test scores in numeracy, literacy, and English on PGI^{EA} and include the residuals from these regressions as our controls for lagged test scores. Hence, we control for all variations in lagged test scores uncorrelated with our variable of interest (PGI^{EA}).

Endnotes

[*] Theoretically, the existence of this effect is ambiguous. Extensive literature has emphasized the importance of dynamic complementarities in skill formation, which suggests higher gains by students with higher initial skills (39). However, consistent with our results, various recent papers have suggested that higher quality schools have more positive effects on students from disadvantaged socioeconomic backgrounds and who have higher genetic predisposition to ADHD and low educational attainment (10–12). Other studies have obtained inconclusive or null findings (9, 40). Some of these inconsistencies could be explained by differences in outcome selectivity (i.e., national tests versus degree completion) and/or differences across institutional contexts (41).

[†] Note that this conclusion only holds when standardizing VA^d with respect to its true SD, which is unobserved. Therefore, we estimate the true SD by the square root of the one-year lag autocovariance, which provides a lower bound on the true within-year SD of VA^d *(22, 42)*. We use this estimate for all standardizations of VA^d.

Ethics

The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from the Regional Committees for Medical and Health Research Ethics. The MoBa cohort is now based on regulations related to the Norwegian Health Registry Act. The current study was approved by the Regional Committees for Medical and Health Research Ethics (#2017/2205).

Acknowledgments

We are grateful to all the participating families in Norway who are taking part in the ongoing MoBa cohort study. We would like to thank seminar audiences at Aarhus University, the University of Oslo, the University of Bristol, the University of Manchester, WU Vienna, the Workshop on Early Investments, Family Well-Being, and Child Development (Rødvig), the European Social Science Genomics Network Conference (Rotterdam), the Genetics and Social Sciences Conference (Oslo), the Research Committee 28 on Social Stratification and Mobility conference (Milano), and the Workshop on Education Economics and Policy (Trondheim) for valuable feedback. All remaining errors remain our own. This research was supported by funding from the European Research Council (#101115949), the Research Council of Norway (#325245), the Research

Council of Norway's Centres of Excellence scheme (#331640), and the European Union's HORIZON-MSCA-2021-DN-01 (#101073237). RC is also supported by the Jacobs Foundation (#2023-1510-00) and European Research Council (#101045526); PH by UKRI Future Leaders Fellowship (#MR/X033333/1). The Norwegian Mother, Father, and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. The Norwegian registry and MoBa data used was from the project SUBPU. The Department of Psychology, University of Oslo, is responsible for the data handling of SUBPU, a Data Protection Impact Assessment (DPIA) has been signed by the head of department, and the project manager is Eivind Ystrom. SUBPU is approved by Committees for Medical and Health Research Ethics (#2017/2205). SUBPU has agreements with the MoBa and Statistics Norway for data linkage and usage. The data access and management costs of SUBPU is financed by the Research Council of Norway (RCN) (#336078, #288083, and #314601), the European Research Council (#101045526, #818425, #101088481, and #818420), and supported by the Department of Psychology (UiO). All data management and analyses were on the secure data "Tjeneste for Sensitive Data" (TSD) facilities, owned by the University of Oslo. Resources provided by Sigma2, the National Infrastructure for High Performance Computing and Data Storage in Norway (UNINETT), was used for analyses (#NS9867S). The authors would like to acknowledge the work of SUBPU data managers Clara Timpe and Oda van Jole.

Bibliography

1. J. S. Coleman, "Equality of Educational Opportunity" (National Center for Educational Statistics, 1966).
2. E. A. Hanushek, S. G. Rivkin, Harming the Best: How Schools Affect the Black–White Achievement Gap. *J Policy Anal Manage* **28**, 366–393 (2009).
3. C. K. Jackson, S. Kiguel, S. C. Porter, J. Q. Easton, Who Benefits from Attending Effective High Schools? *J Labor Econ* **42**, 717–751 (2024).
4. D. J. Deming, J. S. Hastings, T. J. Kane, D. O. Staiger, School choice, school quality and postsecondary attainment. *Am. Econ. Rev.* **104**, 991–1013 (2014).
5. R. Cheesman, *et al.*, Childhood behaviour problems show the greatest gap between DNA-based and twin heritability. *Translational Psychiatry* (2017).
6. E. Krapohl, *et al.*, The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci USA* **111**, 15273–15278 (2014).
7. E. M. Tucker-Drob, T. C. Bates, Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. *Psychol. Sci.* **27**, 138–149 (2016).
8. E. L. de Zeeuw, D. I. Boomsma, Country-by-genotype-by-environment interaction in childhood academic achievement. *Proc Natl Acad Sci USA* **114**, 13318–13320 (2017).
9. S. Trejo, *et al.*, Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health. *Sociol. Sci.* **5**, 513–540 (2018).

10. R. Cheesman, *et al.*, How interactions between ADHD and schools affect educational achievement: a family-based genetically sensitive study. *J. Child Psychol. Psychiatry* **63**, 1174–1185 (2022).

11. R. Cheesman, *et al.*, A population-wide gene-environment interaction study on how genes, schools, and residential areas shape achievement. *NPJ Sci. Learn.* **7**, 29 (2022).

12. B. Arold, P. Hufe, M. Stoeckli, Genetic Endowments, Educational Outcomes, and the Moderating Influence of School Quality. *J Polit Econ: Micro* (2025).

13. J. Angrist, P. Hull, C. Walters, "Methods for measuring school effectiveness" in *Handbook of the Economics of Education*, (Elsevier, 2023), pp. 1–60.

14. L. J. Kirkebøen, School Value-Added and Long-Term Student Outcomes. *SSRN Journal* (2022) https:/doi.org/10.2139/ssrn.4123829.

15. P. Magnus, *et al.*, Cohort profile update: the Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **45**, 382–388 (2016).

16. P. Biroli, *et al.*, The economics and econometrics of gene-environment interplay. *Rev Econ Stud* (2025).

17. A. Kong, *et al.*, The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).

18. A. I. Young, S. Benonisdottir, M. Przeworski, A. Kong, Deconstructing the sources of genotype-phenotype associations in humans. *Science* **365**, 1396–1400 (2019).

19. J. M. Rothstein, Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions. *American Economic Review* **96**, 1333–1350 (2006).

20. J. G. Altonji, R. K. Mansfield, Estimating group effects using averages of observables to control for sorting on unobservables: school and neighborhood effects. *American Economic Review* **108**, 2902–2946 (2018).

21. A. Abdulkadiroğlu, P. A. Pathak, J. Schellenberg, C. R. Walters, Do parents value school effectiveness? *American Economic Review* **110**, 1502–1539 (2020).

22. R. Chetty, J. N. Friedman, J. E. Rockoff, Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* **104**, 2593–2632 (2014).

23. S. Selzam, *et al.*, Comparing Within- and Between-Family Polygenic Score Prediction. *Am. J. Hum. Genet.* **105**, 351–363 (2019).

24. M. G. Nivard, *et al.*, More than nature and nurture, indirect genetic effects on children's academic achievement are consequences of dynastic social processes. *Nat. Hum. Behav.* **8**, 771–778 (2024).

25. M. C. Keller, Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24 (2014).

26. B. Feigenberg, B. Ost, J. A. Qureshi, Omitted Variable Bias in Interacted Models: A Cautionary Tale. *Rev Econ Stat* (2025).

27. J. Angrist, J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, 2009).

28. P. E. Todd, K. I. Wolpin, On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* **113**, F3–F33 (2003).

29. C. Pop-Eleches, M. Urquiola, Going to a better school: effects and behavioral responses. *American Economic Review* **103**, 1289–1324 (2013).

30. U. Bronfenbrenner, S. J. Ceci, Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychol. Rev.* **101**, 568–586 (1994).

31. M. J. Shanahan, S. M. Hofer, Social context in gene-environment interactions: retrospect and prospect. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **60 Spec No 1**, 65–76 (2005).

32. R. Rende, R. Plomin, Diathesis-stress models of psychopathology: A quantitative genetic perspective. *Applied and Preventive Psychology* **1**, 177–182 (1992).

33. A. Salvati, Teacher Instruction, Classroom Composition, and Student Achievement. *mimeo* (2025).

34. E. C. Corfield, *et al.*, The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. *BioRxiv* (2022) https:/doi.org/10.1101/2022.06.23.496289.

35. A. Okbay, *et al.*, Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).

36. B. J. Vilhjálmsson, *et al.*, Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).

37. C. C. Chang, *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015–0047–8 (2015).

38. C. Walters, "Empirical Bayes methods in labor economics" in *Handbook of Labor Economics*, (Elsevier, 2024), pp. 183–260.

39. J. J. Heckman, S. Mosso, The economics of human development and social mobility. *Annu. Rev. Econom.* **6**, 689–733 (2014).

40. K. Stienstra, A. Knigge, I. Maas, Gene-environment interaction analysis of school quality and educational inequality. *NPJ Sci. Learn.* **9**, 14 (2024).

41. G. Ghirardi, F. Bernardi, Compensating or boosting genetic propensities? Gene-family socioeconomic status interactions by educational outcome selectivity. *Soc. Sci. Res.* **129**, 103174 (2025).

42. L. Yariv, B. Gillen, E. Snowberg, Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy*, 701681 (2018).

Figures and Tables

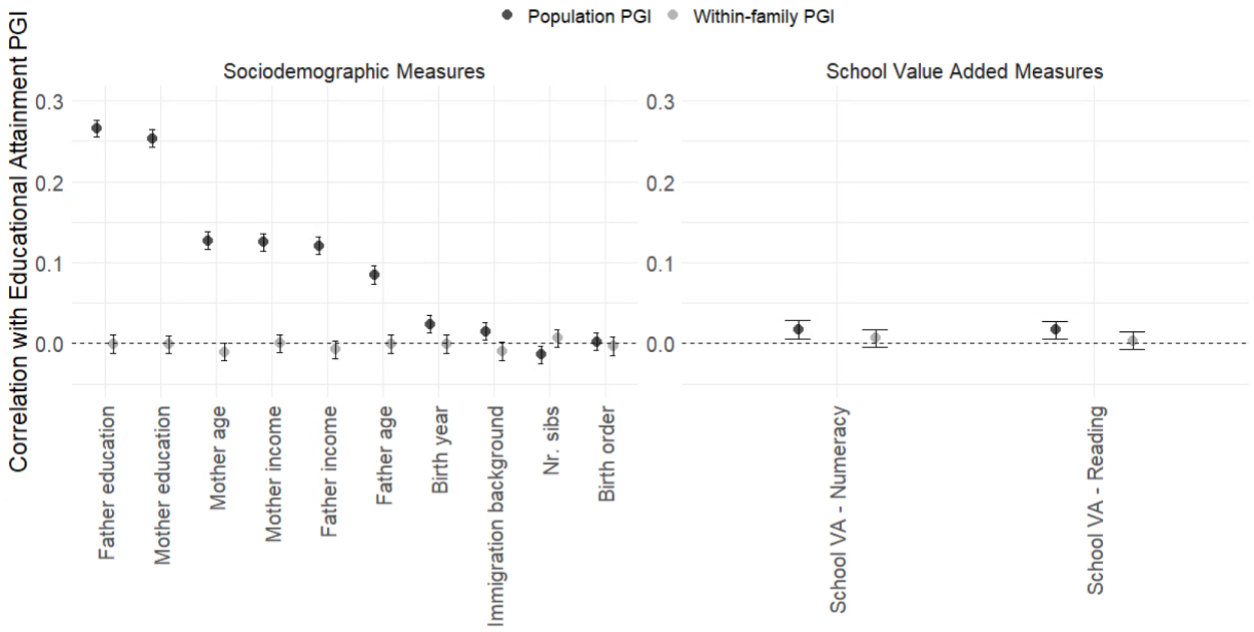

Figure 1. Validation of the exogeneity of within-family PGI^{EA} variation. This plot shows correlations of PGI^{EA} with children's observable sociodemographic background characteristics and school value-added (VA^{d}) in our analysis sample (N=30,939). Dark circles show uncontrolled population-level correlations with children's PGI^{EA}. Light circles show the corresponding within-family correlations after controlling for PGI^{EA} of biological mothers and fathers. Whiskers show 95% confidence intervals. Standard errors are heteroskedasticity robust. Data: Own calculations based on MoBa and Norwegian registers.

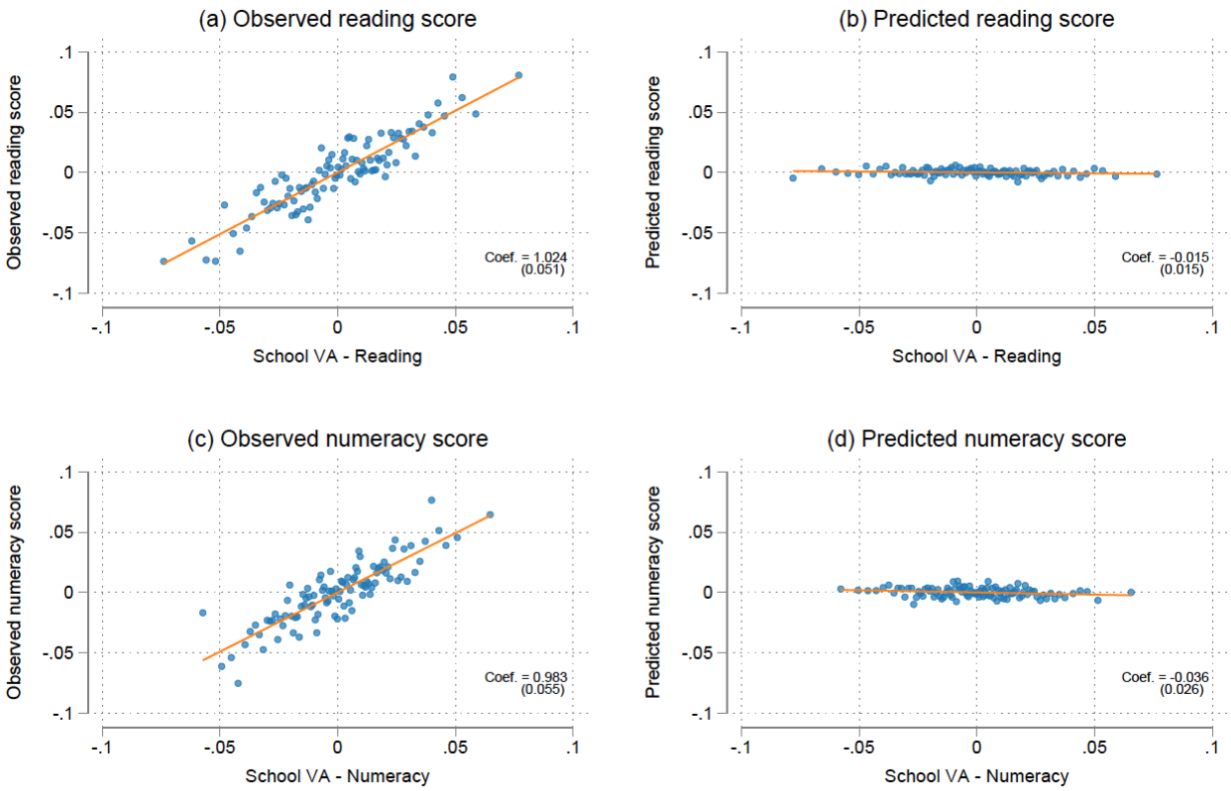

Figure 2. Validation of the exogeneity of VA^d. This figure shows correlations of VA^d with children's observable and predicted test scores in reading (Panels [a] and [b]) and numeracy (Panels [c] and [d]) in grade 9 for the full population of cohorts 1997-2007 (N=508,615). All variables are residualized from control variables Z (see Materials and Methods). Predicted test scores are constructed from children's literacy and numeracy test scores in grade 5, maternal and paternal earnings rank at age 6, and maternal and paternal age at birth. Scatter plots are constructed by binning the VA^d distribution into 100 percentiles. Regression slopes are estimated on the full data. Standard errors are clustered at the school level. Data: Own calculations based on Norwegian registers.

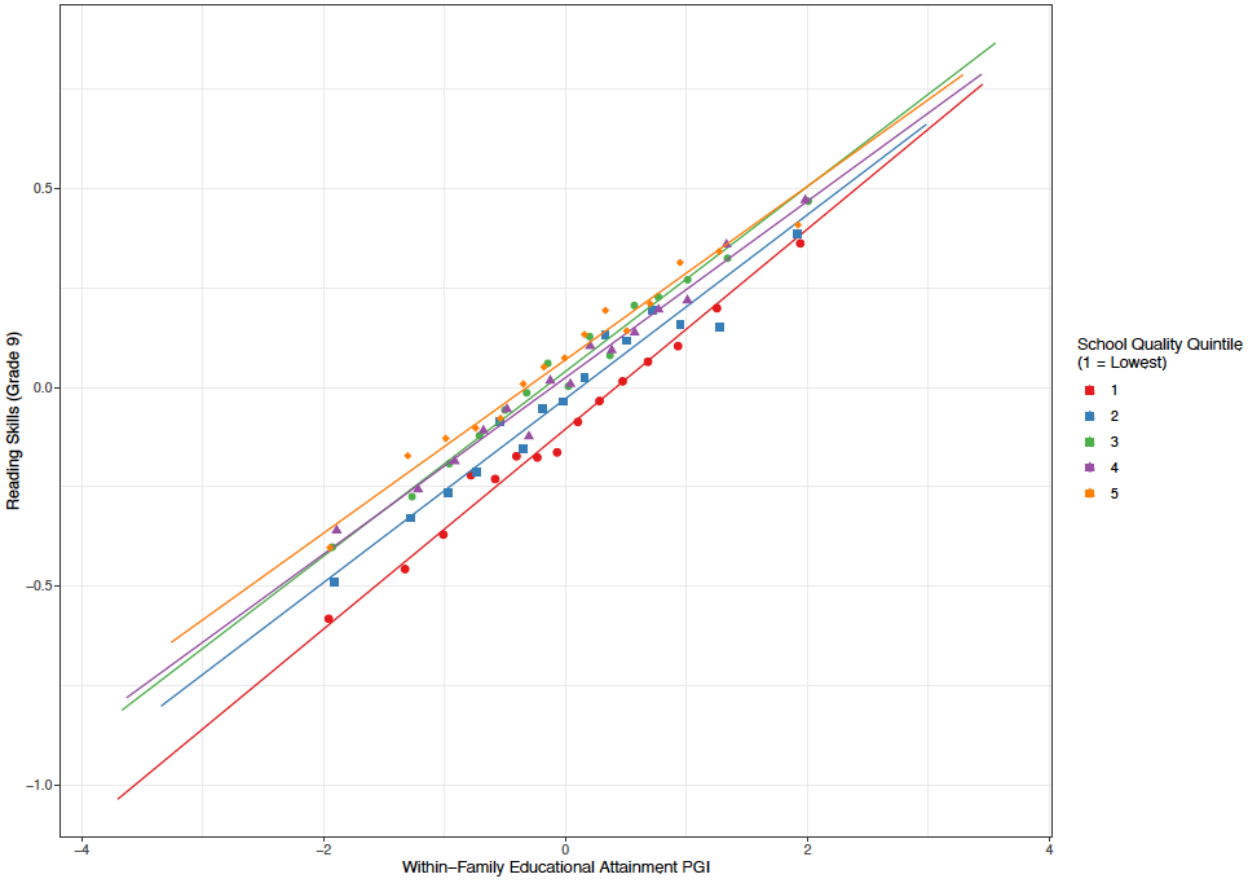

Figure 3. Gene-environment interaction for reading test scores. This figure shows binned scatter plots for the relationship between PGI^{EA} and reading test scores in grade 9 by quintile of the VA^{d} distribution. Scatter plots are constructed by grouping the PGI^{EA} distribution into 15 bins. Regression slopes are estimated on the full data, conditioning on controls matching the focal model (see column 3, Table 2; Materials and Methods). Data: Own calculations based on MoBa and Norwegian registers.

	Analysis sample $N = 30,939$				MoBa (All) $N = 56,533$		Population $N = 331,591$	
	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Mean	St. Dev.
Birth year	2004.9	1.6	2002	2008	2004.8	1.6	2004.5	1.7
Female	0.5	0.5	0.0	1.0	0.5	0.5	0.5	0.5
Migration background	0.1	0.3	0.0	1.0	0.1	0.3	0.2	0.4
Education (Father)	14.6	2.6	7.0	21.0	14.4	2.7	13.7	2.9
Education (Mother)	15.1	2.3	9.0	21.0	15.0	2.4	14.1	2.9
Inc. rank (Father)	58.5	25.6	0.0	99.0	57.1	26.2	50.9	28.3
Inc. rank (Mother)	61.0	25.4	0.0	99.0	59.9	25.7	51.5	27.6
Age (Father)	32.9	5.1	18.0	65.0	33.1	5.3	33.2	6.0
Age (Mother)	30.5	4.4	16.0	47.0	30.6	4.5	30.2	5.1
Reading (Grade 8)	0.3	0.9	-3.2	2.4	0.2	0.9	0.1	1.0
Numeracy (Grade 8)	0.3	0.9	-2.5	2.5	0.2	1.0	0.0	1.0
English (Grade 8)	0.2	1.0	-2.4	2.2	0.1	1.0	0.0	1.0

Table 1. Summary statistics. This table shows descriptive statistics. The first panel focuses on the main analysis sample, i.e., MoBa cohorts 2002-2008 with PGI^{EA} data for biological mothers, fathers, and their children. The second panel also includes MoBa participants with missing PGI^{EA} data for either mothers, fathers, or their children. The third panel focuses on the entire Norwegian population born 2002-2008, irrespective of whether they have participated in MoBa. Parental income ranks are calculated in the full population. Test scores for Reading, Math, and English are standardized on the full population. Data: Own calculations based on MoBa and Norwegian registers.

Outcome: Reading (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.304*** (0.006)	0.230*** (0.008)	0.231*** (0.005)	0.231*** (0.005)
$VA^{Reading}$	0.091*** (0.014)	0.090*** (0.013)	0.052*** (0.007)	0.050*** (0.007)
$PGI^{EA} \times VA^{Reading}$	-0.020* (0.008)	-0.020* (0.008)	-0.013* (0.005)	-0.013 (0.007)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, VA^{d}, X)	×	×	×	✓
R^2	0.096	0.104	0.654	0.657
N	30,939	30,939	30,939	30,939
Skill persistence ρ	–	–	0.462*** (0.006)	0.460*** (0.006)

Table 2. Gene-environment interaction for reading scores. This table shows estimates for the effects of PGI^{EA} and VA^{d} on children's reading scores in grade 9, as well as the corresponding gene-environment interaction. *Genetic controls* include the PGI^{EA} of biological mothers and fathers, and categorical variables for the genotyping batch. *School quality controls* include lagged grade 8 test scores in numeracy, literacy, English, maternal and paternal years of education, second-generation migration status, gender, birth cohort, birth order, number of siblings, and school-cohort averages of all previous controls. *2-way interactions* include all interactions of PGI^{EA} and VA^{d} with the aforementioned controls. Skill persistence ρ indicates the estimate for lagged test scores in reading (grade 8), which is estimated in the model as part of the child controls. Standard errors (in parentheses) are clustered at the school level. Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Data: Own calculations based on MoBa and Norwegian registers.

Supporting Information for

The genetic lottery goes to school: Better schools compensate for the effects of students' genetic differences

Rosa Cheesman[1,2*], Nicolai T. Borgen[3], Astrid M. J. Sandsør[3], and Paul Hufe[4]

1. Department of Psychology, University of Oslo, P.O. box 1094, Blindern, 0317 Oslo, Norway.

2. PsychGen Center, Norwegian Institute of Public Health, Oslo, Norway.

3. Center for Research on Equality in Education (CREATE) and Department of Special Needs Education, University of Oslo, P.O. box 1140 Blindern, 0318, Oslo, Norway.

4. School of Economics, University of Bristol, Bristol BS8 1TU, United Kingdom.

*Correspondence to Rosa Cheesman (r.c.g.cheesman@psykologi.uio.no)

This PDF file includes:

A. Norwegian school system

The Norwegian compulsory school system consists of elementary education (grades 1-7) and lower secondary education (grades 8-10). Nearly all children attend their public neighborhood school, where children are automatically enrolled. Parents can apply to attend a different public school, but this is only possible if there are spaces available, and enrollment is then determined by the principal on a case-by-case basis. Only about 5% of students attend private schools that are either religious, offer alternative pedagogies, or are international. Public schools are free of charge and have a common national curriculum. There is no tracking, and grade promotion or retention is very rare. Most children have the same classmates while attending the same school across grade levels, and teachers will often follow the same students across grade levels as well. Children do not receive grades during elementary school. During lower secondary school, children receive teacher-assessed grades biannually in each subject. Only the last semester's grades in grade 10 are part of the final transcript and are used, together with results on external exams, to apply to upper secondary education. Students are guaranteed a spot in upper secondary education in one of their three preferred programs (5 academic programs and 10 vocational programs), but the final transcript determines program admission and also school admission if the counties have free school choice within the program.

Norway is recognized as an egalitarian society with lower levels of inequality compared to many other Western countries. This relative equality can be attributed to Norway's low levels of income inequality, comprehensive redistributive welfare state institutions, and high rates of intergenerational mobility. In line with other Nordic countries, Norway exhibits smaller between-school differences (1), likely due to less sorting across schools as well as potentially fewer differences in school quality. Schools serving disadvantaged students are often allocated more resources, such as a higher teacher-student ratio in lower-performing schools (2). However, despite being relatively egalitarian, one in ten children grows up in families with persistently low household income (3). Socioeconomic achievement gaps in Norway are narrower than in countries such as the United States, yet the difference is smaller than expected given substantial cultural and economic differences between the countries (4-6).

B. Standardized national tests

National tests in numeracy, reading, and English are taken annually at the beginning of grades 5 and 8, while numeracy and reading are also tested in grade 9. The grade 9 test is the same as that in grade 8 and measures growth during the first year of lower secondary education. The tests are commissioned from experts in test development and psychometrics connected to universities in Norway. The tests are digital, and the results are scored automatically. About 96% of all students in Norway take the tests, although students with special needs and those following introductory language courses may be exempt. The test results are mainly used to collect information about students' skills and to track school development over time. Results are conveyed to teachers, students, and parents, but have no direct consequences for students. In the present study, we used numeracy and reading tests for grade 9 as our main outcome

variables, where we standardized the summed test scores with a mean of 0 and a SD of 1 within each test and year. National tests in grade 8 are used as a control variable, while the ones in grade 5 are used to validate the design.

C. Genotyping

Blood samples were obtained from both parents during pregnancy and from mothers and children (umbilical cord) at birth. Quality-controlled genotyping array data is available for the full 207,569 unique MoBa participants (7). Phasing and imputation were performed with IMPUTE4.1.2_r300.3, using the publicly available Haplotype Reference Consortium release 1.1 panel as a reference. To identify a sub-population of European-associated ancestry, principal component analysis (PCA) was performed with 1000 Genomes phase 1 after LD-pruning. During post-imputation quality control, the following thresholds were used for SNP removal: imputation quality (INFO) score $\leq$ 0.8; MAF<1%; call rate<95%.

D. Supplementary Figures

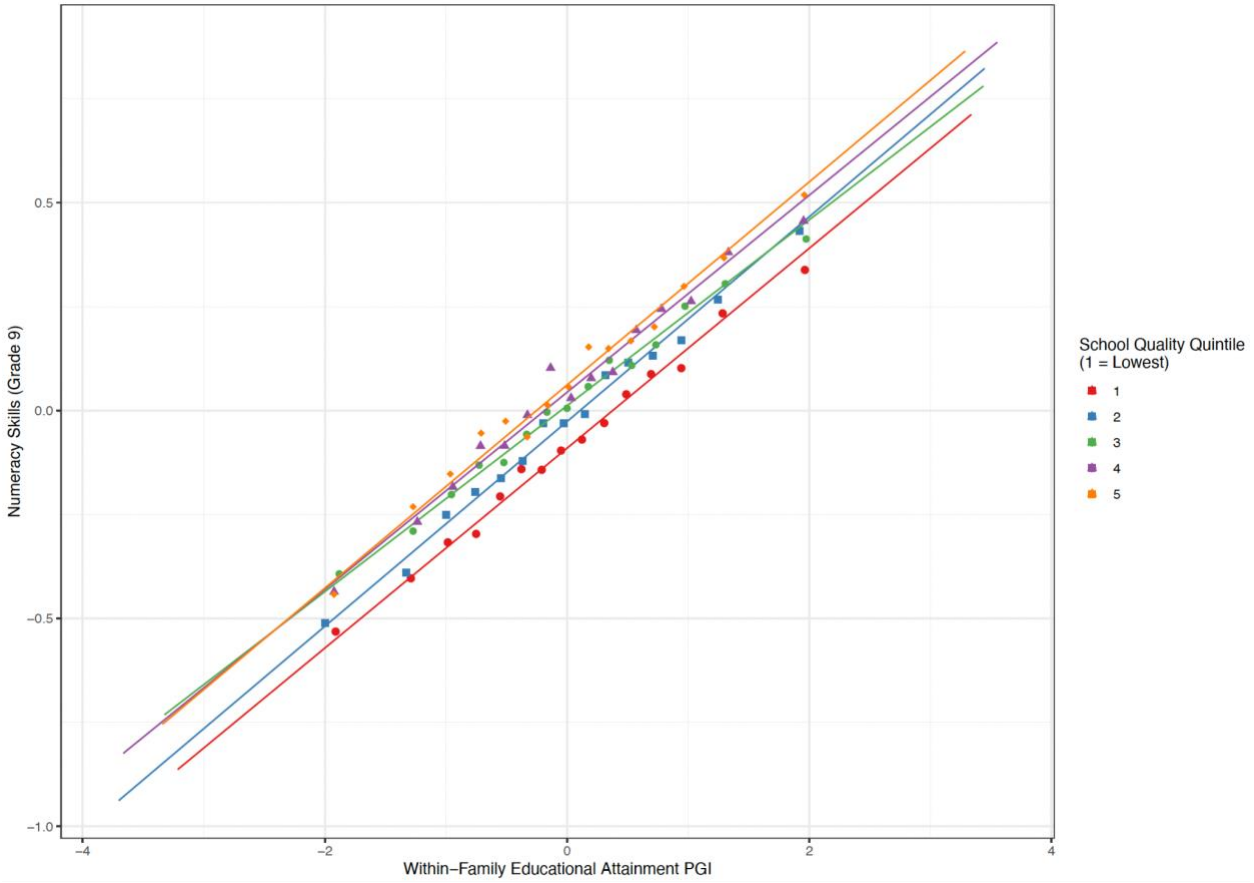

Fig. S1. Gene-environment interaction for numeracy test scores. This figure shows binned scatter plots for the relationship between PGI^{EA} and numeracy test scores in grade 9 by quintile of the VA^{d} distribution. Scatter plots are constructed by grouping the PGI^{EA} distribution into 15 bins. Regression slopes are estimated on the full data, conditioning on controls matching the focal model (see column 3, Table 2; Materials and Methods). Data: Own calculations based on MoBa and Norwegian registers.

D. Supplementary Tables

Outcome: Numeracy (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.314*** (0.006)	0.238*** (0.008)	0.239*** (0.004)	0.239*** (0.004)
$VA^{Numeracy}$	0.076*** (0.013)	0.075*** (0.013)	0.039*** (0.005)	0.040*** (0.005)
$PGI^{EA} \times VA^{Numeracy}$	-0.005 (0.007)	-0.006 (0.007)	-0.000 (0.004)	0.001 (0.005)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, VA^{d}, X)	×	×	×	✓
R^2	0.102	0.109	0.738	0.740
N	30,939	30,939	30,939	30,939
Skill persistence ρ	–	–	0.702*** (0.004)	0.703*** (0.004)

Table S1. Gene-environment interaction numeracy scores. This table shows estimates for the effects of PGI^{EA} and *VA* on children's numeracy scores in grade 9, as well as the corresponding gene-environment interaction. *Genetic controls* include the PGI^{EA} of biological mothers and fathers, and categorical variables for the genotyping batch. *School quality controls* include lagged grade 8 test scores in numeracy, literacy, English, maternal and paternal years of education, second-generation migration status, gender, birth cohort, birth order, number of siblings, and school-cohort averages of all previous controls. *2-way interactions* include all interactions of PGI^{EA} and VA^{d} with the aforementioned controls. Standard errors (in parentheses) are clustered at the school level. Skill persistence ρ indicates the estimate for lagged test scores in numeracy (grade 8), which is estimated in the model as part of the child controls. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Data: Own calculations based on MoBa and Norwegian registers.

References

1. E. Caponera, B. Losito. Context factors and student achievement in the IEA studies: evidence from TIMSS. *Large-scale Assess Educ* **4**, 12 (2016).
2. N.T. Borgen, L. J. Kirkebøen, A. Kotsadam, O. Raaum. Do Funds for More Teachers Improve Student Outcomes? *Journal of Human Resources* (2025)
3. Statistics Norway. Færre barn lever i familier med lavinntekt. [Fewer children live in families with low income.] (2024). https://www.ssb.no/inntekt-og-forbruk/inntekt-og-formue/artikler/faerre-barn-lever-i-familier-med-lavinntekt#:~:text=I%20perioden%202020%2D2022%20tilh%C3%B8rte,sammenlignet%20med%20perioden%202019%2D2021
4. A.K. Chmielewski, S.F. Reardon. Patterns of Cross-National Variation in the Association Between Income and Academic Achievement. *AERA Open*, *2*(3) (2016).
5. A.K. Chmielewski. The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, 84(3), 517-544 (2019).
6. A. M. J. Sandsør, H. D. Zachrisson, L. A. Karoly, E. Dearing.The Widening Achievement Gap Between Rich and Poor in a Nordic Country. *Educational Researcher*, *52*(4), 195-205 (2023).
7. E. C. Corfield, et al., The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. BioRxiv (2022) https:/doi.org/10.1101/2022.06.23.496289.