The genetic lottery goes to school: gene-environment interactions in the school context

Paul Hufe*

* University of Bristol

January 7, 2026

Motivation

- Education is a core determinant of life outcomes (Acemoglu and Autor, 2011; Hanushek and Woessmann, 2008; Krueger and Lindahl, 2001).

- Equity of education systems as a central policy goal:

  Most fundamental, of course, is the question of how well schools reduce the inequity of birth by providing children an equitable foundation of mental skills and knowledge [...].

  Coleman Report, p.36

- Effective education policies require understanding of the production function:

$$Y = f(\underbrace{G}_{\text{Nature}}, \underbrace{I^F, I^S}_{\text{Nurture}}).$$

Motivation

- Education is a core determinant of life outcomes (Acemoglu and Autor, 2011; Hanushek and Woessmann, 2008; Krueger and Lindahl, 2001).

- Equity of education systems as a central policy goal:

  Most fundamental, of course, is the question of how well schools reduce the inequity of birth by providing children an equitable foundation of mental skills and knowledge [...].

  Coleman Report, p.36

- Effective education policies require understanding of the production function:

$$Y = f(\underbrace{G}_{\text{Nature}}, \underbrace{I^F, I^S}_{\text{Nurture}}).$$

Why genes? Why now?

1. Genes account for 40% of variation in years of education (Branigan et al., 2013).
2. Genes matter for distributive justice (Koellinger and Harden, 2018).
3. Recent advances in molecular genetics now allow us to study the role of genes for education (Benjamin et al., 2024).

Why genes? Why now?

1. Genes account for 40% of variation in years of education (Branigan et al., 2013).

2. Genes matter for distributive justice (Koellinger and Harden, 2018).

3. Recent advances in molecular genetics now allow us to study the role of genes for education (Benjamin et al., 2024).

Why genes? Why now?

1. Genes account for 40% of variation in years of education (Branigan et al., 2013).
2. Genes matter for distributive justice (Koellinger and Harden, 2018).
3. Recent advances in molecular genetics now allow us to study the role of genes for education (Benjamin et al., 2024).

A research agenda

Research question

Do better schools increase or decrease the effect of genes on educational attainment?

- **Evidence from the US:**

  B. Arold, P. Hufe, and M. Stoeckli (in press). “Genetic Endowments, Educational Outcomes and the Moderating Influence of School Quality”. *Journal of Political Economy: Microeconomics*

- **Evidence from Norway:**

  N. T. Borgen, R.G. Cheesman, P. Hufe, A.M.J. Sandsor (2025). “The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students Genetic Differences”. *Proceedings of the National Academy of the Sciences* 122 (43), e2511715122

A research agenda

Research question

Do better schools increase or decrease the effect of genes on educational attainment?

- **Evidence from the US:**

  B. Arold, P. Hufe, and M. Stoeckli (in press). "Genetic Endowments, Educational Outcomes and the Moderating Influence of School Quality". *Journal of Political Economy: Microeconomics*

- **Evidence from Norway:**

  N. T. Borgen, R.G. Cheesman, P. Hufe, A.M.J. Sandsor (2025). "The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students Genetic Differences". *Proceedings of the National Academy of the Sciences* 122 (43), e2511715122

A research agenda

Research question

Do better schools increase or decrease the effect of genes on educational attainment?

- **Evidence from the US:**

  B. Arold, P. Hufe, and M. Stoeckli (in press). “Genetic Endowments, Educational Outcomes and the Moderating Influence of School Quality”. *Journal of Political Economy: Microeconomics*

- **Evidence from Norway:**

  N. T. Borgen, R.G. Cheesman, P. Hufe, A.M.J. Sandsor (2025). “The Genetic Lottery Goes to School: Better Schools Compensate for the Effects of Students Genetic Differences”. *Proceedings of the National Academy of the Sciences* 122 (43), e2511715122

A a closely related research agenda (that I will not talk about)

Research question

Does more schooling increase or decrease the effect of genes on educational attainment?

R. Ahlskog, J. Beauchamp, A. Okbay, S, Oskarsson, and K. Thom (2024). Testing for treatment effect heterogeneity: Educational reform, genetic endowments, and family background. Revise and Resubmit at *Nature Communications*.

S. Barcellos, L. Carvalho, and P. Turley (2021). "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages and Later-life SES". *NBER Working Paper* 28750.

A a closely related research agenda (that I will not talk about)

Research question

Does more schooling increase or decrease the effect of genes on educational attainment?

R. Ahlskog, J. Beauchamp, A. Okbay, S, Oskarsson, and K. Thom (2024). Testing for treatment effect heterogeneity: Educational reform, genetic endowments, and family background. Revise and Resubmit at *Nature Communications*.

S. Barcellos, L. Carvalho, and P. Turley (2021). "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages and Later-life SES". *NBER Working Paper* 28750.

A a closely related research agenda (that I will not talk about)

Research question

Does more schooling increase or decrease the effect of genes on educational attainment?

R. Ahlskog, J. Beauchamp, A. Okbay, S, Oskarsson, and K. Thom (2024). Testing for treatment effect heterogeneity: Educational reform, genetic endowments, and family background. Revise and Resubmit at *Nature Communications*.

S. Barcellos, L. Carvalho, and P. Turley (2021). "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages and Later-life SES". *NBER Working Paper* 28750.

A a closely related research agenda (that I will not talk about)

Research question

Does more schooling increase or decrease the effect of genes on educational attainment?

R. Ahlskog, J. Beauchamp, A. Okbay, S, Oskarsson, and K. Thom (2024). Testing for treatment effect heterogeneity: Educational reform, genetic endowments, and family background. Revise and Resubmit at *Nature Communications*.

S. Barcellos, L. Carvalho, and P. Turley (2021). "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages and Later-life SES". *NBER Working Paper* 28750.

Outline

Roadmap

Measuring genetic factors

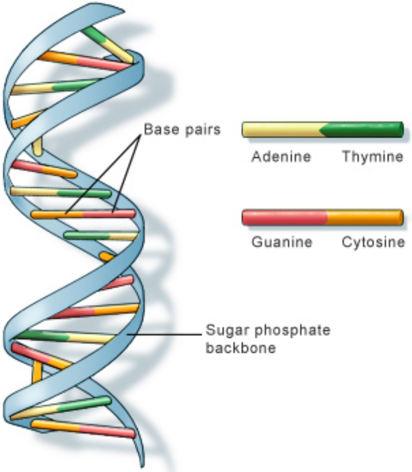
Evidence from the US

Evidence from Norway

Conclusion

Genetics 101

- Human genetic information stored in 23 chromosome pairs.
- Each chromosome consists of a molecule called DNA.
- The "rungs of the ladder" of the DNA are acid-base pairs.
- Genes are sequences of acid-base pairs that are protein-coding.
  - There are 3.3 bn "rungs in the ladder."
  - > 99.5% are the same for all human beings.



▸ Minor and major alleles ▸ Meiosis ▸ GWAS and PGI

Identification

- **Estimation**:

$$Y_i = \alpha PGI_i^{EA} + \beta Q_i + \kappa(PGI_i^{EA} \times Q_i) + \mathbf{X}_i\gamma + \epsilon_i$$

- **Identification**:

Requirement	Potential bias	Affected parameters	Potential solutions
Exogenous PGI^{EA}	indirect genetic effects	α, κ	genetic trios sibling design adoption design ...
Exogenous Q	selection into schools	β, κ	admission lotteries border discontinuities value-added estimates ...
Independent PGI^{EA}, Q	gene-environment correlation	κ	–

Roadmap

Measuring genetic factors

Evidence from the US

Evidence from Norway

Conclusion

Genetic Endowments, Educational Outcomes and the Moderating Influence of School Quality

B. Arold, P. Hufe, and M. Stoeckli

Journal of Political Economy: Microeconomics

National Longitudinal Study of Adolescent to Adult Health (Add Health)

- Initial information for a sample of adolescents ($N = 20,745$) collected in 1994/95.
- Nationally representative sample for students in grades 7-12.
- Follow up waves in 1996, 2001/02, 2008/09, 2016/18.
- We restrict the sample to students of European descent.

Data inputs

- Recall our estimation model:

$$Y_i = \alpha PGI_i^{EA} + \beta Q_i + \kappa(PGI_i^{EA} \times Q_i) + \mathbf{X}_i\gamma + \epsilon_i$$

▸ Educational outcomes Y_i ▸ Genetic factors PGI^{EA} ▸ School quality Q ▸ Control $\mathbf{X}_i$

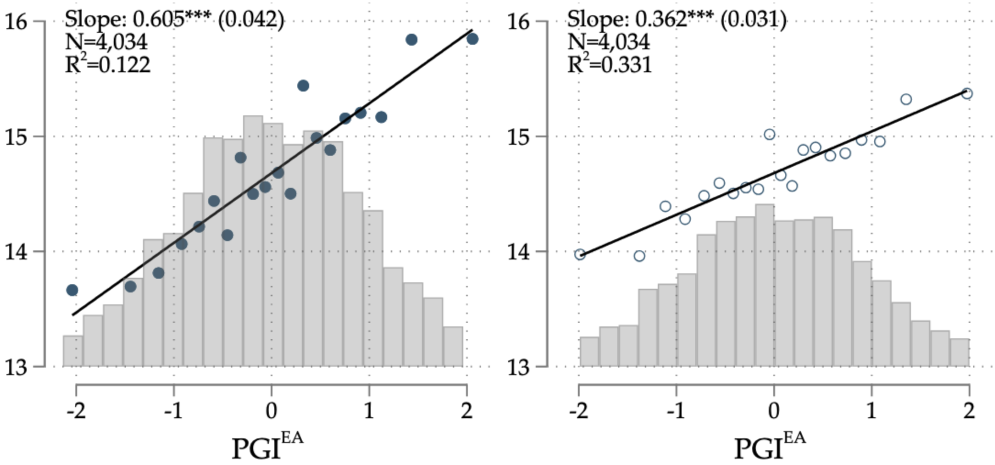
Sample

	N=4,034; High Schools=72			
	Mean	SD	Min	Max
Educational Attainment				
Years of Education	14.68	2.27	8.00	20.00
High School Degree	0.96	0.20	0.00	1.00
2-year College Degree	0.50	0.50	0.00	1.00
4-year College Degree	0.39	0.49	0.00	1.00
Post-Graduate Degree	0.14	0.35	0.00	1.00
Child and Family Characteristics				
PGI^{EA}	0.00	1.00	-4.18	3.40
Female	0.54	0.50	0.00	1.00
Firstborn	0.48	0.50	0.00	1.00
Age in Months (Wave 1)	192.41	19.62	144.00	256.00
Maternal Age at Birth	25.33	4.83	16.00	46.08
Christian	0.82	0.38	0.00	1.00
Education Mother (in Years)	13.54	2.48	0.00	19.00
Education Father (in Years)	13.56	2.68	0.00	19.00
School Quality Indicators				
Q	0.00	1.00	-2.79	1.83
Teacher w/ MA (%)	51.20	24.11	0.00	95.00
Experienced Teacher (%)	66.65	23.43	0.00	98.00
New Teacher (%)	7.88	7.28	0.00	47.00
Class Size	24.40	4.50	12.00	38.00

Recap on identifying assumptions

1. No indirect genetic effects (α, κ).
2. No selection into schools (β, κ).
3. Independent variation in PGI^{EA} and Q (κ).

Identifying genetic effects



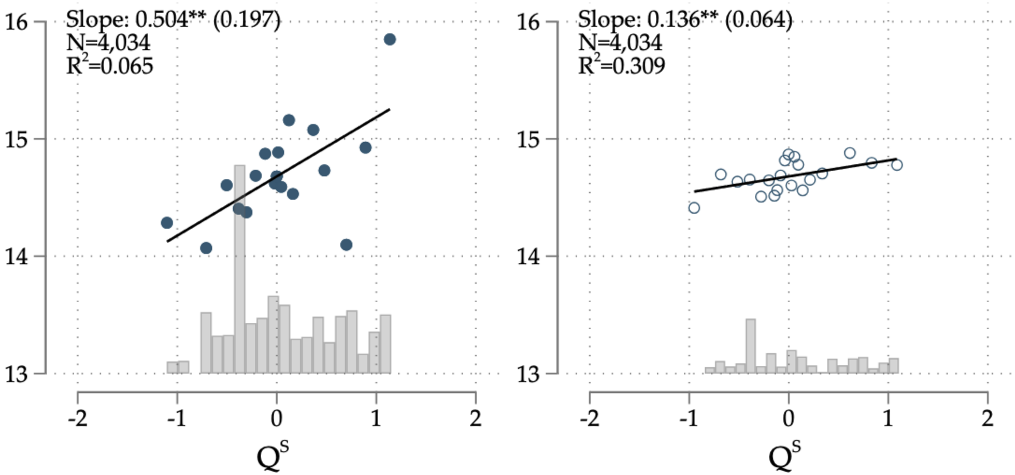
Identifying genetic effects

	Years of Education: Between- vs. Within-Family		Predicted Years of Education: w/o vs. w/ Control Function	
	(1)	(2)	(3)	(4)
PGI^{EA}	0.415^{***} (0.085)	0.432^{***} (0.141)	–	–
Q	–	–	0.264^{**} (0.129)	0.014 (0.050)
Difference in coefficients	-0.017 (0.129) [-0.269, 0.236]		0.250^{**} (0.115) [0.023, 0.476]	
Child Controls	✓	✓	×	×
Family Controls	✓	✓	×	×
Control Function	✓	✓	×	✓
Sibling Fixed Effect	×	✓	×	×
N	677	677	4, 034	4, 034
R^2	0.420	0.795	0.084	0.184
Outcome Mean	14.722	14.722	14.681	14.681
Outcome SD	2.277	2.277	1.163	1.163

Recap on identifying assumptions

- ✓ No indirect genetic effects (α, κ).
2. No selection into schools (β, κ).
3. Independent variation in PGI^{EA} and Q (κ).

Identifying school effects



Identifying school effects

	Years of Education: Between- vs. Within-Family		Predicted Years of Education: w/o vs. w/ Control Function	
	(1)	(2)	(3)	(4)
PGI^{EA}	0.415*** (0.085)	0.432*** (0.141)	–	–
Q	–	–	0.264** (0.129)	0.014 (0.050)
Difference in coefficients	-0.017 (0.129) [-0.269, 0.236]		0.250** (0.115) [0.023, 0.476]	
Child Controls	✓	✓	×	×
Family Controls	✓	✓	×	×
Control Function	✓	✓	×	✓
Sibling Fixed Effect	×	✓	×	×
N	677	677	4, 034	4, 034
R^2	0.420	0.795	0.084	0.184
Outcome Mean	14.722	14.722	14.681	14.681
Outcome SD	2.277	2.277	1.163	1.163

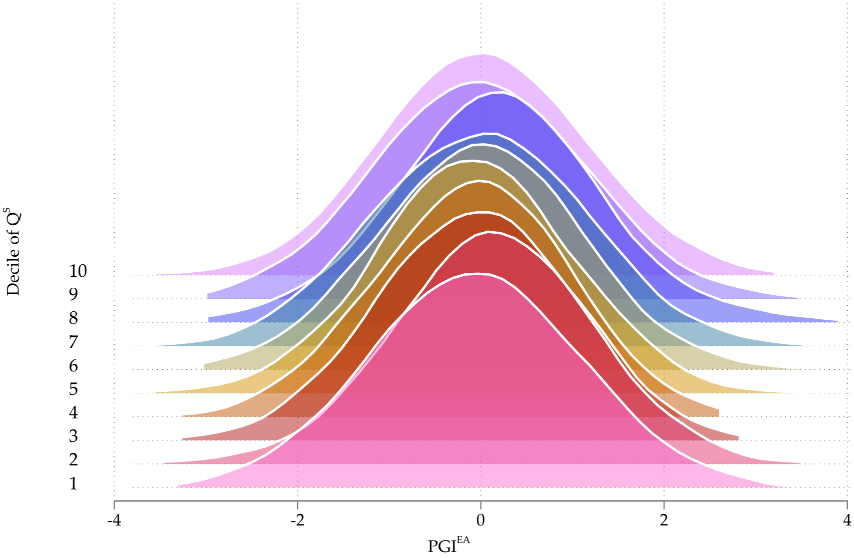
▸ Predictive power ▸ Cinelli and Hazlett (2020)

Recap on identifying assumptions

- ✓ No genetic nurture (α, κ).
- ✓ No selection into schools (β, κ).

3. Independent variation in PGI^{EA} and Q (κ).

Genes and school investments



Recap on identifying assumptions

- ✓ No genetic nurture (α, κ).
- ✓ No selection into schools (β, κ).
- ✓ Independent variation in PGI^{EA} and Q (κ).

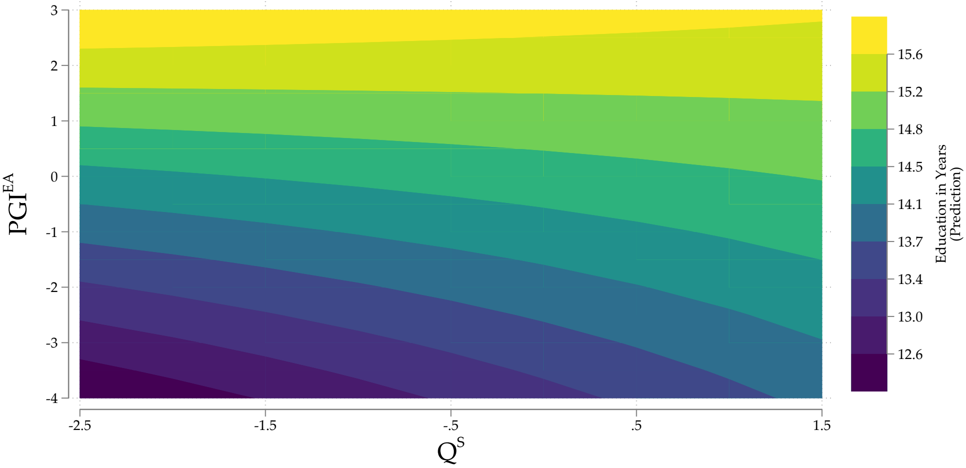
Gene-environment interaction

	Overall		Decomposition of Q			
Outcome: Years of Education	PCA (1)	Anderson (2008) (2)	(3)	(4)	(5)	(6)
PGI^{EA}	0.361^{***} (0.028)	0.361^{***} (0.029)	0.360^{***} (0.027)	0.362^{***} (0.029)	0.362^{***} (0.030)	0.362^{***} (0.030)
Q	0.124^{**} (0.057)	0.098^{**} (0.048)	–	–	–	–
$PGI^{EA} \times$ Q	-0.068^{***} (0.026)	-0.064^{**} (0.028)	–	–	–	–
Teacher w/ MA	–	–	0.166^{**} (0.071)	–	–	–
$PGI^{EA} \times$ Teacher w/ MA	–	–	-0.072^{***} (0.026)	–	–	–
Exp. Teacher	–	–	–	0.069 (0.059)	–	–
$PGI^{EA} \times$ Exp. Teacher	–	–	–	-0.045^{*} (0.026)	–	–
New Teacher	–	–	–	–	-0.020 (0.047)	–
$PGI^{EA} \times$ New Teacher	–	–	–	–	0.038 (0.029)	–
Class Size	–	–	–	–	–	-0.008 (0.044)
$PGI^{EA} \times$ Class Size	–	–	–	–	–	-0.004 (0.032)
Child Controls	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓
Control Function	✓	✓	✓	✓	✓	✓
N	4, 034	4, 034	4, 034	4, 034	4, 034	4, 034
R^2	0.333	0.333	0.334	0.332	0.332	0.331
Outcome Mean	14.681	14.681	14.681	14.681	14.681	14.681
Outcome SD	2.268	2.268	2.268	2.268	2.268	2.268

Gene-environment interaction

Outcome: Years of Education	Overall		Decomposition of Q			
	PCA (1)	Anderson (2008) (2)	(3)	(4)	(5)	(6)
PGI^{EA}	0.361^{***} (0.028)	0.361^{***} (0.029)	0.360^{***} (0.027)	0.362^{***} (0.029)	0.362^{***} (0.030)	0.362^{***} (0.030)
Q	0.124^{**} (0.057)	0.098^{**} (0.048)	–	–	–	–
$PGI^{EA} \times Q$	-0.068^{***} (0.026)	-0.064^{**} (0.028)	–	–	–	–
Teacher w/ MA	–	–	0.166^{**} (0.071)	–	–	–
$PGI^{EA} \times$ Teacher w/ MA	–	–	-0.072^{***} (0.026)	–	–	–
Exp. Teacher	–	–	–	0.069 (0.059)	–	–
$PGI^{EA} \times$ Exp. Teacher	–	–	–	-0.045^{*} (0.026)	–	–
New Teacher	–	–	–	–	-0.020 (0.047)	–
$PGI^{EA} \times$ New Teacher	–	–	–	–	0.038 (0.029)	–
Class Size	–	–	–	–	–	-0.008 (0.044)
$PGI^{EA} \times$ Class Size	–	–	–	–	–	-0.004 (0.032)
Child Controls	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓
Control Function	✓	✓	✓	✓	✓	✓
N	4,034	4,034	4,034	4,034	4,034	4,034
R^2	0.333	0.333	0.334	0.332	0.332	0.331
Outcome Mean	14.681	14.681	14.681	14.681	14.681	14.681
Outcome SD	2.268	2.268	2.268	2.268	2.268	2.268

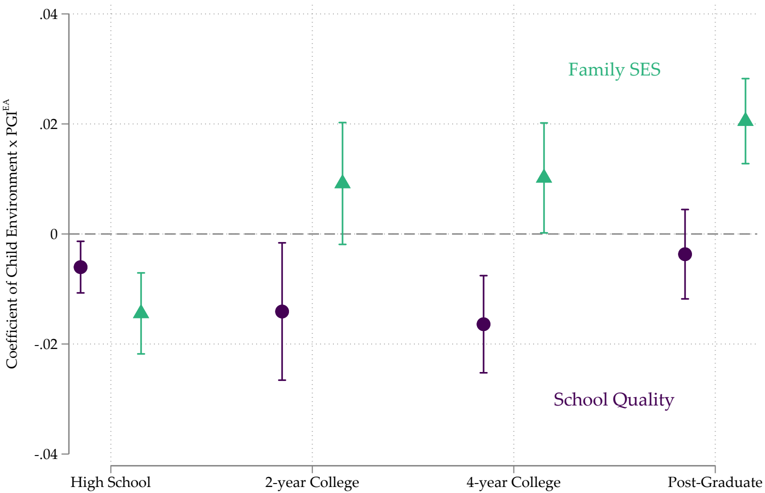
Gene-environment interaction



Robustness

- ✓ Inclusion of other school characteristics and policies, ▸ School characteristics
- ✓ Inclusion of other family/child characteristics, ▸ Family characteristics
- ✓ Inclusion of other PGI, ▸ Other PGI
- ✓ Placebo assignment to schools, ▸ Placebo
- ✓ Exclusion of outlier schools, ▸ Outlier
- ✓ No ceiling effects in educational attainment, ▸ Ceiling effects
- ✓ Sample selection and weighting criteria, ▸ Sample and weighting
- ✓ Correction for measurement error in PGI^{EA}. ▸ Becker et al. (2021)

Educational degrees



Roadmap

Measuring genetic factors

Evidence from the US

Evidence from Norway

Conclusion

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

N. T. Borgen, R.G. Cheesman, P. Hufe, A.M.J Sandsor

Proceedings of the National Academy of the Sciences

Data sources

- MoBa:
  - Initial information for a sample of mothers ($N > 114,000$) from 1999-2008.
  - 44,017 genotyped father-mother-child trios.
  - Linked to Norwegian register data.
  - We restrict the sample to birth cohorts 2002-2008 and students of European descent.
  - Effective sample size $N \approx 31,000$.
- Norwegian registers:
  - Population of students in Norway ($N \approx 60,000$ per cohort).
  - Information on standardized tests in reading and numeracy in grades 5, 8, and 9.
  - We restrict the sample to birth cohorts 1997-2007.
  - Effective sample size $N \approx 670,000$.

Data inputs

- Our estimation model:

$$Y_i = \alpha PGI_i^{EA} + \beta Q_i + \kappa(PGI_i^{EA} \times Q_i) + \mathbf{X}_i\gamma + \epsilon_i$$

▸ Educational outcomes Y_i ▸ Genetic factors PGI^{EA} ▸ School quality Q ▸ Controls $\mathbf{X}_i$

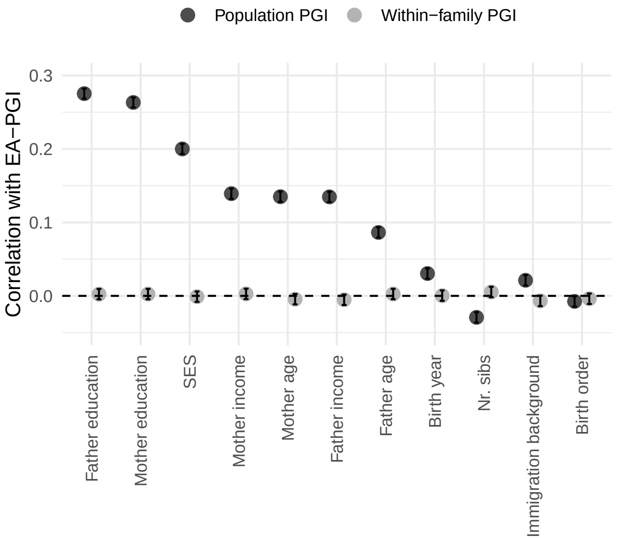
Summary statistics

	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

Recap on identifying assumptions

1. No indirect genetic effects (α, κ).
2. No selection into schools (β, κ).
3. Independent variation in PGI^{EA} and Q (κ).

Identification of genetic effects



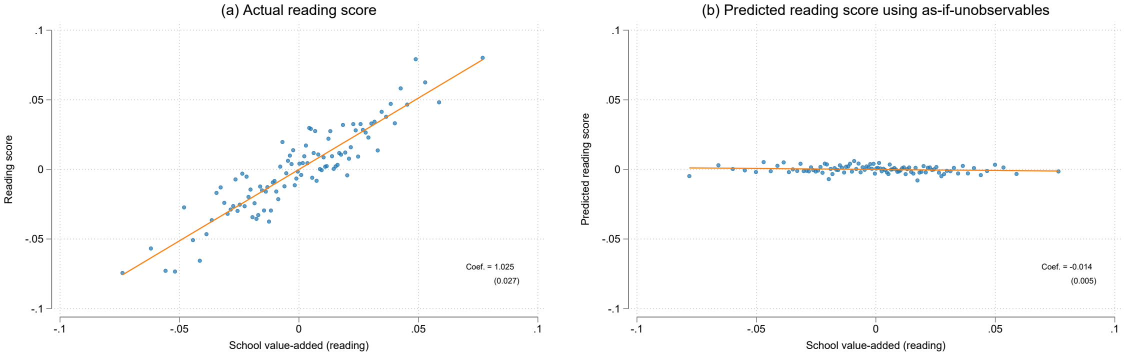
Recap on identifying assumptions

✓ No indirect genetic effects (α, κ).

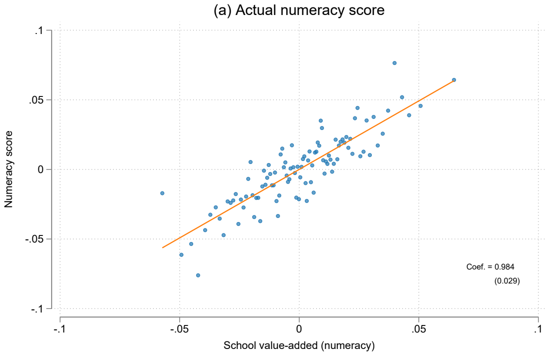
2. No selection into schools (β, κ).

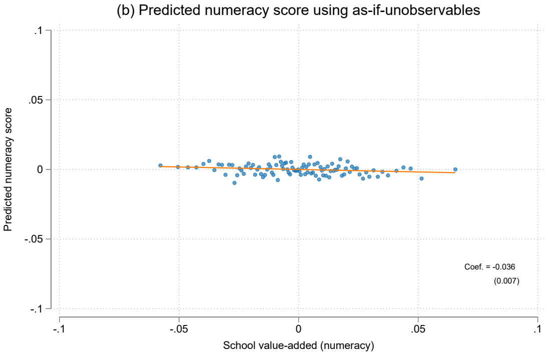
3. No gene-environment correlation in PGI^{EA} and Q (κ).

Identification of school effects (Reading)



Identification of school effects (Numeracy)





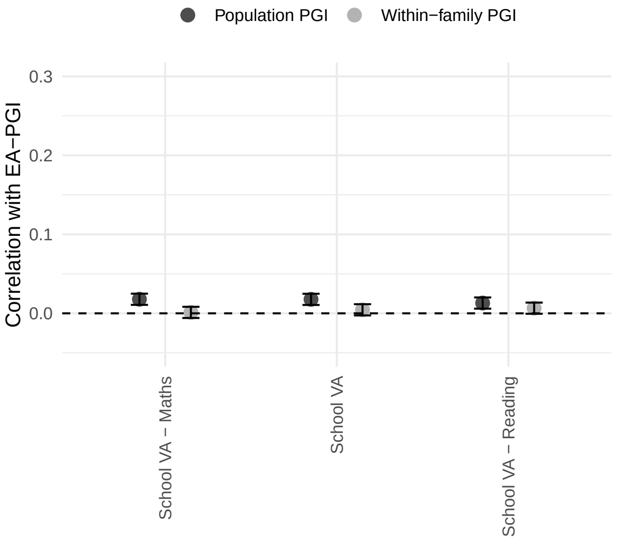
Recap on identifying assumptions

✓ No genetic nurture (α, κ).

✓ No selection into schools (β, κ).

3. No gene-environment correlation in PGI^{EA} and Q (κ).

Gene-environment correlation



Recap on identifying assumptions

- ✓ No indirect genetic effects (α, κ).
- ✓ No selection into schools (β, κ).
- ✓ No gene-environment correlation in PGI^{EA} and Q (κ).

Gene-environment interaction (Reading)

Outcome: Reading (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.304***	0.230***	0.231***	0.231***
	(0.006)	(0.008)	(0.005)	(0.005)
$VA^{Reading}$	0.091***	0.090***	0.052***	0.050***
	(0.014)	(0.013)	(0.007)	(0.007)
$PGI^{EA} \times VA^{Reading}$	-0.020*	-0.020*	-0.013*	-0.013
	(0.008)	(0.008)	(0.005)	(0.007)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, Q, 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***	0.460***
			(0.006)	(0.006)

Note: Own calculations. Standard errors (in parentheses) are clustered at the school level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Gene-environment interaction (Reading)

Outcome: Reading (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.304***	0.230***	0.231***	0.231***
	(0.006)	(0.008)	(0.005)	(0.005)
$VA^{Reading}$	0.091***	0.090***	0.052***	0.050***
	(0.014)	(0.013)	(0.007)	(0.007)
$PGI^{EA} \times VA^{Reading}$	-0.020*	-0.020*	-0.013*	-0.013
	(0.008)	(0.008)	(0.005)	(0.007)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, Q, 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***	0.460***
			(0.006)	(0.006)

Note: Own calculations. Standard errors (in parentheses) are clustered at the school level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Gene-environment interaction (Reading)

Outcome: Reading (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.304***	0.230***	0.231***	0.231***
	(0.006)	(0.008)	(0.005)	(0.005)
$VA^{Reading}$	0.091***	0.090***	0.052***	0.050***
	(0.014)	(0.013)	(0.007)	(0.007)
$PGI^{EA} \times VA^{Reading}$	-0.020*	-0.020*	-0.013*	-0.013
	(0.008)	(0.008)	(0.005)	(0.007)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, Q, 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***	0.460***
			(0.006)	(0.006)

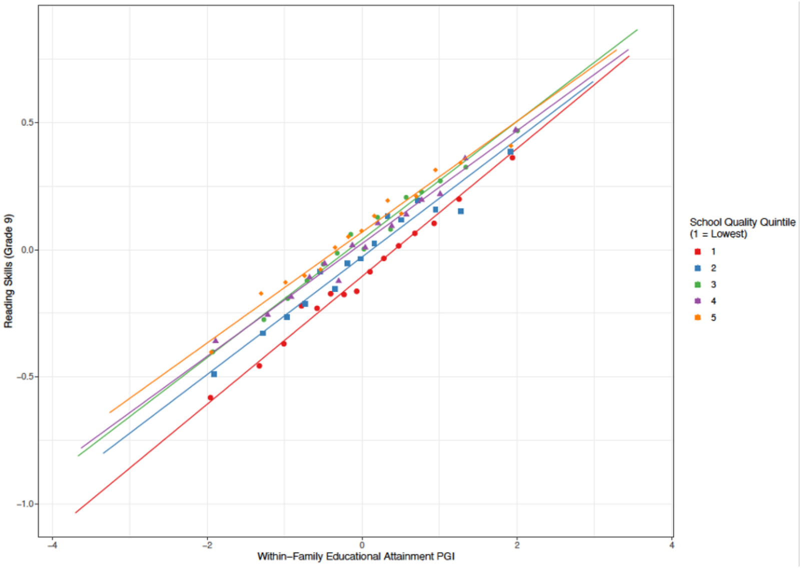
Note: Own calculations. Standard errors (in parentheses) are clustered at the school level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Gene-environment interaction (Reading)

Outcome: Reading (Grade 9)	(1)	(2)	(3)	(4)
PGI^{EA}	0.304***	0.230***	0.231***	0.231***
	(0.006)	(0.008)	(0.005)	(0.005)
$VA^{Reading}$	0.091***	0.090***	0.052***	0.050***
	(0.014)	(0.013)	(0.007)	(0.007)
$PGI^{EA} \times VA^{Reading}$	-0.020*	-0.020*	-0.013*	-0.013
	(0.008)	(0.008)	(0.005)	(0.007)
Genetic controls	×	✓	✓	✓
School quality controls	×	×	✓	✓
2-way interactions (PGI^{EA}, Q, 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***	0.460***
			(0.006)	(0.006)

Note: Own calculations. Standard errors (in parentheses) are clustered at the school level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Gene-environment interaction (Reading)

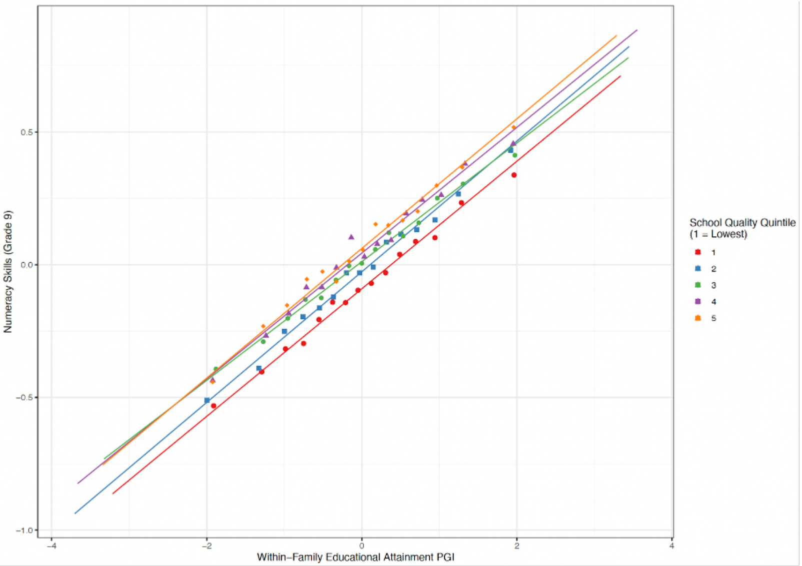


Gene-environment interaction (Numeracy)

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}, Q, 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)

Note: Own calculations. Standard errors (in parentheses) are clustered at the school level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Gene-environment interaction (Numeracy)



Contextualizing effect sizes

- Estimates pertain to a low inequality country. ▸ Inequality in VA
  - Assuming cross-country portability of effects, substitutability would be 10% for grade 9 in Chicago high schools.
- Estimates pertain to one year of schooling.
  - Assuming linear additive effects, substitutability increases to 18% over the course of lower secondary school (grades 8-10) in Norway.
- Estimates can be compared to substituability in other dimensions of advantage:
  - Latent family SES ($\Delta 1 SD$): 2.87% (Jackson et al., 2024).

Roadmap

Measuring genetic factors

Evidence from the US

Evidence from Norway

Conclusion

Summary

- **What we do**:
  - We study the (causal) interplay between PGI^{EA} and school quality.
- **What we find**:
  - Students with lower PGIs benefit more from higher-quality schools.
- **Why it matters**:
  - Investments in schools may help to students to (partially) overcome their draw in the genetic lottery and to reduce unequal opportunities in society.

Open questions

- **Generalizability**:
  - Countries
  - Learning domains
  - Age groups
  - ...

- **Mechanisms**:
  - Features of good schools?
  - Family responses as mediators?
  - ...

Open questions

- **Generalizability**:
  - Countries
  - Learning domains
  - Age groups
  - ...
- **Mechanisms**:
  - Features of good schools?
  - Family responses as mediators?
  - ...

Thank you for your attention! Questions?

✉ paul.hufe@bristol.ac.uk
www.paulhufe.net
paulhufe

References I

Acemoglu, D. and D. Autor (2011). “Skills, Tasks and Technologies: Implications for Employment and Earnings”. Ed. by D. Card and O. Ashenfelter. Vol. 4. Handbook of Labor Economics. Elsevier, pp. 1043–1171.

Ainsworth, R., R. Dehejia, C. Pop-Eleches, and M. Urquiola (2023). “Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences”. *American Economic Review* 113 (4), pp. 1049–82.

Altonji, J. G. and R. K. Mansfield (2018). “Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects”. *American Economic Review* 108 (10), pp. 2902–46.

Anderson, M. L. (2008). “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. *Journal of the American Statistical Association* 103 (484), pp. 1481–1495.

Angrist, J., P. Hull, P. A. Pathak, and C. Walters (2024). “Credible School Value-Added with Undersubscribed School Lotteries”. *Review of Economics and Statistics* 106 (1), pp. 1–19.

Angrist, J., P. Hull, and C. Walters (2023). “Methods for measuring school effectiveness”. Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 7. Handbook of the Economics of Education. Elsevier. Chap. 1, pp. 1–60.

Becker, J. et al. (2021). “Resource profile and user guide of the Polygenic Index Repository”. *Nature Human Behaviour* 5 (12), pp. 1744–1758.

Benjamin, D. J., D. Cesarini, P. Turley, and A. S. Young (2024). “Social-Science Genomics: Progress, Challenges, and Future Directions”. *National Bureau of Economic Research Working Paper Series* 32404.

Branigan, A. R., K. J. McCallum, and J. Freese (2013). “Variation in the Heritability of Educational Attainment: An International Meta-Analysis”. *Social Forces* 92 (1), pp. 109–140.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates”. *American Economic Review* 104 (9), pp. 2593–2632.

References II

Hanushek, E. A. and L. Woessmann (2008). “The Role of Cognitive Skills in Economic Development”. *Journal of Economic Literature* 46 (3), pp. 607–668.

Jackson, C. K. (2013). “Match quality, worker productivity, and worker mobility: direct evidence from teachers”. *Review of Economics and Statistics* 95 (4), pp. 1096–1116.

Jackson, C. K., S. C. Porter, J. Q. Easton, and S. Kiguel (2024). “Who Benefits From Attending Effective High Schools?” *Journal of Labor Economics* 42 (3), pp. 717–751.

Jackson, C. K., S. C. Porter, J. Q. Easton, A. Blanchard, and S. Kiguel (2020). “School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment”. *American Economic Review: Insights* 2 (4), pp. 491–508.

Kirkebøen, L. J. (2022). “School Value-Added and Long-Term Student Outcomes”. *CESifo Working Paper* 9769.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). “Experimental Analysis of Neighborhood Effects”. *Econometrica* 75 (1), pp. 83–119.

Koellinger, P. D. and K. P. Harden (2018). “Using nature to understand nurture”. *Science* 359 (6374), pp. 386–387.

Krueger, A. B. and M. Lindahl (2001). “Education for Growth: Why and for Whom?” *Journal of Economic Literature* 39 (4), pp. 1101–1136.

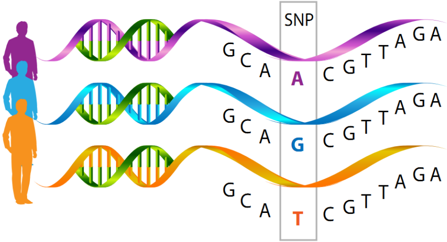
Lee, J. et al. (2018). “Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals”. *Nature Genetics* 50 (8), pp. 1112–1121.

Okbay, A. et al. (2022). “Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals”. *Nature Genetics* 54 (4), pp. 437–449.

Back

Minor and major alleles

- Differences in base pairs across humans: single-nucleotide polymorphism (SNP).
- Some of "rungs in the ladder" are more common than others. They are called minor (major) alleles.
- At each location individuals can have 0, 1, 2 minor (major) alleles.



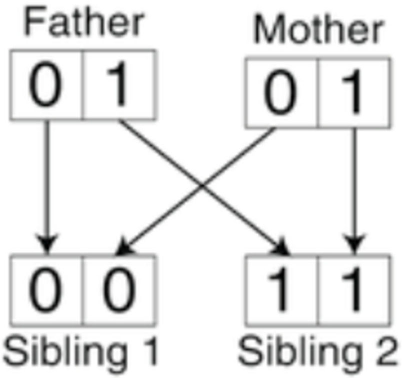
Back

Meiosis

- Parents pass one randomly selected allele to their offspring.
  - Recombination: Parental chromosome pairs cross a random number of times at random loci.
  - Mendelian segregation: For each parent, one of the recombined chromosome pairs is randomly transmitted to the germ cell.

→ Siblings can end up having different alleles from both parents at a SNP.

→ Conditional on the parental genotypes, offspring alleles at a SNP are random.



Back

GWAS and PGI

- Genetic discovery based on genome-wide association studies (GWAS):

$$y_i = SNP_{ij}\beta_j + \sum_{1}^{n} C_i^n \delta_j^n + u_i.$$

- Summary statistics via polygenic scores (PGI):

$$PGI_i = \sum_j SNP_{ij}\hat{\beta}_j.$$

Back

Educational outcomes Y_i

Years of Education

Degree Attainment

High School

College

Postgraduate Degree

Skills

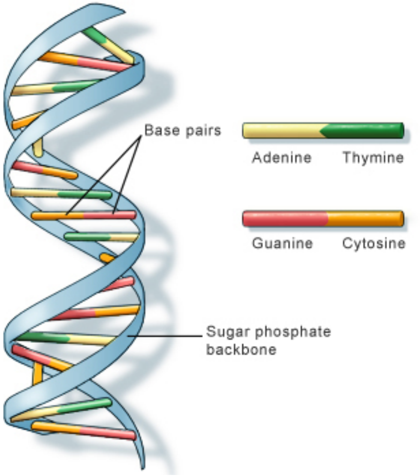
Cognitive

Preferences/Personality

Health

Genetic factors PGI^{EA}

- We use the polygenic index (PGI) for educational attainment from Lee et al. (2018):
  - Discovery sample of 1.1 mn people of European descent.
  - Explains 11% of variation in years of education.



Back

School quality Q

- We use survey and administrative information on teachers:
  - Student-teacher ratio
  - Teacher w/ tenure < 1 year
  - Teacher w/ tenure > 5 years
  - Teacher w/ Master degree

- We aggregate information using PCA or through linear aggregation of standardized variables (Anderson, 2008; Kling et al., 2007).

Back

Control $\mathbf{X}_i$

- Pre-determined characteristics and control function (Altonji and Mansfield, 2018):

Family

- maternal age at birth
- years of education (m/f)
- non-US born (m/f)
- av. potential wage (m/f)
- SD potential wage (m/f)
- religion
- state FE

Child

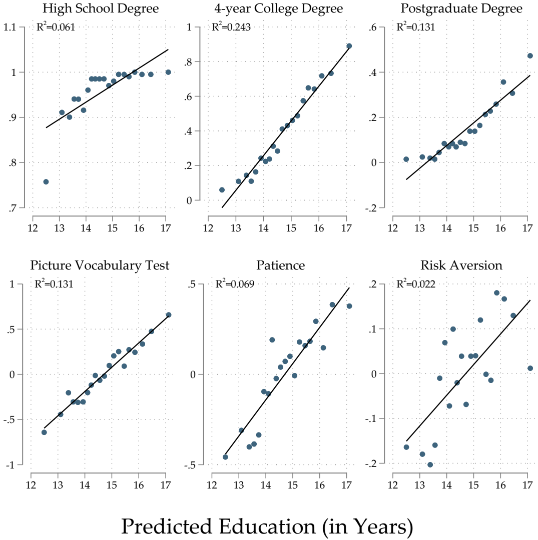
- firstborn
- gender x age in months
- 20 PC of full matrix of genetics data

Control function

- white peers (%)
- single mothers (%)
- education mothers (av.)
- female peers (%)
- migrant peers (%)

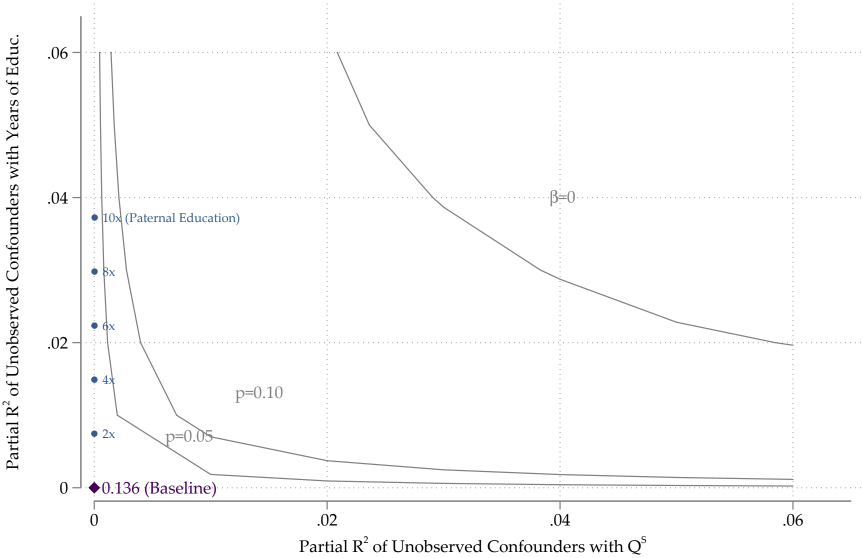
Back

Predictive power



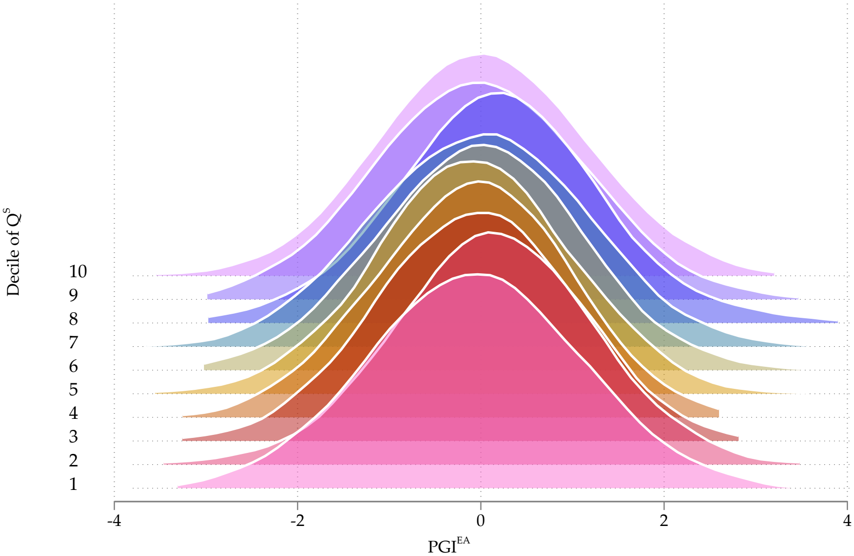
Back

Cinelli and Hazlett (2020)



Back

Residualized distributions



Back

School characteristics

Outcome: Years of Education	Baseline	+ School Policies			+ Private School	+ Teacher Composition		+ School FE
		Retention Policy	Ability Groups	Strict. Index		White Teacher	Female Teacher	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PGI^{EA}	0.361^{***} (0.028)	0.366^{***} (0.029)	0.361^{***} (0.028)	0.362^{***} (0.029)	0.362^{***} (0.029)	0.361^{***} (0.028)	0.360^{***} (0.028)	0.350^{***} (0.029)
Q	0.124^{**} (0.057)	0.116^{**} (0.053)	0.127^{**} (0.058)	0.135^{**} (0.060)	0.144^{**} (0.056)	0.121^{**} (0.055)	0.135^{**} (0.062)	–
$PGI^{EA} \times Q$	-0.068^{***} (0.026)	-0.066^{***} (0.025)	-0.068^{***} (0.025)	-0.064^{**} (0.027)	-0.076^{***} (0.026)	-0.064^{**} (0.026)	-0.065^{**} (0.026)	-0.064^{**} (0.027)
School Characteristic	–	-0.103^{*} (0.060)	0.049 (0.034)	0.062^{*} (0.036)	0.101^{**} (0.043)	-0.013 (0.074)	-0.022 (0.048)	–
$PGI^{EA} \times$ School Characteristic	–	0.034 (0.030)	-0.016 (0.030)	0.019 (0.024)	-0.049^{**} (0.023)	-0.029 (0.033)	0.045 (0.034)	–
Child Controls	✓	✓	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓	✓	✓
Control Function	✓	✓	✓	✓	✓	✓	✓	✓
N	4, 034	3, 969	4, 034	4, 034	4, 034	4, 034	4, 034	4, 034
R^2	0.333	0.334	0.333	0.334	0.334	0.333	0.334	0.343

Back

Family characteristics

Outcome: Years of Education	Baseline (1)	Interacted controls (2)	Non-linearities (3)	Subsample w/ lagged ability measures (4)	(5)	(6)
PGI^{EA}	0.361*** (0.028)	0.353*** (0.029)	0.389*** (0.036)	0.333*** (0.050)	0.334*** (0.048)	0.337*** (0.048)
Q	0.124** (0.057)	0.073 (0.059)	0.139** (0.066)	0.252* (0.144)	0.254* (0.132)	0.255* (0.133)
$PGI^{EA} \times Q$	-0.068*** (0.026)	-0.072** (0.030)	-0.075*** (0.026)	-0.088* (0.047)	-0.086** (0.042)	-0.089** (0.042)
PVT	–	–	–	–	0.223*** (0.047)	0.225*** (0.047)
PVT $\times$ Q	–	–	–	–	–	-0.003 (0.045)
GPA Science	–	–	–	–	0.381*** (0.060)	0.383*** (0.060)
GPA Science $\times$ Q	–	–	–	–	–	0.069 (0.059)
GPA Math	–	–	–	–	0.283*** (0.075)	0.281*** (0.074)
GPA Math $\times$ Q	–	–	–	–	–	-0.004 (0.080)
Child Controls	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓
Control Function	✓	✓	✓	✓	✓	✓
All interactions (Q, PGI^{EA}, $\mathbf{X}$)	×	✓	×	×	×	×
2^{nd} Polynomial (Q, PGI^{EA})	×	×	✓	×	×	×
N	4,034	4,034	4,034	1,039	1,039	1,039
R^2	0.333	0.345	0.334	0.437	0.510	0.511
Outcome Mean	14.681	14.681	14.681	14.520	14.520	14.520
Outcome SD	2.268	2.268	2.268	2.309	2.309	2.309

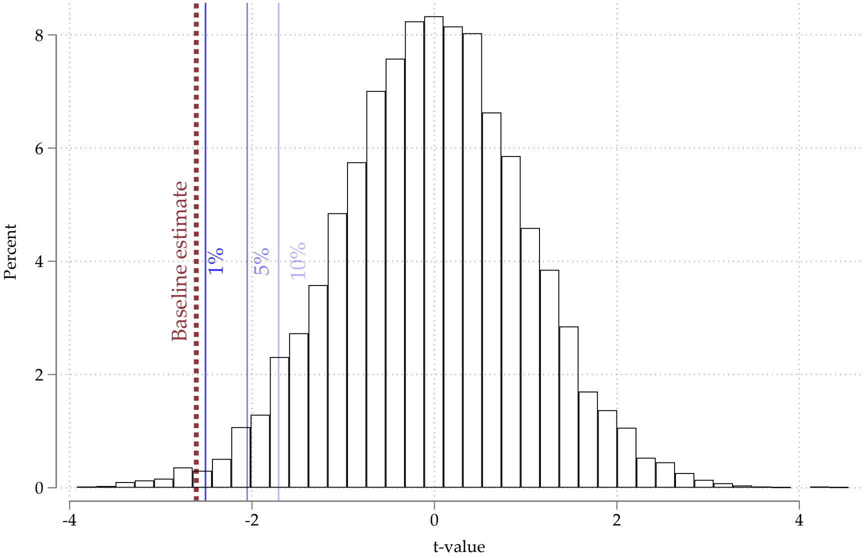
Back

Other PGI

Outcome: Years of Education	Baseline (1)	+ Controls for Other Polygenic Indexes: Body Mass Index (2)	ADHD (3)	Depressive Symptoms (4)	Intelligence (5)	Ever Smoker (6)	Sleep Duration (7)
PGI^{EA}	0.361*** (0.028)	0.341*** (0.031)	0.330*** (0.028)	0.358*** (0.028)	0.349*** (0.031)	0.341*** (0.031)	0.360*** (0.028)
Q	0.124** (0.057)	0.121** (0.056)	0.120** (0.056)	0.120** (0.057)	0.124** (0.057)	0.122** (0.056)	0.124** (0.057)
$PGI^{EA} \times Q$	-0.068*** (0.026)	-0.076*** (0.029)	-0.071*** (0.026)	-0.065** (0.027)	-0.059** (0.028)	-0.067** (0.027)	-0.068*** (0.026)
Other PGI	–	-0.080*** (0.026)	-0.132*** (0.028)	-0.039 (0.030)	0.023 (0.030)	-0.097*** (0.036)	0.026 (0.028)
Other PGI $\times$ Q	–	-0.029 (0.028)	0.003 (0.028)	0.035 (0.029)	-0.018 (0.028)	0.017 (0.033)	-0.003 (0.029)
Child Controls	✓	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓	✓
Control Function	✓	✓	✓	✓	✓	✓	✓
N	4, 034	4, 034	4, 034	4, 034	4, 034	4, 034	4, 034
R^2	0.333	0.334	0.336	0.334	0.333	0.335	0.333

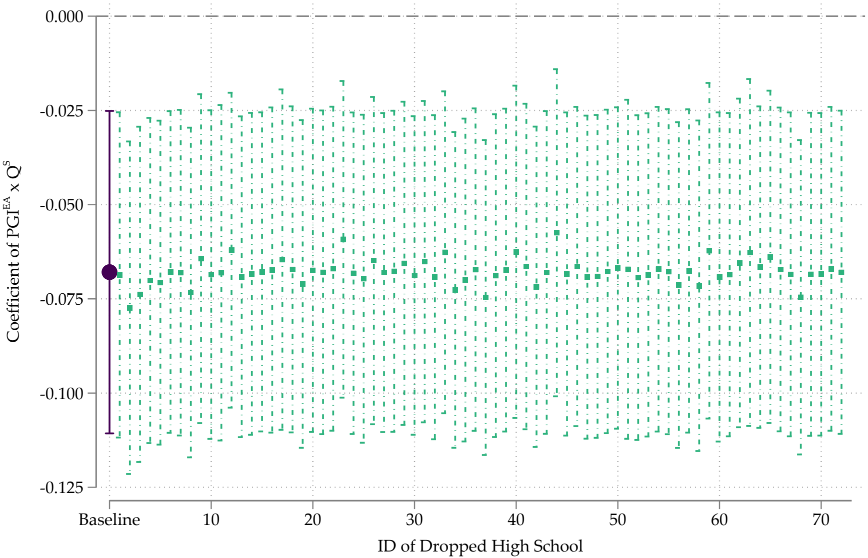
◂ Back

Placebo

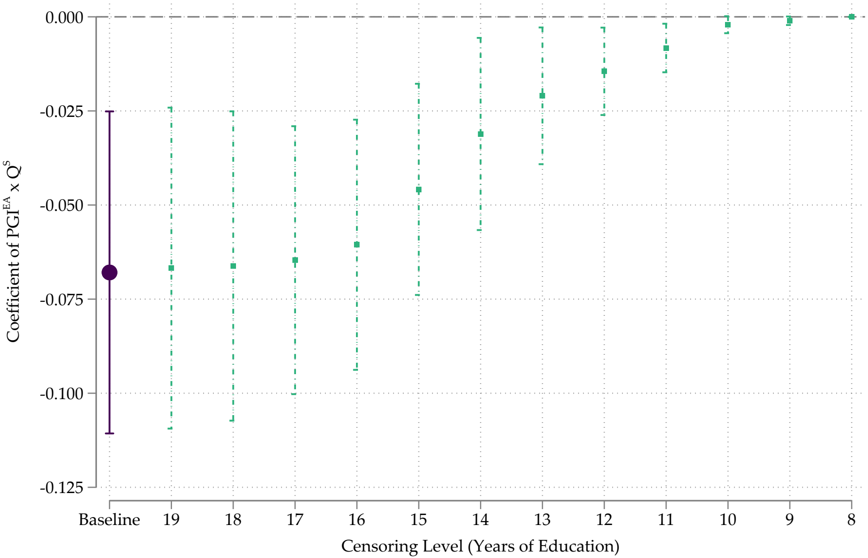


Back

Outlier



Ceiling effects



Back

Sample and weighting

Outcome: Years of Education	Baseline (1)	Alternative Sample Composition: Re-Weighted (2)	Alternative Sample Composition: Excl. (Potential) Movers before High School (3)	Alternative Sample Composition: Excl. (Potential) Movers during High School (4)
PGI^{EA}	0.361***	0.347***	0.353***	0.343***
	(0.028)	(0.031)	(0.035)	(0.040)
Q	0.124**	0.115*	0.159**	0.101
	(0.057)	(0.060)	(0.063)	(0.078)
$PGI^{EA} \times Q$	-0.068***	-0.061**	-0.086***	-0.086**
	(0.026)	(0.027)	(0.032)	(0.038)
Child Controls	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓
Control Function	✓	✓	✓	✓
N	4,034	3,968	2,962	2,439
R^2	0.333	0.313	0.350	0.344

Back

Becker et al. (2021)

	Coefficient	Standard Error	p-value	Substitutability
Baseline				
PGI^{EA}	0.361	0.032	0.000	
Q	0.124	0.064	0.052	
$PGI^{EA} \times Q$	-0.068	0.030	0.023	19%
Add Health (ρ = 1.968)				
PGI^{EA}	0.747	0.074	0.000	
Q	0.097	0.069	0.160	
$PGI^{EA} \times Q$	-0.108	0.061	0.075	15%
Health and Retirement Study (ρ = 1.413)				
PGI^{EA}	0.566	0.051	0.000	
Q	0.111	0.064	0.084	
$PGI^{EA} \times Q$	-0.093	0.043	0.030	16%
Wisconsin Longitudinal Study (ρ = 1.649)				
PGI^{EA}	0.718	0.068	0.000	
Q	0.099	0.068	0.145	
$PGI^{EA} \times Q$	-0.106	0.060	0.076	15%
UK Biobank (ρ = 1.452)				
PGI^{EA}	0.589	0.053	0.000	
Q	0.109	0.064	0.088	
$PGI^{EA} \times Q$	-0.096	0.044	0.030	16%

Back

Educational outcomes Y_i

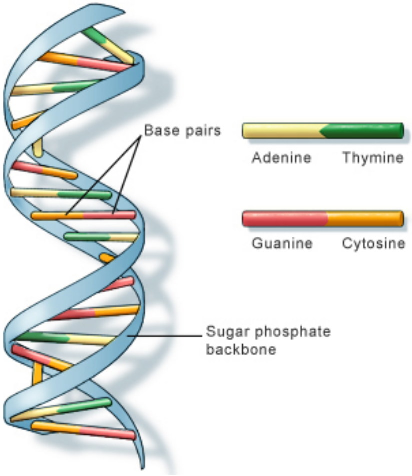
Standardized national tests in reading and numeracy (grade 9)

- Low stakes
  - $\rightarrow$ Communicated to parents and teachers but mostly used to track student development.
- Computer corrected
  - $\rightarrow$ Not affected by teacher biases.
- Taken at beginning of the school year
  - $\rightarrow$ Measure skills accumulated until grade 9.
- Same test as in grade 8
  - $\rightarrow$ Allow mapping for VA calculation.
- Highly predictive of later life-outcomes
  - $\rightarrow$ 1 SD $\uparrow$ in numeracy, increases high school graduation at age 21 by 9.5 p.p.

Back

Genetic factors PGI^{EA}

- We use the polygenic index (PGI) for educational attainment from Okbay et al. (2022):
  - Discovery sample of 3 mn people of European descent.
  - Explains 16% of variation in years of education.
  - $\approx$ 56% of explanatory power due to direct genetic effects.



Back

School quality Q

1. We construct school VA for reading and numeracy in grade 8 (Angrist et al., 2023).

2. We model educational outcomes Y of student i attending school j in cohort c for subject d:

$$Y^d_{ijc} = \beta^d Z_{ijc} + \underbrace{VA^d_{jc} + \epsilon^d_{ijc}}_{=e^d_{ijc}}$$

3. We estimate VA in subject d by averaging over residuals in school-cohort cells:

$$VA^d_{jc} = \sum e^d_{ijc}/N_{jc}$$

4. We apply the Bayesian Shrinkage estimator à la Chetty et al. (2014).

5. Highly predictive of later life-outcomes
   $\rightarrow$ 1 SD $\uparrow$ in VA, increases years of schooling by 0.5-0.8 years (Kirkebøen, 2022).

Back

Controls $\mathbf{X}_i$

Child controls

- Lagged test scores in numeracy, reading, English
- Parental years of education
- Migration status
- Age of arrival in Norway
- # of siblings
- Gender
- Year of birth
- Birth order

School controls

- School-cohort averages of all child background variables

Parental PGI

- PGI^{EA} mother
- PGI^{EA} father

Genotyping controls

- Genotyping center
- Genotyping batch
- Genotyping plate
- Imputation batch

Saturation controls

- Interaction of child background controls, school controls, and parental PGIs with PGI^{EA} and Q

Back

Inequality in VA