Genetics of Educational Attainment

Persistence of Privilege at the Turn of the 21st Century[1]

François Nielsen and J. Micah Roos

Department of Sociology
University of North Carolina
Chapel Hill

[1]Presented at the Southern Sociological Society meetings in New Orleans, LA, 23 Mar 2012

Abstract

We use structural equations methodology with data on 1,576 pairs of variously-related young adult siblings (n pairs = 200 MZ twins, 324 DZ twins, 639 full siblings, 213 half siblings, 68 cousins, and 132 nonrelated siblings) to distinguish the roles of genetic and environmental influences on educational attainment (highest degree earned). We estimate quantitative genetic (ACE) models of attainment controlling for sex and age. While many cognitive and educational outcomes show increasing effects of genes (heritability) and declining effects of the shared environment by late adolescence, we find that the role of genes in educational attainment is relatively weaker (23 percent of the variance in attainment) and the role of the shared family environment stronger (41 percent of the variance for twins and 30 percent of the variance for nontwin siblings) than is typically found for cognitive outcomes (such as IQ) at the same young adult stage in the life course. The pattern of persistent shared environmentality, especially for twins, is not accounted for by the strong degree of assortative mating in the data (parental correlation $\hat{r} = .629$) nor by direct effects of educational attainment of the siblings on each other. We conclude that the empirical pattern reflects a greater impact of available family financial resources on educational attainment than on more purely cognitive outcomes, and the persistence of substantial inequality of opportunity for educational attainment in American society at the turn of the twenty first century.

Environment, Genes and Success

Crucial patterns discovered by Blau and Duncan (1967):

- Education constitutes the main factor in both the upward mobility of individuals and the reproduction of status from generation to generation, so that when educational attainment is statistically controlled, little association between status of origin and destination remains.
- Socioeconomic success through education seems to depend to a considerable extent on characteristics of individuals that are both unmeasured and independent of social origins.

Environment, Genes and Success (cont'd)

- Mainstream social sciences have emphasized environmental characteristics of individuals and their family environment to explain educational and socioeconomic success:
  - Sociologists emphasized material resources and social and cultural capital to explain intergenerational transmission of status.
  - Economists have emphasized human capital investment decisions within families to explain educational continuation (Becker and Tomes 1979, 1986).
- Both sociological and economic research traditions have recognized the potential role of genetic endowment in educational socioeconomic achievement (Duncan, Featherman and Duncan 1972; Solon 2008), but typical research designs used by social scientists do not allow disentangling the roles of genes and the family environment.

The Quantitative Genetic Approach

- We work in a research tradition that flourished briefly in the 1970s in sociology and economics, examining the role of genes in educational and occupational achievement (Eckland 1967; Jencks et al. 1972; Taubman 1976).
- There is a contemporary resurgence of this tradition of genetic thinking in social stratification research exemplified by such works as Behrman and Taubman (1989), Guo and Stearns (2002), Solon (2008), Jencks and Tach (2006), Freese (2008), Nielsen (2006) (and many others).
- The principal motivation for assessing the role of genes in socioeconomic success is that the proportion of variance due to genes constitutes a measure of opportunity, while conversely the proportion of the variance due to the family environment (net of genetic influences) constitute a measure of social ascription (Behrman and Taubman 1989; Jencks and Tach 2006).

The Quantitative Genetic Approach (cont'd)

The quantitative genetic (a.k.a. "behavior genetic") approach was developed in the context of plant and animal breeding, later in psychology and psychiatry. It uses genetically-informative designs, such as twin and adoption studies to distinguish components of the variance in a trait ("phenotype").

- *Heritability* (a^2) is the proportion of the total variance in a trait due to genetic variation.
- *Shared environment* (c^2) is the proportion if the total variance in a trait due to environmental factors that affect siblings in the same way (e.g., social class, neighborhood, ...).
- *Unshared or unique environment* (e^2)is the proportion of the total variance due to factors that are unique to an individual sibling (e.g., a childhood disease, parental preference).

As sociologists we tend to believe that the shared environment is an all-important determinant for many traits.

Genes and Behavior – Two Patterns

- One strong pattern that has emerged is that genes have substantial effects on human traits and behaviors (Freese 2008; Turkheimer 2000).
- Turkheimer (2000: 160) has summarized the findings of many studies with his "three laws of human genetics":
  (1) "All human behavioral traits are heritable"
  (2) "The effect of being raised in the same family is smaller than the effect of the genes"
  (3) "A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families"
- It only *sounds* tongue-in-cheek!

Genes and Behavior – Two Patterns (cont'd)

- A second pattern concerns the evolution of the role of genes in IQ and other cognitive traits.
- Many studies suggest that IQ in childhood depend both on the genetic endowment of the child and the family environment. In adolescence the role of genes in IQ variation increases, and the role of the family environment fades out. Among adults most of the variation in IQ is explained by genes, none by the family environment, and the rest by idiosyncratic (unshared) environmental factors, including measurement error.
- The two patterns of the effects of genes on behavior are illustrated in the next slide.

Quantitative Genetic Estimates for Some Human Traits

Trait (average age)	a^2	c^2	e^2	Ref.
HDL Cholesterol (44)	72	0	28	(b)
Birth Weight (0)	10	55	35	(b)
Smoking [Yes/No], males (18)	66	20	14	(b)
Smoking [Yes/No], females (18)	35	55	10	(b)
Intelligence (5)	30	45	25	(b)
Intelligence (7)	50	25	25	(b)
Intelligence (10)	62	20	18	(b)
Intelligence (16)	58	0	42	(b)
Intelligence (18)	84	0	16	(b)
Intelligence (27)	86	0	14	(b)
Verbal IQ (16)	54	14	33	(c)
Grade Point Average (16)	66	0	34	(a)
College Plans (16)	60	3	37	(c)

SOURCE: (a) Calculated from Add Health siblings data. (b) Boomsma, Busjahn and Peltonen (2002) (c) Nielsen (2006)

Outline of This Paper (cont'd)

- We use a data set (AddHealth) that combines the best features of large social sciences surveys and a genetically-informative design based on twins and siblings.
- We estimate the relative importance of genes and the family environment using a comparison of six types of sibling pairs with known degrees of relatedness: identical (MZ) twins, fraternal (DZ) twins, full siblings, half siblings, cousins, and nonrelated siblings, using classical quantitative genetic models.
- Respondents are young adults (aged 24–32).
- Educational attainment is measured as highest degree earned coded on an ordinal scale and treated as continuous.

Table: Covariances and Correlations of Gender-Corrected Educational Attainment (`w4edut`) by Siblings Type

Siblings Type	N Pairs	Covariance Matrix		Correlation Matrix	
MZ – Monozygotic Twins	200	4.54391		1.0	
		2.69911	3.96422	0.6360	1.0
DZ – Dyzygotic Twins	324	4.31029		1.0	
		2.3685	4.16843	0.5588	1.0
FS – Full Siblings	639	4.18372		1.0	
		1.79389	3.8643	0.4461	1.0
HS – Half Siblings	213	4.10514		1.0	
		.757862	3.63197	0.1963	1.0
CO – Cousins	68	3.26292		1.0	
		1.39484	4.65565	0.3579	1.0
NR – Non Relateds	132	3.41589		1.0	
		1.08983	3.86927	0.2998	1.0
Total	1576				

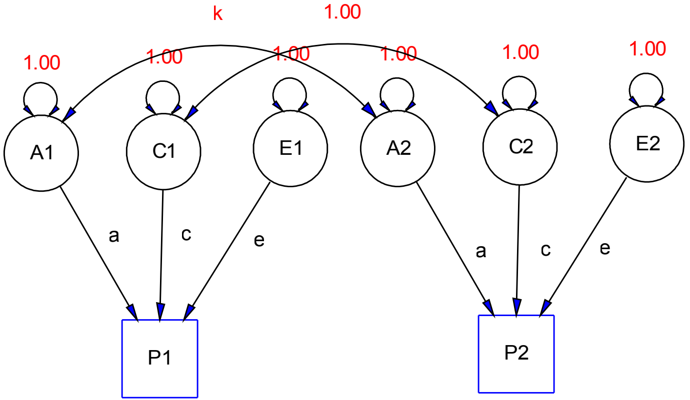

Figure: Basic ACE Model of Quantitative Genetics. Note: *k* (average proportion of genes shared) =1 (MZ twins), .5 (DZ twins and full sibs), .25 (half siblings), .125 (cousins) and 0 (nonrelated sibs)

Table: ACE Models of Educational Attainment – Estimates for Twins Only and for All Siblings Types

	Parameter Estimates			Fit Statistics				
Model	a^2	c^2	e^2	χ^2	*df*	*p*	AIC	RMSEA
Twins Only (2 groups, N = 524 pairs)								
1. ACE	.150	.484	.366	1.685	3	.640	−4.315	.007
2. CE	—[a]	.588	.412	3.917	4	.417	−4.083	.025
3. AE	.675	—[a]	.325	30.081	4	.000	22.081	.132
4. E	—[a]	—[a]	1.0	225.047	5	.000	215.047	.417
All Siblings Types (6 groups, N = 1576 pairs)								
5. ACE	.365	.272	.364	22.804	15	.088	−7.196	.032
6. CE	—[a]	.448	.552	52.108	16	.000	20.108	.090
7. AE	.696	—[a]	.304	64.591	16	.000	32.591	.112
8. E	—[a]	—[a]	1.0	404.000	17	.000	370.000	.259

Note: Parameters are a^2, additive genetic; c^2, shared environment; e^2, non-shared environment.

[a] Parameter fixed to value zero.

Table: ACE Models of Educational Attainment with Special Twins Environments (N = 1576 pairs)

	Parameter Estimates					Fit Statistics			
Model	a^2	c^2_{twi}	e^2_{twi}	c^2_{sib}	e^2_{sib}	$\chi^2(df)$	p	AIC	RMSEA
1. ACE	.228	.411	.361	.298	.474	17.514(14)	.230	−10.485	.013
2. ACE[a]	.365	.272	.364	.272	.364	22.804(15)	.088	−7.196	.023
3. CE	—[b]	.573	.427	.380	.620	25.252(15)	.047	−4.748	.023
4. AE	.696	—[b]	.304	—[b]	.304	64.591(16)	.000	32.591	.093
5. E	—[b]	—[b]	1.0	—[b]	1.0	404.000(17)	.000	370.000	.224

Note: Parameters are a^2, additive genetic; c^2_{twi}, twins shared environment; e^2_{twi}, twins non-shared environment; c^2_{sib}, nontwin siblings shared environment; e^2_{sib}, nontwin siblings non-shared environment.

[a] Model constrained so that $c^2_{\text{twi}} = c^2_{\text{sib}}$ and $e^2_{\text{twi}} = e^2_{\text{sib}}$.

[b] Parameter fixed to value zero.

Table: ML Confidence Intervals for ACE Model of Educational Attainment with Special Twins Environments (N = 1576 Pairs)

		95% Confidence Interval	
Parameter	Estimate	Lower bound	Upper bound
a^2	.228	.068	.390
c^2_{twi}	.411	.270	.538
e^2_{twi}	.361	.304	.428
c^2_{sib}	.298	.217	.376
e^2_{sib}	.474	.363	.589
$\chi^2(14)$	17.514	14.000	35.782
AIC	−10.485	−14.000	7.782
RMSEA	.013	.000	.058

Note: Parameters are a^2, additive genetic; c^2_{twi}, twins shared environment; e^2_{twi}, twins nonshared environment; c^2_{sib}, nontwin siblings shared environment; e^2_{sib}, nontwin siblings nonshared environment.

Alternative Explanation I: Assortative Mating

Table: ACE Models With Assortative Mating Fit to the Educational Achievement Data – Twins Only (N = 524 Pairs)

	Parameter Estimates			Fit Statistics				
Model	a^2	c^2	e^2	χ^2	*df*	*p*	AIC	RMSEA
1. ACE	.167	.467	.366	1.685	3	.640	−4.315	.007
2. CE	—[a]	.588	.412	3.917	4	.417	−4.083	.025
3. AE	.678	—[a]	.322	7.662	4	.105	−.338	.060
4. E	—[a]	—[a]	1.0	225.047	5	.000	215.047	.417

Note: Parameters are a^2, additive genetic; c^2, shared environment; e^2, nonshared environment. Degree of assortative mating is estimated as the parental correlation of educational attainment ($\hat{r} = .629$)

[a] Parameter fixed to value zero.

Assortative mating does not explain the high level of shared environmentality.

Alternative Explanation II: Direct Sibling Interaction

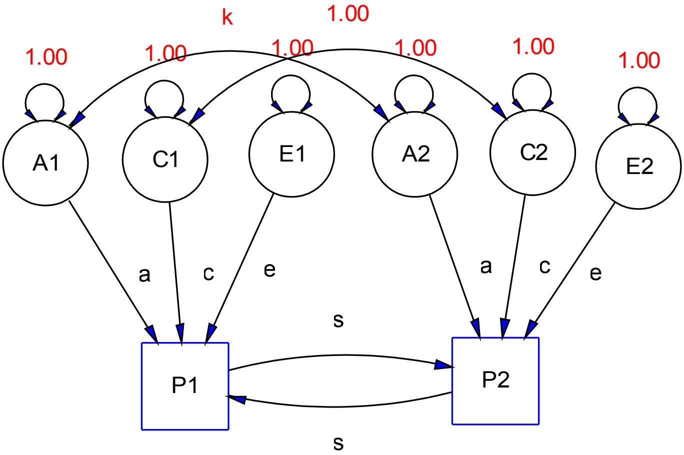

Figure: ACE Model with Direct Sibling Interaction.

Table: ACE+s Models With Siblings Interaction Fit to the Educational Achievement Data – Twins Only (N = 524 Pairs)

	Parameter Estimates				Fit Statistics				
Model	s	a^2	c^2	e^2	χ^2	*df*	*p*	AIC	RMSEA
1. ACE+s	.022	.160	.447	.393	1.684	2	.431	−2.316	.026
2. CE+s	.033	—[a]	.543	.457	3.917	3	.271	−2.083	.035
3. AE+s	.243	.257	—[a]	.744	1.922	3	.589	−4.078	.013
4. E+s	.325	—[a]	—[a]	1.0	3.917	4	.417	−4.083	.025
5. ACE	—[a]	.150	.484	.366	1.685	3	.640	−4.315	.007

Note: Parameters are *s*, phenotypic reciprocal effect; a^2, additive genetic; c^2, shared environment; e^2, nonshared environment.
[a] Parameter fixed to value zero.

Allowing for direct siblings interaction does not improve the fit of the model.

Results Summary

Table: Preferred Model Summary (N = 1576 Pairs)

	Variance Components		
Sibling pair type	a^2	c^2	e^2
Twins	.228	.411	.361
Non-twin siblings	.228	.298	.474

Note: Parameters are a^2, additive genetic; c^2, shared environment; e^2, nonshared environment

Discussion and Conclusion

- We have found a relatively weak role of genetic factors in educational attainment of these young adults (23 percent of the variance in the preferred model), and a relatively stronger role of shared environment factors (41 percent of the variance for twins; 30 percent for nontwin siblings).
- Estimated heritability of 23% is less than in many prior studies of educational attainment (e.g., Behrman and Taubman 1989 based on extensive kinship data, $a^2 = .81$); conversely, the role of the shared environment is greater than in previous studies.
- This finding is at variance with Turkheimer's "laws", as the role of genes is less than the role of the (shared) environment.
- It also does not fit the pattern of declining role of the environment over the life course found for IQ and related cognitive measures.

Discussion and Conclusion (cont'd)

What Does It All Mean?

- Behrman and Taubman write: "The share of the observed variation in schooling that is attributable to across-family variability in environment [c^2] provides a measure of inequality in schooling opportunity" (1989: 1426, cited in Miller, Mulvey and Martin 2001: 214; see also Guo and Stearns 2002; Nielsen 2006; Jencks and Tach 2006; Visscher, Hill and Wray 2008).
- The shared environment c^2 reflects the potential effect on educational attainment of raising the quality of the most disadvantaged family environment to the level of the most advantaged ones. It thus represents an upper bound on improvement in the trait achievable by policy intervention *within the existing range of environmental variation* (Rowe 1994).

Discussion and Conclusion (cont'd)

What Does It All Mean?

- Our finding of a large shared environment component in educational attainment suggests that inequality in educational opportunity is persisting (and has perhaps even increased) at the turn of the twenty-first century.
- Reasons may include:
  - Increasing cost of higher education, such that family financial circumstances become a more important factor in contrast with individual cognitive abilities.
  - The imprecision of the educational attainment measure, as it does not incorporate a quality dimension.
- If confirmed, results suggest there is much room for improving equality of educational opportunity using *traditional policy means* (e.g., financial aid, etc.) of equalizing access to higher education.

Puzzles for Future Work

- How does the quantitative genetics compare for other stages in the socioeconomic achievement process (earnings, occupational prestige, self-assessed SES, ...)?
- Does the quantitative genetic model differ according to the social status of the family of origin ($G \times E$ interaction)?
- If there is a $G \times E$ interaction, is it monotonic (Scarr-Salapatek 1971) or curvilinear (Pareto 1909; see Nielsen 2008)?
- Compare estimates of the quantitative genetic model for different educational transitions *à la* Mare (1980) (but sample sizes decrease as later transitions are considered).

References I

Alderson, Arthur S. and François Nielsen. 2002. "Globalization and the Great U-Turn: Income Inequality Trends in 16 OECD Countries." *American Journal of Sociology* 107:1244–1299.

Blau, Peter M. and Otis Dudley Duncan. 1967. *The American Occupational Structure.* New York: Wiley.

Boomsma, Dorret, Andreas Busjahn and Leena Peltonen. 2002. "Classical Twin Studies and Beyond." *Nature Reviews Genetics* 3: 11 (November): 872–882.

Breen, Richard and Jan Jonsson. 2005. "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility." *Annual Review of Sociology* 31: 223–243.

Duncan, Otis Dudley, David L. Featherman, and Beverly Duncan. 1972. *Socioeconomic Background and Achievement.* New York: Seminar Press.

Fisher, Ronald A. 1918. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society Edimburgh* 52, Part 2: 399–433.

Freese, Jeremy. 2008. "Genetics and the Social Science Explanation of Individual Outcomes." *American Journal of Sociology* 114 Suplement: S1–S35.

Heath, A. C., K. Berg, L. J. Eaves, M. H. Solaas, L. A. Corey, J. Sundet, P. Magnus, and W. E. Nance. 1985. "Education Policy and the Heritability of Educational Attainment." *Nature* 314 (25 April): 734–736.

References II

Hout, Michael and Thomas A. DiPrete. 2006. "What We Have Learned: RC28's Contributions to Knowledge about Social Stratification." *Research in Social Stratification and Mobility* 24: 1–20.

Jencks, Christopher and Laura Tach. 2006. "Would Equal Opportunity Mean More Mobility?" Pp. 23–58 in *Mobility and Inequality: Frontiers of Research from Sociology and Economics,* edited by Stephen Morgan, David Grusky and Gary Fields. Stanford, CA: Stanford University Press.

Moller, Stephanie, Arthur S. Alderson, and François Nielsen. 2009. "Changing Patterns of Income Inequality in U.S. Counties, 1970–2000." *American Journal of Sociology* 114: 4 (January): 1037–1101.

Nielsen, François. 2004a. "The Ecological-Evolutionary Typology of Human Societies and the Evolution of Social Inequality." In *Essays in Honor of Gerhard Lenski* [special issue], edited by Bernice McNair Barnett. *Sociological Theory* 22:292–314.

Nielsen, François. 2006. "Achievement and Ascription in Educational Attainment: Genetic and Environmental Influences on Adolescent Schooling." In *The Linking of Sociology and Biology* [special section], edited by Guang Guo. *Social Forces* 85: 1 (September): 193–216.

Nielsen, François. 2008. "The Nature of Social Reproduction: Two Paradigms of Social Mobility." Symposium / Social Sciences and Natural Sciences – What Connection?, edited by Mario Lucchini and Maurizio Pisati. *Sociologica* Nr 3 / 2008 (November): Pp. 1–35. [Online at `http://www.sociologica.mulino.it/`.]

References III

Pareto, Vilfredo. 1909. *Manuel d'économie politique.* [Manual of Political Economy.] Paris: Giard et Brière.

Rowe, David C. 1994. *The Limits of Family Influence: Genes, Experience, and Behavior.* New York: Guilford.

Scarr-Salapatek, Sandra. 1971. "Unknowns in the IQ Equation." *Science* 174: 1223–1228.

Solon, Gary. 2008. "Intergenerational Income Mobility." Pp. 481–485 in *Class, Race, and Gender in Sociological Perspective* (3e), edited by David B. Grusky. Westview Press.

Turkheimer, Eric. 2000. "Three Laws of Behavior Genetics and What They Mean." *Current Directions in Psychological Science* 5: 160–164.

Quick Intellectual Autobiography

- As an undergraduate at University of Brussels I wrote a thesis on Vilfredo Pareto
- I was intrigued by Pareto's discovery that the distribution of income obeyed a power law ("Pareto distribution")
- Distributions of income of different societies look like crystals formed from the same chemical substance:
  - there are large ones and small ones, but they all have the same shape
- Regularity of shape suggests deep, unseen mechanisms generating the distribution of income (and more generally the social structure) in all societies, even though quantitative parameters may vary

Title

- I also became