Antecedents of Educational Attainment: Parental Involvement, Genetics, and Non-Cognitive Skills

Student:
Domante Gerciene

Supervisors:
Melinda C. Mills
David M. Brazel



A thesis presented for the degree of
Doctor of Philosophy in Sociology

Word count: approx. 50,000 words

Department of Sociology
University of Oxford

Trinity Term 2021

Abstract

Education-driven social inequalities are increasing in the advanced countries. Educational attainment is to large extent driven by one's early social circumstances and then in turn has a significant impact on one's later life chances. Sociologists have been successful at documenting many processes relating to educational attainment. However, one important antecedent has not been fully integrated into the sociological toolbox of mechanisms - genetic transmission of educational propensity. My thesis aims to use the methodological advances from the field of social science genetics, namely the educational attainment polygenic score, to better understand how disparities in genetic predispositions, through their interplay with social environments, contribute to the divergence of educational and occupational outcomes.

Over the course of three empirical chapters, I explore the dynamic influence of genetic propensity to education over the life course: from adolescence to early and, eventually, late adulthood. Using the Add Health dataset, I investigate the relationship between parental involvement and child's genetic potential and how it shapes educational outcomes. The Wisconsin Longitudinal Study helps me to look under the hood of the educational attainment polygenic score by allowing me to investigate the mediating role of non-cognitive skills. I also look into how the complex interconnection of social background, genetics, and education influences occupational attainment.

This thesis makes several contributions to the existing body of research in the fields of social stratification and social science genetics. First of all, I embed the genetic mechanism in the broader sociological framework of attainment. My hypotheses and the construction of variables are theory-driven. I qualify my general findings by exploring their heterogeneity by gender and socio-economic status. Second, I open up the black box of the educational attainment polygenic score, which is widely used in the literature but whose true meaning remains elusive. I study the mediating mechanisms in the form of non-cognitive skills which can help one realise their genetic potential. Third, I show that genetic propensity to education dynamically interacts with social environments and spans the whole life course.

Acknowledgements

I would not describe my DPhil experience as particularly easy or enjoyable. Although I never seriously contemplated abandoning this endeavour, I doubt I would have written this thesis without the extraordinary support, help, and encouragement I received from many people.

I would like to thank my supervisors Prof. Melinda Mills and Dr. David Brazel. It goes without saying that Prof. Mills is a real role model. I still don't fully understand how she does it but I hope I can now emulate some of her best practices. Dr. Brazel's patience with my unceasing requests to run some additional analyses on the UNC cluster requires a special mention. Our bi-monthly Sociogenome reading group with Evelina Akimova, Xuejie Ding, Riley Taiji, Felix Tropf, Stine Mollegard, to name a few, served as a safe space to share my ideas, present preliminary work and seek feedback (which was always constructive and extremely kind).

I would also like to express my gratitude to Nuffield College. Its generosity not only funded my DPhil but also left an indelible mark on my bookshelves as my research allowance to a large extent was dedicated to expanding my collection of research books. Every member of the Nuffield's staff made me feel at home. Nothing can beat Nuffield's lunches. And my office was objectively the best in college. What a view!

I am also hugely grateful to ESRC for awarding me the Advanced Quantitative Methods scholarship to fund my research. It is probably the best social science funding scheme in the UK. During the low moments I would remind myself that I should feel lucky that someone believes in my potential.

The Department of Sociology at Oxford also proved to be a home away from home. I managed to tackle any recurring administrative chores only because of the help from Kristiana Dahl, Natasha Cotton, Khalid Omer, and Maria Moreno. The department's master programme informed how I approach sociology and its academic community continued to shape my thinking throughout my DPhil. I would like to thank my Transfer of Status, Confirmation of Status, and future Viva Voce assessors — Prof. Jan O. Jonsson, Prof. Nan Dirk de Graaf, Dr. Charles Rahal, and Prof. Nicola Barban — for all the advice I received and will receive. A special thanks goes to Prof. Richard

Breen who was my college advisor and who patiently listened to my often incoherent musings about the academic world.

I consider myself extremely lucky to have found amazing friends in Oxford. Roberto, Said, Ruairi, Peer - who said discussing politics until 3am is not cool? Artur - you are a true cinephile and thanks to you I watched some real masterpieces (Das Boot is very high on that list). My girls' lunch club - Fijnanda, Florianne, Victoria, Marta, and Sarah - was the best part of my day. Fijnanda - you are my greatest inspiration; you could row and do your PhD and I could do neither of those things. Florianne - your sense of humour and passion for shoes is everything. Victoria - your never-ending pipeline of cat memes always cheered me up. Marta - I could switch from crying to laughing in your office in 5 minutes. Sarah - your sense of calm and maturity is exemplary. And Hannah - you fed me, you empathised with me, and you brought me to the best piano concerts.

And nothing would have been possible without my family and friends in Lithuania. I used every possible opportunity to sneak out of Oxford and travel back home. The six months I spent there after the Covid pandemic started gave me so much perspective and helped me reconsider my life choices so that in the future I can be closer to those dear to me. Novile, Dovile, Andre, Auguste, Viktorija, Indre, Milda, Goda B., Goda R., Ryte, Gabriele - I could not ask for better friends. I am also forever grateful for the unconditional love showed by my mom and grandma. Life was never easy for you but you made it much easier for me. And to Arnas - as we embark on our forever together I thank you for all the math help I received from you starting in high school (which proved invaluable even in social sciences), an absolutely egalitarian division of household labour, and for tolerating my emotional outbursts during a painful episode of Covid-related wedding postponement.

Contents

1 Chapter 1. Introduction **1**
- 1.1 Education in the age of rising inequality . . . 2
- 1.2 Research on social stratification in education: towards the understanding of mechanisms . . . 3
- 1.3 Social science genetics: what we know so far . . . 5
  - 1.3.1 Behaviour genetics . . . 5
  - 1.3.2 Molecular genetics . . . 6
- 1.4 Genetic mechanisms in educational stratification: overview of the thesis . . . 10
  - 1.4.1 Chapter 2: Parental involvement . . . 11
  - 1.4.2 Chapter 3: Cognitive and non-cognitive skills . . . 13
  - 1.4.3 Chapter 4: Occupational attainment over the life course . . . 14
- 1.5 Data & methods . . . 17
  - 1.5.1 Data and samples . . . 17
  - 1.5.2 Utilisation of polygenic scores . . . 19
- 1.6 Summary . . . 21

2 Chapter 2. Parental Involvement: Correlates and Consequences **22**
- 2.1 Introduction . . . 24
- 2.2 Theoretical framework . . . 26
  - 2.2.1 Definition of parental involvement . . . 26
  - 2.2.2 Dimensions of parenting practices . . . 27
  - 2.2.3 Consequences and effectiveness of parenting practices . . . 29
  - 2.2.4 Determinants of parental involvement . . . 30
- 2.3 Data . . . 37
  - 2.3.1 Dataset . . . 37
  - 2.3.2 Sample . . . 38
- 2.4 Measurement . . . 39
  - 2.4.1 Parental involvement . . . 39
  - 2.4.2 Educational attainment . . . 42
  - 2.4.3 Family characteristics . . . 42
  - 2.4.4 Individual characteristics . . . 43
- 2.5 Analytical strategy & methods . . . 47
  - 2.5.1 Explaining parental involvement . . . 47
  - 2.5.2 Explaining educational attainment . . . 48
- 2.6 Results . . . 50
  - 2.6.1 Explaining parental involvement . . . 50
  - 2.6.2 Explaining educational attainment . . . 55

2.7 Conclusion/Discussion . . . 60
2.A Appendix . . . 67

3 Chapter 3. Genetic Predisposition to Education: The Role of Cognitive and Non-Cognitive Skills 84
3.1 Introduction . . . 86
3.2 Theoretical framework . . . 88
3.2.1 Genetic predisposition to educational attainment . . . 88
3.2.2 Potential mediators: cognitive skills and non-cognitive Skills . . . 89
3.2.3 Heterogeneity by gender and SES . . . 94
3.3 Data . . . 96
3.3.1 Dataset . . . 96
3.3.2 Sample size . . . 97
3.4 Measurement . . . 97
3.4.1 Outcome variable — educational attainment . . . 97
3.4.2 Explanatory variable — cognitive skills . . . 98
3.4.3 Explanatory variables — non-cognitive skills . . . 99
3.4.4 Educational attainment polygenic score . . . 102
3.4.5 Covariates . . . 103
3.4.6 Principal components . . . 104
3.5 Analytical strategy & methods . . . 105
3.6 Results . . . 107
3.6.1 Mediating role of non-cognitive skills . . . 107
3.6.2 Heterogeneity of mediation by gender . . . 113
3.6.3 Heterogeneity of mediation by socio-economic status . . . 115
3.7 Discussion/Conclusion . . . 116
3.A Appendix . . . 121

4 Chapter 4. The Role of Genetic Propensity to Education for Status Attainment 146
4.1 Introduction . . . 148
4.2 Theoretical framework . . . 150
4.2.1 Traditional status attainment models . . . 150
4.2.2 Extending traditional models: the educational attainment polygenic score and the job market entry . . . 151
4.2.3 Occupational attainment over the life course . . . 155
4.3 Data . . . 158
4.4 Measurement . . . 159
4.4.1 Occupational attainment . . . 159
4.4.2 Educational attainment . . . 160
4.4.3 Educational attainment polygenic score . . . 160
4.4.4 Cognitive skills . . . 161
4.4.5 Parental SES . . . 161
4.5 Analytical strategy . . . 162
4.6 Results . . . 164
4.6.1 Descriptive results: the polygenic score, occupational attainment and gender over the life course . . . 164
4.6.2 The role of the polygenic score at the job market entry . . . 168
4.6.3 Predictors of occupational success over the life course . . . 171

4.7 Conclusion/Discussion . . . 177
4.A Appendix . . . 182
4.A.1 Operationalisation of educational attainment . . . 184
4.A.2 Additional analysis: Women and labour force participation . . . 189
4.A.3 Additional analysis: residuals . . . 190

5 Chapter 5. Conclusion **191**
5.1 Main contributions . . . 192
5.2 Summary of findings . . . 193
5.3 Limitations . . . 195
5.4 Areas of future research . . . 197
Bibliography . . . 199

List of Figures

2.1 Hypothesised mechanisms. . . . 31
2.2 Dimensions of parental involvement. . . . 39
2.3 Comparing estimated coefficients of the associations between parental involvement measures with individual- and family-level variables using the full available sample (OLS) and the sibling sample (fixed effects). . . . 54
2.4 Comparing estimated coefficients of the associations between years of education and parental involvement measures using the full available sample (OLS) and the sibling sample (fixed effects). . . . 59
2.A.1 Reduction of the full sample in the Add Health. . . . 67
2.A.2 Reduction of the sibling sample in the Add Health. . . . 68
2.A.3 Pattern of missing data in the Add Health. . . . 69
2.A.4 Distributions of items of school-based activities in the Add Health. . . . 76
2.A.5 Distributions of items of academic socialisation with mother in the Add Health. . . 77
2.A.6 Distributions of items of academic socialisation with father in the Add Health. . . 78
2.A.7 Distributions of items of parental control in the Add Health. . . . 79
2.A.8 Distribution of years of education (for respondents, for mothers, and for fathers) in the Add Health. . . . 80
2.A.9 Diagnostic plots for the model associating years of education with school-based activities using the Add Health. . . . 81
2.A.10 Diagnostic plots for the model associating years of education with academic socialisation with mother using the Add Health. . . . 81
2.A.11 Diagnostic plots for the model associating years of education with academic socialisation with father using the Add Health. . . . 82
2.A.12 Diagnostic plots for the model associating years of education with parental control using the Add Health. . . . 82
2.A.13 Modeling distribution of the association between each dimension of parental involvement and years of education estimated using the Add Health. . . . 83

3.1 Theoretical causal graph of mediating mechanisms. . . . 89
3.2 Theoretical causal graph of mediating mechanisms with the moderation by gender. 95
3.3 Theoretical causal graph of mediating mechanisms with the confounding by parental characteristics. . . . 96
3.1 Estimated coefficients of the associations between the educational attainment polygenic score and cognitive/non-cognitive skills in the WLS. . . . 108
3.2 Estimated coefficients of the associations between years of education and the educational attainment polygenic score/cognitive skills/non-cognitive skills in the WLS. . . . 109

3.3 Proportions of the total association between the educational attainment polygenic score and years of education mediated by cognitive and non-cognitive skills in the WLS. . . . 110
3.4 Direct and indirect effects of the educational attainment polygenic score on years of education, mediated by cognitive and non-cognitive skills in the WLS. . . . 111
3.5 Estimated coefficients of the associations in the mediation model moderated by gender. . . . 114
3.6 Estimated coefficients of the associations in the mediation model moderated by SES. . . . 116
3.A.1 Sample reduction before and after imputation. . . . 121
3.A.2 Pattern of missing data in the WLS. . . . 122
3.A.3 Boxplots of the polygenic score distributions. . . . 125
3.A.4 Distributions: respondents education, father's education, IQ, EA3 PGS. . . . 129
3.A.5 Histograms of unstandardised measures of non-cognitive skills in the WLS. . . . 130
3.A.6 Associations between a categorical measure of educational attainment and the main explanatory variables in the WLS. . . . 132
3.A.7 Associations between EA3 PGS quintiles and other variables in the WLS. . . . 133
3.A.8 Pearson's correlations between cognitive/non-cognitive skills and the polygenic score/years of education in the WLS. . . . 134
3.A.9 Association between educational attainment and EA3 PGS. . . . 135
3.A.10 Distributions of years of education, father's years of education, and IQ by gender in the WLS. . . . 137
3.A.11 Distributions of non-cognitive skill measures by gender in the WLS. . . . 138
3.A.12 Distributions of years of education and IQ by SES. . . . 139
3.A.13 Distributions of non-cognitive skills by SES. . . . 140
3.A.14 Coefficient sizes of the association between cognitive/non-cognitive skills and years of education by EA3 PGS. . . . 145

4.1 Theoretical causal graph. . . . 155
4.2 The evolution of proportions of respondents in labour force, not in labour force or with missing data by gender in the WLS. . . . 157
4.1 The evolution of average occupational education, occupational income and Siegel prestige scores by gender over time in the WLS. . . . 165
4.2 Boxplots of occupational attainment scores by gender and the educational attainment polygenic score tercile over the life course in the WLS. . . . 167
4.3 Summary of the estimated coefficients for the association between the polygenic score (EA3 PGS), educational attainment (EA) and family socio-economic status (SES) using OLS regressions in the WLS. . . . 171
4.4 Associations between 1 standard deviation increase in the educational attainment polygenic score (EA3 PGS) and a corresponding change in occupational attainment in the WLS. . . . 176
4.5 Adjusted R-squared of different OLS regressions explaining occupational attainment in the WLS. . . . 176
4.A.1 Sample reduction by gender and respondent type (graduate or sibling) in the WLS. 182
4.A.2 Distributions of years of education attained by three time points — 1975, 1992, and 2003 — in the WLS. . . . 183

4.A.3 Boxplots of residuals from the OLS regressions explaining differences in the occupational education scores with educational attainment (purple) and explaining differences in the occupational education score with educational attainment and the educational attainment polygenic score (green). 190

List of Tables

1.1 Thesis outline. . . . 16

2.1 Summary of hypotheses. . . . 32
2.2 Descriptive statistics of the Add Health dataset. . . . 46
2.3 Estimated coefficients from the OLS regressions associating parental involvement with individual- and family-level characteristics. . . . 51
2.4 Estimated coefficients from the OLS regressions associating parental involvement with individual-level characteristics using sibling fixed effects. . . . 53
2.5 Summary of how the results correspond to the hypotheses. . . . 55
2.6 Estimated coefficients from the OLS regressions associating years of education with parental involvement and control variables. . . . 57
2.7 Estimated coefficients from the OLS regressions associating years of education with parental involvement and control variables using sibling fixed effects. . . . 58
2.A.1 Measures of parental involvement. . . . 70
2.A.2 Correlations between individual-level explanatory factors. . . . 71
2.A.3 Descriptive statistics of the items of parental involvement in the Add Health. . . . 72
2.A.4 Correlations between school-based activities in the Add Health. . . . 73
2.A.5 Correlations between items of academic socialisation with mother in the Add Health. . . . 73
2.A.6 Correlations between items of academic socialisation with father in the Add Health. 73
2.A.7 Correlations between items of parental control in the Add Health. . . . 74
2.A.8 Correlations between composite measures of parental involvement constructing using the Add Health. . . . 74
2.A.9 Sibling correlations in the Add Health sibling sample. . . . 74
2.A.10 Interclass correlations (ICC) of parental involvement measures between siblings by socio-economic status in the Add Health. . . . 75

3.1 Summary of the definitions of non-cognitive skills. . . . 92
3.1 Descriptive statistics of the Wisconsin Longitudinal Study. . . . 105
3.1 Direct and indirect effects associated with the EA3 PGS, cognitive, non-cognitive skills, and educational attainment (years of education). . . . 112
3.A.1 Comparing respondents with missing and non-missing data. . . . 126
3.A.2 WLS items for non-cognitive measures. . . . 127
3.A.3 Comparing descriptive statistics of standardised variables between unimputed and imputed samples using the WLS. . . . 128
3.A.4 Correlations between individual-level explanatory factors. . . . 131
3.A.5 Estimated coefficients of direct and indirect effects: sensitivity analysis. . . . 136
3.A.6 Estimated coefficients of direct and indirect effects of openness measures. . . . 141

3.A.7 Estimated coefficients of direct and indirect effects: moderation by gender. . . . 142
3.A.8 Standardised path coefficients, direct and indirect effects: moderation by SES. . 143
3.A.9 Moderation of the association between cognitive/non-cognitive skills and years of education by educational attainment in the WLS. 144

4.1 Descriptive statistics for graduates. 162
4.1 Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the first job held by the graduates with explanatory variables. 169
4.2 Estimated coefficients from the fixed effects regressions associating occupational education, occupational income and the Siegel prestige scores of the first job with explanatory variables between siblings. 170
4.3 Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1970 job held by the graduates with explanatory variables. 173
4.4 Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1975 job held by the graduates with explanatory variables. 174
4.5 Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1992 job held by the graduates with explanatory variables. 175
4.A.1 Descriptive statistics by respondent type (graduate or sibling) and gender (male or female) of the WLS. 187
4.A.3 Counts of sibling pairs by gender composition in the WLS. 188
4.A.2 Traditional measures of the Wisconsin status attainment model in the WLS. . . 188
4.A.4 Estimated coefficients (odds ratios) from the logistics regressions associating women's propensity of not being in the labour force with explanatory variables. . 190

Chapter 1. Introduction

1.1 Education in the age of rising inequality

In many industrialized countries inequality has been increasing since the 1970s (DiPrete, 2020). The earnings of college graduates in the US are now on average 90% higher compared to the individuals with only a high school diploma (Autor et al., 2008), contributing to the income inequality levels not seen since the 1940s (Chetty et al., 2014; Reardon, 2011). Similar inequality patterns are also found in most of the EU countries (Blanchet et al., 2019). Education is at the center of this social and economic problem. Growing earnings returns to higher education can be largely explained by the skill-based technological change (Goldin and Katz, 2007; Blank, 2011). However, the access to education and, subsequently, economic fortunes remain highly unequal (Ellwood and Kane, 2000; Haveman and Smeeding, 2006; Belley and Lochner, 2007). If we want to successfully tackle social inequalities, we need to start by understanding how educational inequalities are produced. The seminal report by Coleman (1966) established that family is the most important factor in explaining educational disparities between individuals. Instead of family influences being neutralised by the democratisation of education, parents in the 21st century are as important for children's academic success as ever. McLanahan (2004) uses the term "diverging destinies" to describe the phenomenon of children from privileged backgrounds obtaining increasing levels of monetary and parenting resources while those from disadvantaged backgrounds are losing them. Grusky and Hill (2018) also talk about the rigidification of social inequalities and its potential to impede social mobility. Assortative mating and the growing differentiation of family resources by parental education levels is widening the educational gap between children. At the same time, those who can take advantage of more resources and attain more and higher-quality education reap disproportionate benefits because of increasing returns to education, leading to "an even more unequal and economically polarised society" (Reardon, 2011, p. 116). Moreover, even though the educational gap between men and women has been closing (DiPrete and Buchmann, 2013), women still experience social inequalities in unique ways in both family and institutional settings (Dumais, 2002). Bringing a gender perspective into the study of educational stratification is, thus, essential. The role of sociologists in this context is to identify and study the mechanisms that perpetuate inequalities. One important mechanism of family influences in education has long been overlooked

— that of genetic transmission. According to Heckman (2013), life success and failure should be understood as a product of the interaction between family investments, social environment, and genetic predispositions. My dissertation investigates the influence of all these three components in a specific context — that of the US, which places a great emphasis on the equality of opportunity. In particular, I use two American datasets — the Add Health and the Wisconsin Longitudinal Study — to explain disparities in educational attainment by taking into account various social factors as well as genetic propensity to education, measured as the educational attainment polygenic score. I study parental involvement and how it can be shaped within families in response to children's genetic educational potential. I also look at the development of skills and how the genetic influences on human capital production are contingent on social environments. I am also interested in the realisation and returns to genetic potential over one's life course and their differential patterns by gender. My goal is to incorporate new knowledge and tools produced in the area of social science genetics in order to bring more nuance into our understanding of social stratification.

1.2 Research on social stratification in education: towards the understanding of mechanisms

The study of inequalities has always been at the heart of the sociological discipline. The starting point of the systematic research on inequality and mobility dates back to the classic work by Perrin (1904) and Sorokin (1927). Since then, the role of education has been extensively studied within the social stratification literature (Blau and Duncan, 1967; Sewell et al., 1969; Bourdieu and Passerson, 1977; Featherman and Hauser, 1978; Teachman, 1987; Coleman, 1988; Hout, 1988; Mare, 1993). To use the terminology of Ganzeboom et al. (1991), the first "generation" of social stratification research mainly used mobility tables to identify descriptive patterns across different time points and countries. The big question of interest to sociologists at the time was the impact of political structure on intergenerational occupational correlations (Glass, 1954; Svalastoga, 1959; Lipset and Bendix, 1959). The second generation social stratification researchers moved beyond description. They developed more systematic ways to code occupations, socio-economic indices, prestige scales;

employed path models and multivariate linear regressions to understand the factors mediating the association between father's and son's occupational attainment. Educational attainment became a key variable of interest as a driving force behind intergenerational correlations (Blau and Duncan, 1967; Haller and Portes, 1973; Sewell and Hauser, 1972). In other words, the goal of such research became to provide explanations to statistical regularities by identifying more fine-grained intervening explanatory factors (Ganzeboom et al., 1991). This line of thinking served as a basis for later developments in sociology known as methodological individualism (Elster, 2012; Goldthorpe, 2015) and analytical sociology (Hedstrom and Bearman, 2009). What characterises these sociological traditions is viewing social processes as made of constituent components. In other words, the purpose of sociological research is seen as not simply documenting statistical patterns but rather as identifying and then empirically testing the mechanisms which bring social phenomena into existence (Hedstrom, 2005). Many ongoing debates over the relative importance of different explanatory mechanisms can be found within sociology: economic versus cultural explanations of inequality (Jonsson, 1987), psychosocial versus structural constraints on mobility (Morgan and McKerrow, 2007), primary versus secondary effects in educational choices (Domina et al., 2011). Instead of perceiving these mechanisms as competing and mutually exclusive, sociologists should use a "toolbox of mechanisms" which spans various substantive debates (Hedstrom and Bearman, 2009). An essential element of this sociological toolbox should be genetics or, as Heckman (2013) calls it, "the accident of birth". There is growing evidence that genetics is one of the mechanisms responsible for intergenerational transmission of social positions (Belsky et al., 2016). New knowledge in the area of social science genetics helps us appreciate the interplay of genetic and environmental causes, which complicate or even defy our causal understanding of the social world (Freese and Shostak, 2009) and informs debates about socioeconomic inequality (Conley, 2016). Historically, the notions of heredity and genetic transmission of social advantages were mentioned from time to time in the social science literature (Jencks, 1972a, 1979, 1980; Scarr and McCartney, 1983; Becker and Tomes, 1979; Bowles and Gintis, 1976; Haveman and Wolfe, 1995; Teachman, 1996). However, the suggestions that social scientists should integrate these concepts in their research were often met with skepticism or even outright refusal: the field of genetics was portrayed as deterministic and, thus, directly opposed to what sociology stands for and, in general, outside the purview of the sociological discipline (Anderson, 1967). The lack of tools to measure heredity directly rather

than as a latent construct also precluded social scientists from venturing into the area of genetics. These days, fortunately, newly developed molecular genetic tools can be used to revisit and elaborate existing sociological theories of social attainment (Belsky et al., 2018). My thesis aims at enriching the traditional toolbox of mechanisms found in sociology by bringing in perspectives and methods from social science genetics. My goal is not to demonstrate the omnipotence of genetic influences but rather to investigate their causal interdependence with the social factors traditionally of interest to sociologists: social background, parental involvement, gender, educational and occupational attainment. I intend to help expand the formulation of mechanisms in the field of social stratification, which, as some claim, has become "excessively narrow" (Ganzeboom et al., 1991).

1.3 Social science genetics: what we know so far

Our current ever-developing understanding of the role that genetics play for the outcomes which social scientists are interested in — such as education, assortative mating, fertility and reproduction, smoking, substance abuse — comes from two main scientific sources: behaviour genetics and molecular genetics.

1.3.1 Behaviour genetics

Behaviour genetics has provided social scientists with the first tools to help quantify their hypotheses about the heritability of different phenotypes. Heritability is defined as the proportion of variation of a trait in a population that can be attributed to genetic differences (Mills and Tropf, 2020). Heritability is not an individual-level estimate but rather a population-level statistic. Low levels of heritability do not imply the absence of genetic effects but rather point at a low genetic variance of a given trait within a population of interest (Visscher et al., 2008). Twin studies serve as a natural experiment which allows us to measure heritability. Such studies compare a correlation of a trait between identical, or monozygotic (MZ), twins, who share almost 100% of their genotypes, with a

correlation between fraternal, or dizygotic (DZ), twins, who on average share 50% of their genetic makeup. After making strong assumptions about equal relevant environments shared by MZ and DZ twins and no assortative mating, higher correlations among MZ twins compared to those among DZ twins are attributed to the former's higher levels of genetic similarity, allowing us to disentangle the phenotypic variance by additive genetic effects (heritability), common (shared) and unique (non-shared) environments (Mills et al., 2020a). Behaviour genetics has so far been the biggest contributor to our understanding on how genetic and environmental influences shape outcomes (Mills and Tropf, 2020). Heritability estimates exist for social outcomes as diverse as wealth (Black et al., 2019), savings (Cronqvist and Siegel, 2015), voting behaviour (Cesarini et al., 2014), crime (Hjalmarsson and Lindqvist, 2013), entrepreneurship (Lindquist et al., 2015), or risk-taking (Black et al., 2017; Beauchamp et al., 2017). Of particular interest to social scientists are numerous educational attainment twin studies which estimate its heritability at around 60% (Branigan et al., 2013). In general, behaviour genetics has provided strong evidence for two claims: first, many traits have heritability of around 50%; second, this conclusion holds not only for biological attributes but also for social outcomes such as ability, personality, and preferences (Benjamin et al., 2012; Polderman et al., 2015). In other words, most human traits and behaviours seem to be highly heritable (Turkheimer, 2000). Unfortunately, heritability estimates produced in twin studies do not isolate specific genetic variants. These estimates, as a consequence, are inherently unsatisfying to sociologists whose main interests lie in understanding mechanisms (Boardman et al., 2010; Visscher et al., 2008). Fortunately, recent developments in molecular genetics now allow us to measure genetic influences directly instead of modelling them as latent constructs (Conley, 2016).

1.3.2 Molecular genetics

The development of genotyping and DNA sequencing tools allowed researchers to study genetic variants explicitly. Theory-driven candidate-gene studies were very promising but did not succeed in producing many substantively significant and reproducible results. The sequencing of the whole human genome in the early 2000s produced a breakthrough: now many genetic variants could be tested against different outcomes without *a priori* knowledge of their potential functions.

Candidate-gene studies

Candidate-gene studies use one or a handful of genetic variants to study their effects on traits of interest (Conley, 2016). The focus is on studying genetic polymorphisms (loci where the DNA sequences differ between individuals). The possible variants of such polymorphisms are known as alleles, or alternative forms of a gene. Alleles, in turn, can have different forms but often they are measured as single nucleotide polymorphisms (SNPs). SNPs represent variations between individuals at specific positions along the genome which are present in a sufficiently large portion of the population, 1% being the most common threshold (Mills et al., 2020a). Candidate SNPs are chosen based on their suspected biological function. This method can be useful for traits and diseases such as Alzheimer's (APOE) or obesity (FTO), where a single gene can indeed explain a sizeable part of the trait's heritability (Mills and Tropf, 2020). However, other complex social and behavioural outcomes are most often polygenic. As a consequence, candidate-gene studies produced false positive results, which could not be replicated and which were later debunked (Rietveld et al., 2014). The numerous reasons for such a failure to replicate include an incorrect specification of hypotheses, a publication bias towards positive findings, statistical significance thresholds not being stringent enough, studies being underpowered. The journal *Behaviour Genetics* invites researchers to critically reevaluate the evidence which has been previously produced by such studies (Hewitt, 2012).

Genome-wide association studies

The lack of established large effects for common alleles suggested that most of the complex traits are probably polygenic — many genetic variants have an effect on a particular phenotype and each of such effects is small. Consequently, a more a-theoretical, hypothesis-free line of research became perceived as potentially more productive. The realisation of such research agenda is the genome-wide association study (GWAS), first developed in 2005 and made possible by the development of genotyping arrays and the rapidly decreasing costs of DNA sequencing as well as the growing availability of large-scale genetic data (McCarthy et al., 2008). A GWAS tests whether there

are any associations between genotyped SNPs and a phenotype of interest in order to identify significant hits. Since running a GWAS essentially entails testing numerous hypotheses (for each SNP), the resulting estimates are Bonferroni-corrected for multiple testing (Mills et al., 2020a). A given SNP is considered a hit if it meets a stringent significance threshold of $p < 5 \times 10^{-8}$. GWASs are often comprised of meta-analyses of results estimated in many different datasets. The most recent GWASs can be as large as 1 million individuals (Lee et al., 2018). In addition to numerous biological functions, many GWASs already exist for various social traits as well. Some of the most famous examples include GWASs for educational attainment (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018), fertility (Barban et al., 2016; Mills et al., 2021; Mathieson et al., 2020), well-being (Baselmans et al., 2019), risk preferences (Karlsson Linner et al., 2019), tobacco and alcohol use (Liu et al., 2019).

Polygenic scores

One of the most important innovations in the research area, which built upon the GWAS, was the development of polygenic scores (PGSs) — composite indicators representing linear combinations of SNPs, whose weights are derived from GWASs (Mills and Tropf, 2020). The premise of such scores is the previously discussed fact that most of the complex human traits are inherently polygenic (Conley, 2016). A PGS is intended as a quantitive measure of an individual's genetic propensity to a trait (Mills et al., 2020a). Upon the release of the first educational attainment GWAS and the corresponding polygenic score (Rietveld et al., 2013), the focus was on testing its predictive power on new samples of data (Krapohl and Plomin, 2016; Domingue et al., 2017). Nevertheless, the sheer predictive power, increasing with every subsequent GWAS, slowly became of secondary importance. The research focus shifted to trying to understand the black-box nature of the polygenic score and interpreting its strengths as well as weaknesses. One of the central questions surrounding the educational attainment polygenic score is what exactly it measures — why that particular combination of genetic variants is predictive of educational outcomes. The first possible explanation for such an association is actual biological causal effects that, for example, can influence brain function, which, in turn, can shape learning ability, cognitive function, memory, or personality

(Belsky et al., 2018). Some of the intervening mechanisms found in the existing research include dyslexia, attention problems (Cesarini and Visscher, 2017; Domingue et al., 2015; de Zeeuw et al., 2014), reading performance at different ages (Selzam et al., 2017), school performance at different ages (Selzam et al., 2017; de Zeeuw et al., 2014), behavioural problems, self-efficacy, and well-being (Krapohl and Plomin, 2016). One of the goals of this dissertation is to study another potential mediating mechanism — non-cogntive skills. However, it is important to note that such detected associations can capture not just biological pathways but also be driven by environmental confounding and assortative mating (Mills and Tropf, 2020). Children of highly educationally-predisposed individuals might have relatively higher educational attainment polygenic scores as well as higher than average educational outcomes, which could, to a large extent, be influenced by one's social environment. Consequently, the resulting relationship between the polygenic score and the educational outcomes might not be entirely causal. Similarly, population stratification, or differences in allele frequency across populations, might create spurious correlations between the polygenic scores and phenotypes of interest (Belsky et al., 2016; Robison et al., 2017). A careful controlling for population stratification and socio-economic background as potential confounders, is a crucial task in this thesis. Indirect genetic effects, also known as genetic nurture, are also possible. Kong et al. (2018a) demonstrate that genetic variants, not transmitted from parents and children, can still shape the environments in which those children grow up and influence their academic and, more generally, life trajectories. This line of research is very promising but, unfortunately, beyond the scope of this thesis as I do not have the data which would contain multigenerational genetic measures. Another important concept to take into consideration when utilising polygenic scores is pleiotropy — when one gene is associated with two or more traits (Mills and Tropf, 2020). Research shows that this phenomenon is wide-spread and that genetic architecture of many complex traits is interconnected (Watanabe et al., 2019). As a consequence, the educational attainment polygenic score can potentially be employed in trying to understand the outcomes that are not strictly educational. In other words, this composite indicator can capture genetic variants which are associated with socioeconomic success more broadly such as one's occupation, assets, credit problems, social welfare (Belsky et al., 2018). In this dissertation I explore this possibility by extending the models of attainment from purely educational ones to those concerning occupational status over one's life course.

1.4 Genetic mechanisms in educational stratification: overview of the thesis

Educational attainment is shaped by interactive social, structural, historical and genetic forces (Duncan et al., 2004). The potential of the educational attainment polygenic score lies not in its predictive power but rather in its promise to help us start asking and answering complex social questions (Domingue et al., 2015). The introduction of the notion of genetic predisposition into the area of sociology of education should not be seen as an attempt to upend the sociological status quo but rather as a gentle push to make our theories more nuanced, even if less tractable, empirical research richer, and biases (due to genetic confounding) lower. In this dissertation my hypotheses and theories are inspired by the rich literature of sociology and the tools of genetics are employed to find a new angle to hard-pressing questions.

More specifically, my thesis aims at connecting social origins, family processes and education-linked genetics to skill formation, educational attainment and labour market outcomes in the US. In the process, I aim to both open up the black box of the educational attainment polygenic score and to better understand the aetiology of education. The role of genetic propensity in the educational context is important both from a substantive point of view but also as a methodological improvement that potentially allows us to make our environmental estimates less biased. As succinctly put by Freese and Shostak (2009), genetic and social causes are so interdependent that the study of one is necessarily confounded or incomplete without the other. This thesis also attempts to span the whole life course of an individual: origins (family background), skill formation (education), destinations (occupation).

My first broad goal is to better understand the relationship between parental socio-economic background as well as parental involvement in childhood and education-linked genetics in producing educational disparities. In Chapter 2 I explore the potential evocative gene-environment correlation: does child's educational predisposition evoke parenting of certain intensity? At the same time, I seek to reduce possible genetic confounding in the association between parenting and years

of education. In Chapter 4 I study passive gene-environment correlation, or the association between the genotype a child inherits from parents and the environment where she is raised in. My second goal is to move beyond correlation to mediation and examine the genetic aspect of cognitive and non-cognitive skill formation as well as its role in the process of social stratification (Chapter 3). My third goal is to cover not only origins and early-life human capital development but also destinations to better understand the genetic architecture of status attainment. I focus on the reinforcing and ever-changing roles of social origins, genetics and educational attainment in the creation of occupational inequalities over one's career (Chapter 4). My fourth goal is to incorporate contextual factors into our understanding of how education-linked genetics operate. The Wisconsin Longitudinal Study follows the respondents through the era of liberalising gender attitudes and, in a way, provides an opportunity to study the double gene-environment interaction: how the genetic influences differ by gender and how those gender differences are moderated by both life-course developments and structural shifts over the years (Chapter 3, Chapter 4). The fifth goal is to bring in the sociological tradition of rich, nuanced and multi-faceted measurements to social science genetics. I explore multiple angles of social phenomena: parental involvement with its different attributes (Chapter 2), a rich set of non-cognitive skills (Chapter 3), and competing ways to measure occupational attainment (Chapter 4). Below I discuss each empirical chapter in more detail (also summarised in Table 1.1).

1.4.1 Chapter 2: Parental involvement

Parenting and, as a consequence, the scientific understanding of it has never been so important. More than 60% of today's income comes from wages, which are to a large extent determined by one's human capital. Those skills are acquired with the help of parental inputs (Doepke et al., 2019). Researchers document a growing "parenting divide" and "developmental gradient": better-educated, better-off parents are investing their time into emotional and scholastic support at an increasingly higher rate than those from a lower socio-economic background (Reeves and Howard, 2013; Kalil et al., 2012). Already in the 1970s sociologists noticed that parents with different educational levels use different strategies to raise their kids (Hill and Stafford, 1974; Lareau, 1989): whereas more

educated parents often endorse a very intensive, time-consuming approach of "concerted cultivation", less privileged families are not as involved and are largely characterised by an attitude of "accomplishment of natural growth" (Lareau, 2002). It is very important to understand family-level dynamics of parenting because parenting preferences based on the socio-economic background affect children's welfare and later attainment (Doepke and Zilibotti, 2017) as well as social mobility and long-run inequality more generally (Doepke et al., 2019). Within-family patterns of how parenting decisions are made should be thoroughly studied, as differences in parenting are observed not only between families but also between children growing in the same family. In particular, the role of genetics in this context has been essentially neglected: while we know that individuals with higher genetic propensities to education tend to achieve better educational outcomes, we don't know whether parents react to child's genetic propensities, adjust their parenting accordingly, and, thus, help actualise their children's potential. I use the Add Health dataset to try to bridge these gaps in the literature.

In Chapter 2 I conceptualise parental involvement as having three dimensions: school-based activities, academic socialisation, and parental control. In the first part of the chapter, I investigate individual- and family-level predictors for each parental involvement dimension. In addition to the child's genetic propensity to education, measured as the educational attainment polygenic score, I also explore the role of age, gender, achievement, cognitive skills, self-efficacy, and behaviour problems (individual factors) as well as parental education, income, and father absence (family-level factors). In the second part of the chapter, I ask whether parental involvement, in turn, is associated with the ultimate educational outcome — completed years of education. I run both OLS regressions and sibling fixed effects, which allows to control for unmeasured family-level confounding, and compare the results produced using the two approaches. I find that the child's genetic propensity to education is not a significant factor driving parental involvement. Parental education seems to be associated with a certain cultural logic which encourages an involved type of parenting. Overall, parental involvement, once important observable individual and family-level confounders are taken into account, does not exert a strong influence over one's educational attainment.

1.4.2 Chapter 3: Cognitive and non-cognitive skills

Parenting and other forms of cultural transmission influence the formation of cognitive and non-cognitive skills (Bjorklund et al., 2005). However, we know much less about their genetic channels of formation. Non-cognitive skills can be defined as personality characteristics and social behaviours which help to maximise life chances (Pearce et al., 2016; Borghans et al., 2008). Although already briefly mentioned in the 1970s by Jencks (1979) and Bowles and Gintis (1976), non-cognitive skills gained traction in the social science literature only in the early 2000s. The concept was popularised by Heckman and his colleagues (Heckman and Rubinstein, 2001; Heckman, 2006; Heckman and Mosso, 2014). Whereas traditionally individual's economic returns were understood as a function of cognitive abilities, the pioneering research on non-cognitive skills showed that human capital in the educational and economic realm is much more multi-faceted (Murnane and Levy, 2006; Feinstein, 2000). Moreover, researchers came to suspect that non-cognitive skill formation must have both environmental and genetic antecedents (Heckman et al., 2006; Cesarini et al., 2010; Knudsen et al., 2006). Father-son correlations of non-cognitive skills are around 0.43 — almost as high as those of cognitive skills, which hover around 0.48 (Grönqvist et al., 2010). Twin studies have demonstrated that cognitive and non-cognitive skills are highly heritable (Tucker-Drob et al., 2016).

Chapter 3 aims at extending the existing knowledge by integrating novel molecular genetic tools into the investigation of how cognitive and non-cognitive skills are formed and how this type of formation affects educational attainment. To this aim I use the Wisconsin Longitudinal Study and employ mediation analysis to answer the question of whether non-cognitive skills mediate the association between the educational attainment polygenic score and years of education. I am also interested in comparing the mediation process by gender and socio-economic status since the previous literature has suggested significant heterogeneity (Herd et al., 2019). My results show that cognitive and non-cognitive skills together account for approximately 60% of the association between one's genetic predisposition to education, measured by the educational attainment polygenic score, and years of education. Cognitive skills, however, seem to play a much more pronounced role. Among non-cognitive skills, only openness displays a consistent and substantial mediating role. I

also uncover interesting gendered patterns: non-cognitive skills have higher explanatory power for women's education than for men's but the opposite is true for cognitive skills.

1.4.3 Chapter 4: Occupational attainment over the life course

Cognitive and non-cognitive skills, developed via the interplay of environment and genetics, have an impact on individuals' educational achievements (Heckman, 2013). Moreover, as already discussed, genetic variants often display a high degree of pleiotropy — ability to influence multiple traits (Mills and Tropf, 2020). Can the educational attainment polygenic score explain life outcomes beyond schooling? Is it associated with occupational attainment through pathways independent of educational attainment? The existing literature suggests that this is indeed the case (Belsky et al., 2018). However, some knowledge gaps are still present and I aim at filling some of them in the Chapter 4 of this thesis. I connect one's origins to destinations by combining intergenerational and intragenerational perspectives (Bukodi and Goldthorpe, 2011). In other words, I investigate whether the role of genetic propensity to education and social background wanes or grows over one's career.

I use the Wisconsin Longitudinal Study (WLS) to study these patterns because it has detailed longitudinal data on occupational attainment. I am also interested in whether males and females experience these changes differently. I operationalise occupational attainment with three different measures: occupational education, occupational income, and Siegel prestige score. The results produced using OLS regressions and sibling fixed effects analyses show that education-linked genetics are associated with occupational outcomes of the first job but only through their effect on educational outcomes. When the whole career is considered, the role of genetic propensity decreases for men but slightly increases for women. This pattern can result from a mixture of factors: sample selection due to life-course dynamics (some women are not in the labour force during child-rearing years) or structural shifts in the labour market and how it rewards female workers. I also document the differences in the results produced using the three different measurements for occupational attainment. Such heterogeneity by operationalisation reveals the relevance of rich sociological understanding of social phenomena when investigating not just their environmental but also genetic

foundations.

Chapter	Research Questions	Main variables	Data & Methods	Contributions
Chapter 2. Parental Involvement: Correlates and Consequences	1. What are individual- and family-level predictors of parental involvement in adolescence? 2. Does parental involvement predict educational attainment?	Dependent variables: parental involvement, years of education. Independent variables: parents' education, parental income, father's absence, age, gender, GPA, educational attainment polygenic score (EA3 PGS), cognitive ability, self-efficacy, behaviour problems.	Data: Add Health. Methods: OLS regression, sibling fixed effects.	1. Distentangling of family and individual-level influences on parental involvement. 2. Understanding the relationship between genetic propensity to education and parental involvement/years of education within families.
Chapter 3. Genetic Predisposition to Education: The Role of Cognitive and Non-Cognitive Skills	1. Do non-cognitive skills mediate the association between the educational attainment polygenic score and years of education? 2. Does the mediation process differ by gender and socio-economic background?	Dependent variables: years of education. Independent variables: educational attainment polygenic score (EA3 PGS), cognitive skills (IQ), non-cognitive skills (openness, conscientiousness, perseverance, self-efficacy, self-esteem.	Data: Wisconsin Longitudinal Study (WLS). Methods: mediation analysis.	1. Exploration of the role of non-cognitive skills for educational attainment and its genetic influences. 2. Uncovering differences by gender and SES in the mediation of genetic propensity to education.
Chapter 4. Occupational Attainment Over the Life Course: The Role of Genetic Propensity to Education	1. Can the educational attainment polygenic score predict occupational outcomes over the life course over and above years of education? 2. Is this association moderated by gender? 3. How do these patterns change over one's life course?	Dependent variables: occupational education, occupational income, Siegel prestige scores. Independent variables: educational attainment polygenic score (EA3 PGS), cognitive skills (IQ), years of education, parental SES, previous occupational attainment.	Data: Wisconsin Longitudinal Study (WLS). Methods: OLS regression, sibling fixed effects.	1. Documenting the role of genetic propensity to education in the context of occupational attainment. 2. Exploration of how gender produces unique occupational experiences.

Table 1.1: Thesis outline.

1.5 Data & methods

In this dissertation I use two observational data sources: the National Longitudinal Study of Adolescent to Adult Health (Add Health) for Chapter 2 (Harris et al., 2019) and the Wisconsin Longitudinal Study (WLS) for Chapter 3 and Chapter 4 (Herd et al., 2014). Overall, these data sources were chosen because of the availability of genetic information, the rich sets of information on one's family origins, the longitudinal designs, and the measure of completed education. Other datasets popular in the area of social science genetics could not be used because either the genotypic information is about parents and not children (Understanding Society), participants have not completed their education (Millennium Cohort Study), the information on socioeconomic background, parental involvement or skill formation is not available (UK Biobank) or is not at the right level of granularity (Health and Retirement Study). The Add Health and the WLS have their limitations but they are the most optimal data choices given the particular substantive questions I raise in this thesis. The strengths and weaknesses of these two datasets as well as the different ways in which the educational attainment polygenic score is constructed in each dataset are discussed below.

1.5.1 Data and samples

Chapter 2 of this thesis utilised the information provided by the Add Health. The Add Health is a nationally representative longitudinal study which started in 1994 and is currently in the process of gathering the sixth wave of data (Harris et al., 2019). It is mandated by the National Institutes of Health and administered by the University of North Carolina at Chapel Hill. The main goal of this survey is to study health and risk behaviours of adolescents in the US. In the first phase which took place from September 1994 to April 1995, around 90,000 school students from 134 schools across the country were interviewed using a short questionnaire. In the spring and summer of 1995 more detailed in-home interviews were conducted with approximately 20,000 children and one of their parents. The subsequent waves follow these individuals over the years. The main strength of the Add Health is the richness of its various data components. It contains ecological information on residence location, school characteristics via the school-administrator

questionnaire, sibling data, biomarker and genetic information, friendship social networks, school transcripts. Most importantly, it has quite detailed data (coming both from children and parents) about child-parent interactions and parenting practices. However, the main goal of the Add Health is understanding how different behaviours influence teenagers' health and not education (which is my primary interest). Educational attainment and academic achievement is measured but not extensively. The dataset does not provide any information about child's capabilities (for example, an official IQ measure) and achievement (grades or test scores) prior to adolescence, making it difficult to model education both as an antecedent to other behaviours and as an outcome affected by adolescent's traits and changing habits. Furthermore, no questions are asked about adolescents' more concrete career aspirations regarding certain jobs/fields or cultural capital available to them through their parents. The impressive sample size of almost 20,000 individuals is significantly reduced in my analyses as I include only the individuals who have genetic information available and who report their educational attainment (the main outcome of interest) in Wave IV.

The WLS is used to test my hypotheses in Chapter 3 and Chapter 4 of this dissertation. It is one of the longest running study in the US concerned with life-course and aging (Herd et al., 2014). In the late 1950s the officials of the state of Wisconsin decided to launch a large-scale survey which would include all individuals graduating from high school in 1957. The goal of this study was to understand who is choosing to go to college and what factors are driving their choices. Seven years after the initial questionnaire, William H. Sewell, Professor of Sociology at the University of Wisconsin-Madison, decided to incorporate it into a larger and more general research project on social stratification. He designed a follow-up survey which included a randomly selected sample of one third of the Wisconsin's high-school graduate population of 1957. The questionnaire focused on family background and aspirations. The biggest contribution of the WLS is the famous Wisconsin status attainment model, which introduced psychosocial mechanisms, namely the influence of significant others on students' aspirations and expectations, to the study of intergenerational transmission of educational inequalities. The strengths of the WLS include having information on siblings and family formation history of graduates themselves, the administrative data on IQ, income and employment, the rich operationalisation of job measures, the battery of questions pertaining to personality, psychology, and well-being, the life-course perspective. Nevertheless, the

WLS also have some inherent limitations. It has a limited number of questions relating to early childhood conditions, posing challenges to my ability to study the malleability of predisopositions from early childhood to adulthood (Heckman, 2013). Furthermore, the WLS is often criticised for its homogeneous sample: almost all participants are white, have at least a high school diploma, have mostly rural origins and were born in the late 1930s. On the one hand, the insights drawn from using the WLS then can be generalised only to older white Americans, dramatically limiting their external validity. On the other hand, the WLS can serve as a case study whose homogeneity minimises potential confounding as fewer unobserved variables are present. Studying this particular cohort over time allows to track how life-course dynamics and rapidly changing structural factors interact with genetic propensities.

1.5.2 Utilisation of polygenic scores

A polygenic score of an individual i is equal to the sum of the allele counts a_{ij}(0,1 or 2) from each SNP j = 1,..M, multiplied by a weight w_j: $PGS_i = \sum_{j=1}^{m} w_j a_{ij}$. A PGS is intended as a quantitive measure of an individual's genetic propensity to a trait (Mills et al., 2020a). Such scores go beyond the candidate-gene model because they capture a genetic continuum (Belsky et al., 2016). They have served as a building block for sociogenomics (or social science genetics) since social scientists can finally incorporate a direct genetic measure into their models.

This dissertation heavily utilises the educational attainment polygenic score. The latest educational attainment GWAS included more than 1.1 million individuals in its discovery sample and produced a polygenic score which could predict up to 9-11% of the variance in years of education - an estimate in its predictive power comparable to many key measures used by social scientists (Lee et al., 2018).

The methodological improvements and conceptual innovations surrounding the utilisation of the polygenic scores is ongoing. The field has matured a lot since the early days when the predictive power of the score was its sole success metric. As a consequence, more complex modeling techniques have become common. Datasets containing information on family members across generations coupled with richer and more sophisticated phenotypic information encourage researchers to explore

complex pathways of the intergenerational genetic transmission of advantages and disadvantages (Liu, 2018). Another big focus is understanding the previously discussed genetic nurture: the fact that parental genotype exerts influence over children not only through transmitted alleles but also through the non-transmitted ones (Kong et al., 2018a). The study of gene-environment interaction (GxE) is also a key area of interest. GxE demonstrates how contingent genetic effects are on the environments and truly challenges the notion of nature and nurture as something that can be disentangled. The GxE tradition started in the era of family studies and helped to demonstrate that in more equalised environments the importance of genetics, instead of decreasing, actually increases (Heath et al., 1985; Turkheimer et al., 2003; Guo and Stearns, 2002; Rowe et al., 1999). This approach has been extended to studying the interaction between the polygenic scores and macro-level (Rimfeld et al., 2018) or more proximate circumstances (Guo et al., 2015). Researchers have also employed more advanced statistical techniques such as polygenic scores as instrumental variarables in the Mendelian Randomisation framework (Davey Smith, 2005) and genomic structural equation modeling (SEM) which allows to closely investigate assumed causal models (Grotzinger et al., 2019).

The educational attainment polygenic scores utilised in this dissertation are not constructed by me but rather are directly borrowed from the datasets that I use. Both the Add Health and the WLS release their own educational attainment polygenic scores constructed using the summary statistics from the third educational attainment GWAS (Lee et al., 2018). The very first versions of the polygenic scores were essentially linear combinations of SNPs. However, this method does not take into account the effect of linkage disequilibrium (LD) — a correlation of alleles at different locations in the genome (Slatkin, 2008). Two different methods of developing polygenic scores aim at addressing these concerns: PRSice (used in the Add Health) and LDPred (used in the WLS). PRSice is known as a pruning and thresholding technique. It removes redundant SNPs which are in very high or complete linkage disequilibrium with other SNPs. It then produces the polygenic score using different p-value thresholds for SNPs and selects the one with the highest predictive capacity — the most precise (Euesden et al., 2015). PRSice is a very flexible method. However, when pruning, potentially informative markers are discarded, thus, limiting its overall predictive accuracy. In contrast, LDPred includes all measured SNPs and weighs them by their

conditional effects given other SNPs (Vilhjamsson et al., 2015). This Bayesian method takes linkage disequilibrium information from an external reference panel. LDPred has been shown to produce results both of higher predictive power and less bias with advantages increasing in higher sample sizes (Mills et al., 2020a). Nevertheless, some limitations of LDPred compared to the PRSice approach are also worth mentioning. It is not as flexible and its advantages highly depend on the quality of the reference panel, which needs to undergo quality controls (Vilhjamsson et al., 2015). Moreover, since the assumptions about the prior distribution of causal SNPs are most often complex, different priors need to be specified. The process of calculating linkage disequilibrium weights can be very computationally intensive (Mills et al., 2020a). All in all, the literature on the best ways to construct polygenic scores is rapidly developing. PRSice and LDPred are both standard approaches widely used in the field despite their limitations.

1.6 Summary

To recapitulate, this thesis aims at understanding the interplay of genetic and social factors in explaining educational attainment in the US. The introductory chapter outlined the key elements of the thesis. I explained how the rising social inequalities serve as a motivation for this dissertation. I also sketched a theoretical sociological framework which informs the direction my thesis takes and identified the gaps in the literature to be filled. I also provided the basics of social science genetics and outlined the integration of the educational attainment polygenic score in my thesis. I then briefly summarised the three empirical chapters, which use either the Add Health or the Wisconsin Longitudinal Study data, and focus on parental involvement, non-cognitive skills, and status attainment over the life course. Finally, I turned to the discussion of the datasets and methods I use. Overall, my thesis will explore the role the genetic propensity plays in educational attainment and how the genetic transmission mechanism is embedded within a larger social structure.

Chapter 2. Parental Involvement: Correlates and Consequences

Parental Involvement: Correlates and Consequences

Abstract

In recent decades the attention paid to parenting has dramatically increased as more and more intensive parental strategies have become widely promoted. However, we know surprisingly little whether middle-class parenting strategies used in adolescence truly translate into better educational outcomes. This paper intends to fill this gap by exploring the correlates and consequences of three parental involvement dimensions — school-based activities, academic socialisation, and parental control. First, I investigate and compare family-level and individual factors (including the educational attainment polygenic score) which are correlated with different dimensions of parental involvement. Second, I ask a question whether, after taking into account these factors, parental involvement still matters for academic success, measured as educational attainment, or whether these factors confound the relationship between parental involvement and years of education. To answer these questions I use the Add Health dataset to compare individuals between families using OLS regressions and within families using sibling fixed effects. I find a high socio-economic gradient for school-based activities and academic socialisation but not for parental control. The strongest individual-level explanatory factors of parental involvement are child's academic performance, cognitive ability, and behaviour problems. Interestingly, a novel indicator of genetic propensity to education, or the polygenic score, is not associated with parental involvement measures. Finally, my results show that these individual and family characteristics indeed confound the relationship between parental involvement and years of education as its association size decreases by two thirds once those confounders are included in my models. Although not precisely causal, my results show that in this context parental involvement has limited explanatory power for education. It calls for a critical reevaluation of the existing parenting literature which champions a single parenting paradigm detached from structural and cultural features of a particular family.

Keywords: parenting, educational attainment, siblings, polygenic score, Add Health

2.1 Introduction

Parental involvement, or investment of parental effort and time into child's development and academic progress, has long been understood by economists to be an important input for child's human capital production (Becker, 1964; Heckman et al., 2013; Doepke and Zilibotti, 2017). Sociologists should be interested in the topic for two additional reasons. First, the attention paid to parental involvement has increased significantly as the separation between school and family roles has been narrowing for the past five decades (Reardon, 2011). More and more parents feel that in order to give their children a good chance at succeeding in an increasingly competitive landscape of college admission and labour market, they need to get heavily involved in their children's extracurricular and academic lives. Time use research confirms a decline over the years in permissive parenting and a growing popularity of very intensive parenting (Doepke et al., 2019), which the media have been quick to label "helicopter parenting", "the rug rat race", or "perfect madness" (Ramey and Ramey, 2009; Warner, 2005). Second, sociologists should also pay attention to parental involvement because it represents one of the stratification processes. Existing literature documents a growing socio-economic gradient in parental investments (Kalil et al., 2016). In other words, the growth in parenting activities has mostly occured among highly-educated mothers, leading to an increasing parenting divide both in terms of quantity of parental time and support (Guryan et al., 2008; Reeves and Howard, 2013) and the effectiveness of how these resources are used (Kalil et al., 2012). While rational from the perspective of advantaged parents who want to benefit from increasing returns to having highly skilled children (Schaub, 2010; Cunha and Heckman, 2008; Ramey and Ramey, 2009), such inequalities in parenting can impede social mobility in the long run (Doepke et al., 2019). In reaction to these trends, there has been a push from policy-makers in the US to make parental involvement federally mandated (Hill and Tyson, 2009). However, before making such prescriptions, the causal link between different parenting practices and children's outcomes should be established. While such evidence exists for early childhood from randomised interventions (Kalil, 2014), less is known about the effectiveness of parental involvement in adolescence (Harris, 1998).

The task of establishing a causal link between parenting behaviours and child outcomes is complicated by three sources of confounding (Brooks-Gunn and Markman, 2005). First, social, economic and educational conditions can influence both the choice of parenting strategies and child's academic success. The resulting correlation between the two might not be causal. Consequently, simply transplanting parental practices of high-SES families might not prove effective in other social contexts. Second, parental involvement, especially in adolescence, is a two-way relationship: parental involvement cannot be driven simply by cultural logic and preferences of parents but has to be negotiated with the child, responding to her personal characteristics and needs. In other words, child's individual characteristics might influence both the types of parenting used and the child's educational success. Finally, genetic similarities between parents and children might also complicate the relationship between parental involvement and various outcomes. What is deemed to be a successful investment into children by their middle-class parents can reflect the advantages they share in their genetic endowments. This paper will aim at at least partially addressing all of these three issues.

A lot of existing research has been interested in how different parental strategies affect child outcomes but does not sufficiently account for all the major confounders. In other words, it uses parental involvement as an explanatory tool. The first goal of this paper, however, is to better understand the determinants of parental involvement using the Add Health data. Such findings could inform us both about the potential confounders in the association between parental involvement and child outcomes and about how parents make parenting choices and how one could modify these choices. Once the determinants of parental involvement are better understood, I will then look at whether parental involvement can explain some of the disparities in educational attainment between individuals. Moreover, this paper will focus on adolescence. With notable exceptions (Plenty and Jonsson, 2021; Benner et al., 2016), much less empirical sociological research has been done on parenting decisions during adolescence compared to early childhood and primary-school years and it is not firmly established whether a socio-economic gradient is as visible in this developmental period.

First of all, I will investigate family-level characteristics which determine different levels of parental

involvement in three separate dimensions during adolescence: school-based activities, academic socialisation, and parental control. Second of all, I will look into individual-level determinants, i.e. personal characteristics which parents react to and adjust their parental involvement accordingly, using both OLS and family fixed effects approaches. Seven individual characteristics will be used for this purpose: age, gender, educational attainment polygenic score, GPA, Peabody Picture Vocabulary Test score, self-efficacy index, and behaviour problem index. Third of all, I will use the newly acquired knowledge about individual-level influences to account for their potential confounding and explain educational attainment.

This paper will make several contributions to the existing literature. First, it will seek to understand determinants and consequences of parental involvement in adolescence rather than in earlier developmental periods. Second, it will operationalise parental involvement in a more fine-grained way than has been done before by distinguishing between three different dimensions of parental involvement — school-based activities, academic socialisation, and parental control. Third, the paper will supplement the between-family analysis with a within-family design which can effectively deal with potential confounding by unmeasured parental social, educational, financial, and genetic endowments. Finally, a rich set of individual characteristics will be used as predictors, including the educational attainment polygenic score — an innovative measure for one's genetic propensity to education.

2.2 Theoretical framework

2.2.1 Definition of parental involvement

An active academic interest in parental involvement originated in the 1960s among child development and socialisation researchers (Baumrind, 1966; Maccoby and Martin, 1983). Parental involvement is generally understood as strategies employed by parents to promote their children's academic success (Hill and Taylor, 2004). In contrast to more monetary parental investments and transfers, parental involvement is a two-way time-intensive interaction between parents and children

and between parents and schools. There are two main ways to conceptualise parental involvement: as parenting styles and as parenting practices. The typology of parenting styles was developed by Baumrind (1966). Three styles — permissive, authoritative and authoritarian — broadly represent the emotional climate (Darling and Steinberg, 1993) and strategies that parents think are best for raising their children (Doepke et al., 2019). Permissive parents, also known as indulgent, are not very demanding and let their children make their own academic decisions (Doepke and Zilibotti, 2017). Authoritarian parents, on the other hand, are very demanding but not at all responsive — they restrict the set of choices their child can make (Bronte-Tinkew et al., 2006). Authoritative parents combine both demandingness and responsiveness (Baumrind, 1966) and instead of making decisions for their children or letting their children go free they try to adjust children's preference to parents' liking (Doepke and Zilibotti, 2017). The adoption of one of these parenting styles is dependent on class- and country-related contextual factors such as cultural logic (Lareau, 2002), beliefs (von Otter, 2014), or social and cultural capital (McNeal Jr., 2001). The second way to conceptualise parental involvement is to investigate concrete parental practices. In contrast to rather abstract overarching parenting styles, parental practices are specific behaviours used by parents to socialise with their children (Darling and Steinberg, 1993). This paper will focus on parenting practices instead of parenting styles because the latter is less culturally deterministic and more flexible and reactive to child's individual characteristics within families (Hill and Taylor, 2004).

2.2.2 Dimensions of parenting practices

The types of parenting practices and the time spent pursuing those practices depends on the child's age. In early childhood, parental involvement amounts to fulfilling child's basic needs, such as feeding, clothing, or stimulating her cognitive development through playing and reading (Kalil et al., 2012). During elementary school years, home-based academic activities become the most important way through which parents become involved. Such home-based activities are reading books, helping with homework, and assisting in school projects (Gonzalez, 2004; Epstein, 2001). In middle school, home-based parental involvement focuses on inculcating good study habits and interest in school (Pomerantz et al., 2007). During this period parents are also involved in planning, organising, and

monitoring the child's life outside the home (Kalil et al., 2012). In high school, however, parental involvement can no longer be solely driven by parents' cultural logic and preferences but rather needs to be negotiated with their child. Adolescents, who will be the focus of this paper, assert their independence, leading to a decline in parental involvement and an increase in the influence of one's peers (Ryan and Stiller, 1991). Home-based activities and other educational support practices become less important. However, parents of adolescents still want to influence their children's behaviour and support their skill formation (Agostinelli et al., 2020). They can achieve this by employing practices covering three broad parental involvement dimensions, which the literature shows are the most important during adolescence: school-based activities, academic socialisation, and parental control.

School-based activities. School-based activities denote parental involvement in a child's life by attending teacher-parent conferences and being involved in school decision-making, volunteering at school, running for leadership positions, organizing fundraisers, being in regular contact with teachers about available educational programs, student progress and expectations (Grolnick and Slowiaczek, 1994; Hill and Tyson, 2009; Benner et al., 2016; Epstein, 2001; Hill and Taylor, 2004; Spera, 2005). School-based activities can be effective in producing positive child outcomes by allowing parents to learn new information through communication with school and by building consensus with school about child's appropriate behaviour (Hill and Taylor, 2004).

Academic socialisation. Academic socialisation is another dimension of parental involvement in adolescence which involves parent-child discussions about education (McNeal Jr., 2001). Through socialising with their children, parents can communicate the expectations about desired outcomes they have for them and, thus, alter children's academic goals and aspirations (Bandura, 1989; Eccles and Wigfield, 1995; Ford, 1992; von Otter, 2014). Values, expectations and aspirations are very important in adolescence when planning the future after high school becomes salient. Academic socialisation, which can be both direct (verbal communication) and indirect (role-modeling), becomes a very important source for children to get much needed school curriculum and college application advice as well as a glimpse into what is possible and attainable (Benner et al., 2016).

Parental control. As teenagers assert their autonomy and become susceptible to their peers'

influence, "keeeping track" of them and parental watchfulness becomes more and more relevant (Dubow et al., 2009; Brooks-Gunn and Markman, 2005). Parental control, or monitoring, activities involve establishing bedtime routine, limiting TV and internet usage, approving or disapproving of child's friends, limiting child's exposure to violence, crime, and substance use (Kalil et al., 2012; Dishion et al., 1999).

2.2.3 Consequences and effectiveness of parenting practices

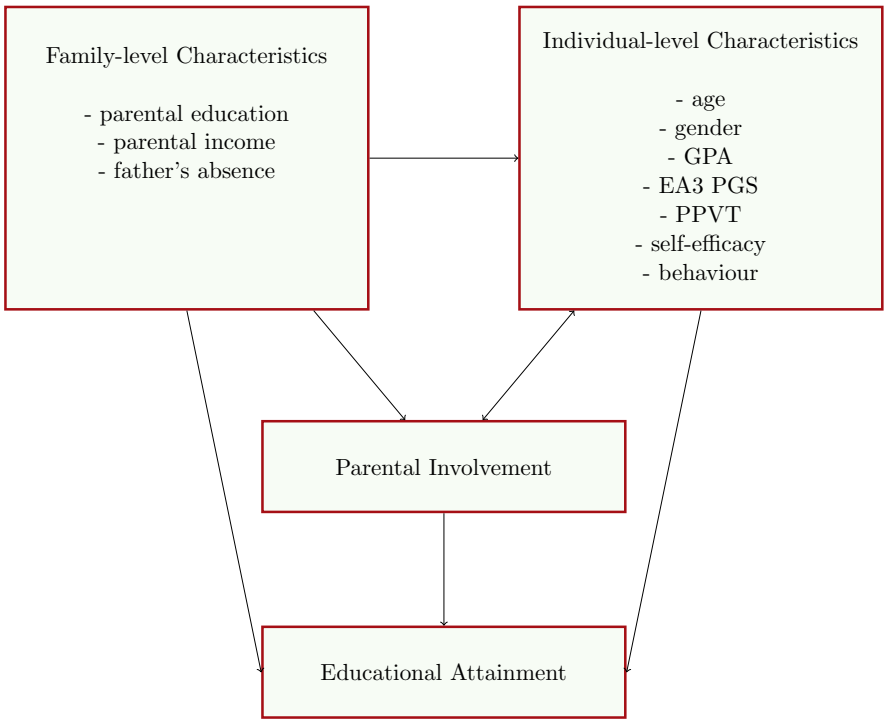
Research shows that parental involvement is very important for academic outcomes throughout the schooling process (Wang and Sheikh-Khalil, 2014; Hampden-Thompson et al., 2013; Hill and Tyson, 2009; Jeynes, 2005; Pomerantz et al., 2007; Hango, 2007; Gonzalez, 2004; Epstein, 2001). For young children parental time investment is found to be more important than the monetary one in helping them develop cognitive and non-cognitive skills (Doepke et al., 2019; Heckman, 2007; Waldfogel and Washbrook, 2011; Price, 2008, 2010). More involved parents help with school task completion (Guryan et al., 2008; Sayer et al., 2004; Mani et al., 2013; Shah et al., 2012), ensure language stimulation (Hart and Risley, 1995; Phillips, 2011), and organise academically enriching summer activities (Downey et al., 2004; Burkam et al., 2004; Chin and Phillips, 2004). A meta-analysis of studies on the effects of parental involvement for adolescents also show that parenting practices matter for their academic achievement in high school (Jeynes, 2007). The most important dimension of parental involvement during those years is consistently shown to be academic socialisation and encouragement of educational aspirations (Neuenschwander et al., 2007; von Otter, 2014). Unfortunately, a lot of research in this area suffers from insufficient control of important confounders. For example, Bronte-Tinkew et al. (2006) find that authoritarian parenting is associated with child's delinquent activities and substance abuse. The authors then recommend a less demanding and more responsive parenting style to avoid such problems. However, they do not take into account potential confounders such as socio-economic status: both behaviour problems and more authoritarian parenting are more prevalent in low-SES, working-class contexts. Such research design limitations abound in the literature on parental involvement. Eleanor Maccoby, one of the pioneers in the area, herself became rather disillusioned with the validity of results

produced by parenting research (Harris, 1998). Consequently, before measuring the effectiveness of parental involvement strategies for various child outcomes, one should take a step back and try to understand what explains differences in parental involvement strategies in the first place. Two sets of predictors might influence parental involvement: family and child characteristics (Brooks-Gunn and Markman, 2005).

2.2.4 Determinants of parental involvement

Factors influencing parental involvement can be grouped into two broad categories: family-level factors and individual, or child, characteristics. In this paper I would like to understand not only how all these characteristics are related to parental involvement and, in turn, educational attainment, but also how these two categories are associated with one another. Figure 2.1 depicts such interrelationships and highlights the importance of studying family and individual factors in tandem. As we can see from the graph, family-level variables might confound the relationship between individual characteristics and parental involvement as children from higher-SES household are more likely to, for example, have higher grades and be recipients of higher levels of parental involvement. In other words, we cannot quantify the importance of one category of factors without taking into account the other. Moreover, if we want to estimate the importance of parental involvement in the context of educational attainment, we need to carefully control for family and individual characteristics as they can both confound the relationship of interest.

Figure 2.1: Hypothesised mechanisms.



Note: GPA - Grade Point Average, EA3 PGS - educational attainment polygenic score, PPVT - Peabody Picture Vocabulary Test

In the paragraphs below, hypotheses about how all these variables are related will be specified. To facilitate the reading of the following theoretical sections, the hypotheses are already summarised in Table 2.1.

Table 2.1: Summary of hypotheses.

		A	B	C	D
		School-based activities	Academic socialisation (mother)	Academic socialisation (father)	Parental control
H1.1	Education	+	+	+	-
H1.2	Income	+	+	+	-
H1.3	Father's absence	+	+	NA	-
H2.1	Age	-	-	-	-
H2.2	Female	+	+	+	-
H2.3	GPA	+	+	+	-
H2.4	EA3 PGS	+	+	+	-
H2.5	PPVT	+	+	+	-
H2.6	Self-efficacy	+	+	+	-
H2.7	Behaviour problems	-	-	-	+

Notes: '+" mark an expected positive association between parental involvement measures and predictors and "-" - a negative association. GPA = Grade Point Average, EA3 PGS = educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test.

Family-level characteristics

The intensity and types of parental involvement depend on family characteristics. A desire to become involved is not enough — parents also need to possess the right skills, knowledge, time and energy (Catsambis, 2016). Parenting is one of the cultural reproduction processes which tends to be correlated across generations and be influenced by one's social milieu (Duncan et al., 2005). We already know that parents' socio-economic environment determines parenting practices directed towards young children (Kalil et al., 2012; Ermisch, 2008). However, little is known about the social gradient in parental involvement during adolescence. Sociological literature often uses an adjective "middle-class" to describe intensive, involved, educational goal-oriented practices of parents (Lareau, 2002). "Class" in the American context, which is the empirical focus of this paper, should not be understood in the classical sense of employment relations (Erikson and Goldthorpe, 1992) but rather as a much looser term proxying for occupation and education (Mills et al., 2006). As a consequence, here it will not be treated as a separate mechanism from the three family-level channels which might be producing parenting inequalities during adolescence: parental education, income, and family composition.

Parents who have higher levels of education can be more effective in school-based activities (Ream and Palardy, 2008). They possess more social capital because of their membership in networks and broader social structures (Portes, 1998). As portrayed by Lareau (2003), middle-class parents are generally willing to advocate for their children at school and more at ease when communicating with school administration as well as attending parent-teacher conferences. In contrast, working-class parents are more likely to trust teachers' authority and delegate their children's educational progress to teachers' discretion (Lareau, 2003, 2002; Hoover-Dempsey and Sandler, 1997). With regards to academic socialisation, highly-educated parents can serve as role models and communicate their past educational experiences to better inform their children's academic choices (Doepke et al., 2019). Income, although closely related to parental education, can also exert independent effects on the choice of parenting practices. Parents from low-income households often have inflexible work schedules and face transportation issues, which prevent them from attending school events (Hornby and Lafaele, 2011). Furthermore, such families also often live in dangerous neighborhoods, where higher levels of parental control and vigilance are rational practices (Baldwin et al., 1990). Family composition, especially in the US context where father absence is very high, is also worth mentioning. Research shows that single-parent and step-family homes are associated with generally lower levels of parental involvement because two-parent households tend to be more affluent and because time-intensive responsibilities can be shared by both parents (Dornbusch et al., 1987; Jones and White, 2000; Henry et al., 2011; Catsambis, 2016). Finally, literature also demonstrates that a minority status in terms of race also influences parental practices (Spera, 2005; Auerbach, 2007; Catsambis, 2001). However, in this chapter race will not be studied as the analysis that use the educational attainment polygenic score require limiting the sample to only individuals of European ancestry. Overall, I expect parents who have lower levels of education and income, or whose partner is absent from a household, to be on average less involved in school-based activities and academic socialisation of their children. However, I expect the opposite relationship with family factors and parental control. My hypothesis is that higher-SES households will tend to employ more permissive parenting marked by less parental control and that, in contrast, minority (non-white) parents will monitor their children more frequently even after conditioning on other characteristics (Hypotheses 1.1 - 1.3 in Table 2.1). In this table, "+" sign indicates that these family characteristics are expected to be positively associated with a given parental involvement measure while "-" shows

that the hypothesised direction of the relationship is negative. "NA" marks the fact that in the households where the father is absent there is no recorded academic socialisation with father and, thus, the association between these two variables cannot be measured.

Individual characteristics

As was already mentioned, not only family characteristics but also individual child factors can evoke particular types and levels of parental involvement. Since relationships between parents and children during adolescence can be difficult, it is likely that parents, rather than following their preferred parenting strategy, renegotiate and adjust them according to their child's needs (Gutman and Eccles, 2007). In this section we can draw on evidence from both between- and within-family studies. Getting at the associations between individual factors and parental involvement using between-family data will require careful control of family-level characteristics, which were discussed in the previous section. Unfortunately, some unmeasured family characteristics might still bias the results. To overcome this problem researchers have exploited a natural experiment of sibling comparison by looking at how parents make parenting decisions based on their children's differences. Existing research shows that parents distribute parental investments unequally between children based on their differences in endowments and behaviour (Gratz, 2018; Becker and Tomes, 1976; Behrman et al., 1982; Griliches, 1979; Aizer and Cunha, 2012; Behrman et al., 1994; Datar et al., 2010; Del Bono et al., 2012; Frijsters et al., 2013; Hsin, 2012; Restrepo, 2016; Rosales-Rueda, 2014; Yi et al., 2015). Little is known, however, whether and to what extent parents react differently to siblings in adolescence. Literature on intra-household allocation of parental investments also does not agree which direction this reaction takes: whether parents reinforce the difference by investing in better-endowed or better-behaved child (Becker and Tomes, 1976; Aizer and Cunha, 2012; Hsin, 2012; Frijsters et al., 2013; Gratz and Torche, 2016), compensate by diverting more resources towards a more disadvantaged child because of equity concern (Behrman et al., 1982; Pitt et al., 1990; Bharadwaj et al., 2018) or that parents are more or less consistent in the way they allocate investments to their children and do not react to children's individual characteristics (Almond and Currie, 2009; Lynch and Brooks, 2013).

I will investigate associations between six characteristics, which emerged from surveying the literature on parenting, and a novel, so far underexplored seventh characteristic — one's genetic predisposition to educational attainment measures as the educational attainment polygenic score. The first individual characteristic whose association with different parenting practices will be tested in this paper is child's age. The evidence on the direction of the association between parental involvement and age is mixed (deCastro et al., 2009). On the one hand, the older child is, the less likely parents become are to directly communicate with teachers or attend school events. Older children are also less likely to share details about school life limiting opportunities for a productive parent-child interaction. Parents of older children might also feel that their children's personalities, habits, and behaviour have fully formed and are, thus, less conducive to parental influences (Black et al., 2017). In other words, adolescents often demand autonomy and after a period of heightened monitoring parents tend to relent. On the other hand, older child's age might mean intensified academic socialisation about strategic course-taking and college applications (Hill and Tyson, 2009). Overall, however, I expect age to be negatively associated with school-based activities and academic socialisation. With regards to parental control, I expect the association to also be negative after controlling for other characteristics such as behaviour problems (Hypothesis 2.1 in Table 2.1). Gender can also have an effect on how parents allocate their parental resources. Girls generally are more likely to graduate from college, have a higher GPA, and display fewer behavioural problems (Buchmann and DiPrete, 2006; Blau and Kahn, 2000). Given these facts, parents might reinforce gender differences by focusing their time and effort on girls, who are more likely to succeed in education (Quadlin, 2016), or instead might compensate by allocating more parental involvement to boys (Rosin, 2012; Tyre, 2008). I, nevertheless, hypothesise that because parental involvement results from a negotiation between a child and a parent, it is more likely that girls, who tend to be emotionally closer than boys with parents, encourage and, consequently, receive more parental involvement. However, I expect parents to exert more control for boys than girls because, even controlling for child's actual behaviour, boys are more likely to start misbehaving in the future (Hypothesis 2.2 in Table 2.1). The status attainment model suggests that parents might reinforce educational disparities through active socialisation of expectations for higher levels of education to those performing better at school (Blau and Duncan, 1967; Featherman and Hauser, 1978; Sewell et al., 1969) because achievement provides evidence about children's underlying cognitive skills and

ultimate educational attainment (Becker, 1964; Becker and Tomes, 1976). The relationship can go in a different direction as well: those receiving higher levels of parental involvement might later have better academic performance, measured as GPA. While given the data I cannot distinguish between these two mechanisms, quantifying an association between parental involvement and GPA is still descriptively interesting. Using sibling fixed effects might also help us come a step closer to estimating an effect of performance on parental involvement. It is hard to imagine parents who would direct more parental involvement toward one child for no reason. I expect that even after controlling for child differences in age, gender, or verbal ability, academic achievement will still serve as a predictor for higher levels of parents' school-based activities and academic socialisation. Better performing children should also evoke higher levels of parental trust and, consequently, lower levels of parental control (Hypothesis 2.3 in Table 2.1). Cognitive ability, net of academic achievement, can also be associated with parental involvement a child receives from her parents. They might react to higher ability by either investing more time and resources to maximise the child's potential or, in contrast, by relaxing their parenting approach because of an expectation that a gifted child will succeed without much parental help. Twin studies conducted in early childhood, however, show that highly educated parents provide more cognitive stimulation to higher-ability children, thus, reinforcing ability differences (Gratz and Torche, 2016; Dendir, 2014). As a consequence, we should expect cognitive ability to be positively associated with school-based activities and academic socialisation, controlling for other family and individual factors. As is the case with academic performance, I expect parents to be less controlling with higher-ability children (Hypothesis 2.5 in Table 2.1). Parents might react not only to child's cognitive ability but also to her non-cognitive skills when deciding on a particular level of parental involvement. Higher non-cognitive skills, such as self-efficacy, might facilitate the relationship between a parent and a child and, thus, encourage parents to be more involved. On the other hand, children with lower levels of non-cognitive skills might be perceived by their parents as in need of more nurturing and involvement (Leight and Liu, 2018). Limited existing research gives mixed evidence on the relationship between non-cognitive skills and parenting. I will, nevertheless, hypothesise that their association follows the same pattern as above mentioned child performance measures, i.e. they are positively related with school-based activities and academic socialisation and negatively with more authoritarian parenting strategies such as close control and monitoring (Hypothesis 2.6 in Table 2.1). Research on how parents react

to child's behavioural challenges is also scant. Externalising behavior might elicit more authoritarian parenting and, consequently, more parental control. On the other hand, more permissive parenting in the first place might result in more conduct problems later on. In general, I expect parents, conditioning on other factors, to devote more time to parental control when faced with child's challenging behaviour, potentially at an expense of school-based activities and academic socialisation, leading to a positive association between behaviour problems and parental control and a negative association with the rest of my parental involvement measures (Hypothesis 2.7 in Table 2.1). Finally, I will also explore the association between parental involvement and child's genetic endowment. Unlike many other individual characteristics I am interested in, such as academic performance or cognitive ability, the educational attainment polygenic score is exogenous to parental involvement if sufficiently controlled for family-level factors or compare siblings within families. The polygenic score (EA3 PGS), which measures individual's genetic propensity to education, has been used in multiple studies to explain disparities in education attainment both across and within families (Selzam et al., 2019; Domingue et al., 2015; Lee et al., 2018; Belsky et al., 2018). However, only few researchers have looked at whether there is an evocative gene-environment correlation between the educational attainment polygenic score and parental involvement a child received (Wertz et al., 2019; Sanz-de Galdeano and Terskaya, 2019). In other words, higher genetic propensity to education, which can lead to higher ability and interest in academic pursuits, might elicit more parental investment in terms of involvement in school matters and educational communication. I also expect parents to trust their better endowed children more and, thus, have a more relaxed approach to parental control (Hypothesis 2.4 in Table 2.1).

2.3 Data

2.3.1 Dataset

The data used in this paper come from The National Longitudinal Study of Adolescent to Adult Health (Add Health) administered by the University of North Carolina, Chapel Hill (Harris et al., 2019). The collection of the data was mandated by the U.S. Congress in order to study health from

a longitudinal perspective. The study contains five waves. Wave I was collected in 1994-1995. In-school questionnaires were administered to 90,118 students clustered in a nationally representative random sample of 80 schools. Between September 1994 and April 1995 in-home interviews lasting for 45-60 minutes were conducted with a random subsample of 20,745 students from the original sample of 90,118. For Wave IV in 2008 respondents were genotyped using Oragene saliva collection. The dataset is uniquely suited to study the topic of parental involvement because of its extensive set of questions concerning parenting activities that can be found both in the parental and the student questionnaires. Moreover, rich information on students' individual characteristics such as academic achievement, behaviour and non-cognitive skills is also available. Add Health also has a sizeable sample of siblings allowing us to study intra-household dynamics in adolescence.

2.3.2 Sample

My sample was significantly reduced due to missing values in my main dependent parental involvement variables and because only around half of all the respondents have been genotyped and, thus, have their educational attainment polygenic scores. A full graphical description of sample reduction can be seen in Figure 2.A.1 in the Appendix. For example, 6,602 respondents have information on their parents' school-based activities. Out of these 6,602 individuals, 3,847 respondents have their educational attainment polygenic scores. To address the issue of other explanatory variables having missing values I used the technique of multiple imputation with chained equations. As recommended by von Hippel (2007), I do not use imputed values of the dependent variables (parental involvement and educational attainment). Such method is known as "multiple imputation, then deletion" (MID): multiple imputation uses all available cases but then the cases with imputed dependent values are deleted. MID has been shown to improve efficiency and reduce the bias from misspecifying the outcome imputation model (von Hippel, 2007). However, Sullivan et al. (2015) demonstrate that MID might not appropriate if an imputation model uses "auxiliary" variables in addition to those used in a substantive model: MID cannot benefit from a potential of those auxiliary variables to predict missing values in an outcome variable. Since I use the same variables both for modelling and imputation, the von Hippel's approach is considered to be the best imputation

method (Mustillo, 2012). Consequently, my analytical samples have different sizes depending on a specific dependent variable: 3,847 for school-based activities, 8,579 for academic socialisation with mother; 6,455 for academic socialisation with father, and 8,941 for parental control.

To run the within-family analyses, I restricted my samples to full siblings and dizygotic twins to ensure variation in their educational attainment polygenic scores. The sample reduction for my sibling subsample can be seen in Figur 2.A.2 in the Appendix. As is the case for the full sample, I use different samples depending on a specific dependent variable I am explaining: 760 siblings for school-based activities, 1,618 - for academic socialisation with mothers, 1,281 for academic socialisation with father, and 1,668 for parental control. Overall, all full and sibling models I also run on the unimputed dataset producing substantively similar results to the imputed models. More details about the imputation process can be found in the Appendix.

2.4 Measurement

2.4.1 Parental involvement

Three broad categories of parental involvement are measured and investigated in this paper: school-based activities, academic socialisation, and parental control (Figure 2.2).

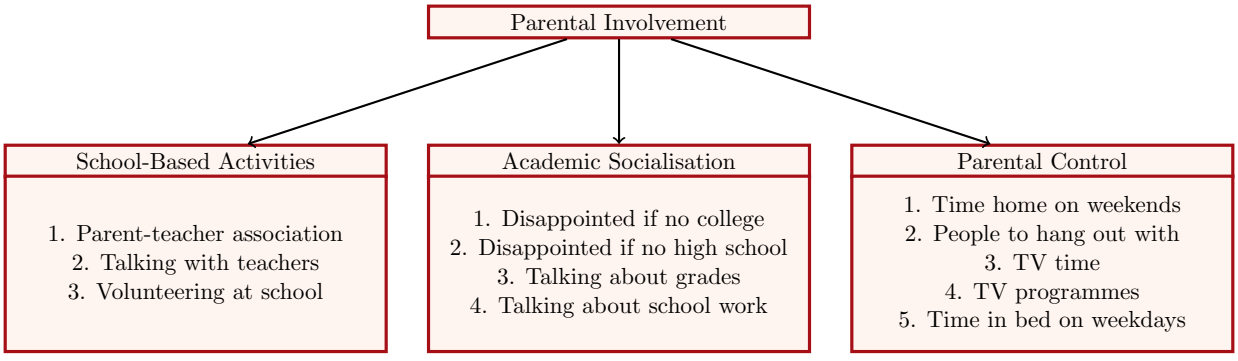


Figure 2.2: Dimensions of parental involvement.

School-based activities measure the extent to which parents actively participate in their child's

school life. Three questions capture this dimension of parental involvement: 1) *Are you a member of a parent/teacher organisation?*; 2) *Have you talked with your child's teachers about his/her school work this school year, either informally or in a regularly scheduled parent-teacher conference?*; 3) *During this school year, have you participated in school fund-raising or done volunteer work for your child's school, such as supervising lunch, chaperoning a field trip, etc.?*. The responses to each of the three questions are binary — either yes (1) or no (0). 34% of the parents are members of parent/teacher organizations; 75% report talking regularly with child's teacher; 35% volunteer at child's school (Table 2.A.3). The items are then added up and, hence, the composite indicator of school-based activities can range from 0 to 3. A standardised version of the indicator (mean = 0, SD = 1) will be used in the analyses. The choice of items and their summation to construct a composite indicator is borrowed from the existing literature (Wang and Sheikh-Khalil, 2014; Benner et al., 2016). Cronbach's alpha for this measure is 0.61. The high missingness for school-based activities might hint at the sample selection on this variable, which might bias the results: parents with missing information on school-based activities might be even less involved than those that have reported low levels of involvement. However, after running a logistic regression to explain the missingness, I found no evidence that families with the missing data on this variable were systematically different from other families.

The second dimension of parental involvement is academic socialisation. It is measured with the following questions in Add Health: 1) *On a scale of 1 to 5, where 1 is low and 5 is high, how disappointed would your parent be if you did not graduate from college?*; 2) *On a scale of 1 to 5, where 1 is low and 5 is high, how disappointed would your parent be if you did not graduate from high school?*; 3) *Have you talked about your school work or grades with your parent in the past 4 weeks?*; 4) *Have you talked about other things you're doing in school with your parent in the past 4 weeks?*. These questions are asked about one's mother and father separately. Descriptive statistics for them can be seen in Table 2.A.3. The selection of items to be included in this composite indicator is theoretically driven as it is supposed to cover educational aspirations or representations of desired states and outcomes (such as disappointment if low levels of education are attained) and regular meaningful communication about concrete academic actions and progress through which those desired states can be expressed (Benner et al., 2016; Bandura, 1989; Ford, 1992). Both mothers

and fathers on average would be disappointed if their child did not attend college and seriously disappointed if the child did not graduate from high school. 63% of mothers report talking about grades with their children while 52% of fathers report doing so. Fewer but still around half of the parents talk about other school work with their children: 52% of the mothers and 44% of the fathers do so. The first two questions were made into binary ones where 1 encodes 5 or 4 in the original scale. The dummmy answers to the four questions were added to construct an index of academic socialisation with mother and fathers separately and then standardised. Cronbach's alpha is 0.63 and 0.6 for fathers and mothers respectively. Factor analysis shows that the four items represent two facets: aspirations (item 1 and 2) and communication (item 3 and 4). However, following the example of Wang and Sheikh-Khalil (2014), who see academic socialisation as a product of both a higher-level transmission of aspirations and every-day communication, I combine all four items into one variable.

The last dimension of parental involvement which is explored in this paper is parental control, in the literature also known as monitoring or management. Five questions capture the level of control parents exert on their children: 1) *Do your parents let you make your own decisions about the time you must be home on weekend nights?*; 2) *Do your parents let you make your own decisions about the people you hang around with?*; 3) *Do your parents let you make your own decisions about how much television you watch?*; 4) *Do your parents let you make your own decisions about which television programs you watch?*; 5) *Do your parents let you make your own decisions about what time you go to bed on week nights?*. A value of 1 encodes children not being able to make their own decisions regarding these matters and, thus, a higher level of parental control. The proportions of children who answer these questions negatively can be found in Table 2.A.3. The composite indicator of parental control is produced following Fish et al. (2020) by adding dummy values of each of the five questions and then standardising it to have mean of 0 and a standard deviation of 1. Cronbach's alpha for this measure is 0.67. Factor analysis confirms the existence of a single factor structure. The summary of all dimensions with the corresponding items in the questionnaire can be found in Table 2.A.1 in the Appendix.

2.4.2 Educational attainment

Respondents' educational attainment is measured in Wave IV (in 2008). In the dataset this measure is provided as a categorical variable. For the analyses this variable has been recoded to a numerical one to facilitate analysis and interpretation of results. Here is the list of all categories in the original variable and how it has been recoded to years of education: 1) *8th grade or less* - 8 years; 2) *some high school* - 10 years; 3) *high school graduate* - 12 years; 4) *some vocational/technical training (after high school)* - 13 years; 5) *completed vocational/technical training (after high school)* - 14 years; 6) *some college* - 16 years; 7) *completed college (bachelor's degree)* - 16 years; 8) *some graduate school* - 17 years; 9) *completed a master's degree* - 18 years; 10) *some graduate training beyond a master's degree* - 19 years; 11) *completed a doctoral degree* - 20 years; 12) *some post baccalaureate professional education* - 18 years; 13) *completed post baccalaureate professional education* - 19 years. Average years of education in our sample is 14.87 years while the median is 16. The standard deviation is 2.33 years and values range from 8 to 20. The distribution of years of education is shown in Figure 2.A.8 in the Appendix.

2.4.3 Family characteristics

Parents' education. The measures of parental education in Add Health are provided as a categorical variable. In this paper the recoded numerical values will be used. Here is the list of categorical values and the years of education each category corresponds to: 1) *8th grade or less* - 8 years; 2) *less than 8th grade/didn't graduate from high school* - 10 years; 3) *business/trade/vocational school instead high school* - 10 years; 4) *high school graduate* - 12 years; 5) *GED* - 12 years; 6) *business/trade/vocational school after high school* - 13 years; 7) *attended college but didn't graduate* - 14 years; 8) *graduated from college/university* - 16 years; 9) *professional training beyond 4-year college/university* - 18 years; 10) *never went to school* - 0 years. As can be seen from Table 2.2, mothers on average have 13.31 years of education (with a standard deviation of 2.60 years) and father have slightly more education on average - 13.49 years (with a standard deviation of 2.78 years).

Household income. Household income is measured with a question in the parental questionnaire: *About how much total income, before taxes did your family receive in 1994? Include your own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources.* The values of income range from 0 to 999 (in thousands of dollars) in the sample. The mean income is 47.7 while the median is 40, showing that the distribution is slightly skewed to the right. The standard deviation is 56.35 (Table 2.2).

Father's absence. I use a dummy variable which measures whether a father figure is present or absent in a given household. Although father absence is not explicitly measured with a distinct variable, I can get this information from the father's educational attainment variable because absent fathers are coded as a legitimate skip in this question. Overall, in 30% of the families fathers or father figures are absent.

2.4.4 Individual characteristics

Age. The age of the respondents ranges from 11 to 21 with a mean and a median of 16 (Table 2.2). The participants of the Add Health are enrolled in 7 to 12 grades roughly corresponding to being 13 to 18 years old. However, there are some students who have skipped or repeated grades, thus, increasing the range of possible ages.

Gender. Gender is self-reported and binary. The sample contains 51% girls and 49% boys (Table 2.2).

Educational attainment polygenic score. A polygenic score is a single aggregate measure of the multiple genetic effects correlated with educational attainment. It includes only linear combinations of the effects of genetic variants. A polygenic score of an individual i is equal to the sum of the allele counts a_{ij}(0,1 or 2) from each SNP $j = 1$,..M, multiplied by a weight w_j: $PGS_i = \sum_{j=1}^{m} w_j a_{ij}$. The weights are association coefficients produced in the educational attainment GWAS. Once constructed, polygenic scores can be calculated for any individual in any sample containing molecular genetic data. Individual's educational attainment polygenic score can then be interpreted as a

single quantitative measure of one's genetic propensity for educational attainment. In the analysis for this paper the polygenic score from the latest educational attainment GWAS will be used (Lee et al., 2018). I do not construct the polygenic score myself but rather use the one directly provided by the Add Health. The Add Health-produced polygenic score is constructed using the PRSice method (Chang et al., 2015; Euesden et al., 2015). This method uses pruning and thresholding techniques. First, it selects independent SNPs based on linkage or, in other words, discards SNPs in linkage with other SNPs. The PRSice method then specifies different p-value thresholds and for each threshold includes only SNPs below that threshold. Finally, the polygenic score which has the highest predictive power is chosen (Mills et al., 2020a). In my sample the polygenic score is standardised to have a mean of 0 and a standard deviation of 1. My analyses also include 10 principal components to control for potential population stratification.

Academic achievement. Academic achievement is measured as GPA. Respondents self-report letter grades (A, B, C, or D) for four subject: English, mathematics, science, and history. I assign a numerical grade to each letter grade (A - 4, B - 3, C - 2, D - 1) and calculate an average of all four subjects to get a GPA measure. The mean GPA in the Add Health dataset is 2.80 with a standard deviation of 0.76 (Table 2.2). In the analyses GPA is standardised to have a mean of 0 and a standard deviation of 1.

Cognitive skills. Cognitive skills are not extensively measured in the dataset. Apart from academic achievement, the only other proxy measure is verbally administered Peabody Picture Vocabulary Test (PPVT) which measures students' language ability and literacy. The total PPVT score in my sample ranges from 13 to 146 with a mean of 100 and a standard deviation of 15 (Table 2.2).

Self-efficacy. Self-efficacy is measured with seven binary questions: whether the respondent 1) deals well with problems; 2) does not get easily upset by problems; 3) does not just use gut feeling when reaching decisions; 4) uses facts when solving problems; 5) thinks of different ways to solve problems; 6) thinks about different alternatives when making important decisions; 7) analyses consequences when a problem is resolved. Cronbach's alpha for these items is 0.69. I extract a factor whose values range from -2.04 to 0.8 and have a mean of 0 and a standard deviation of 0.80.

Behaviour problems. Behaviour problems are measured with ten binary questions: whether the respondent 1) has been away from home without permission; 2) has been absent without an excuse; 3) has been suspended; 4) has had sex; 5) has smoked; 6) has painted a graffiti; 7) has damaged property; 8) has lied to parents; 9) has participated in a physical fight; 10) has acted unruly. Cronbach's alpha for these items is 0.72. I extract a factor whose values range from -1.06 to 2.65 and have a mean of 0 and a standard deviation of 0.82.

Table 2.2: Descriptive statistics of the Add Health dataset.

	Full sample					Sibling sample				
Variables	Mean	Median	SD	Range	N	Mean	Median	SD	Range	N
School-based activities	0	-0.35	1.00	-1.37 - 1.71	8,561	0.04	-0.35	1.01	-1.37 - 1.71	1,367
Academic socialisation (mother)	0	0.20	1.00	-2.57 - 1.12	19,316	0.01	0.20	1.02	-2.57 - 1.12	2,983
Academic socialisation (father)	0	0.36	1.00	-2.28 - 1.20	14,340	0.03	0.36	1.01	-2.28 - 1.24	2,344
Parental control	0	-0.44	1.00	-1.22 - 2.70	20,218	0.06	-0.44	1.00	-1.22 - 2.70	3,080
Child's age	15.66	16.00	1.75	11.00 - 21.00	20,728	15.62	16.00	1.70	11.00 - 21.00	3,134
Child's gender					20,743					3,136
Female	0.51				10,480	0.50				1,554
Male	0.49				10,263	0.50				1,582
EA3 PGS	0	0.02	1.00	-4.13 - 3.40	6,011	0.01	0.015	1.00	-3.00 - 3.40	1,141
Years of education	14.19	14.00	2.21	8.00 - 21.00	15,697	14.19	14.00	2.25	8.00 - 21.00	2,631
GPA	2.80	2.75	0.76	1.00 - 4.00	15,380	2.90	3.00	0.74	1.00 - 4.00	2,369
PPVT	99.66	100.00	15.17	13.00 - 146.00	19,713	99.21	99.00	14.70	13.00 - 146.00	3,012
Self-efficacy	0	0.12	0.8	-2.04 - 0.80	20,365	0	0.20	0.81	-2.00 - 0.81	3,083
Behaviour problems	0	-0.18	0.82	-1.06 - 2.65	19,881	0	-0.21	0.82	-1.01 - 2.77	3,011
Parental income	45.73	38.00	51.62	0 - 999.00	15,351	46.00	38.00	52.00	0 - 999.00	2,412
Parents' years of education										
Mother	13.14	12.00	2.68	0 - 18.00	18,413	13.10	12.00	2.68	0 - 18.00	2,843
Father	13.37	14.00	2.85	0 - 18.00	13,442	13.41	14.00	2.83	0 - 18.00	2,222
Number of siblings	1.63	1.00	1.49	0 - 14.00	20,429	2.30	2.00	1.60	1 - 13.00	3,123
Father's presence					20,745					
Present	0.70				14,463	0.75				2,361
Absent	0.30				6,282	0.25				775

Notes: GPA = Grade Point Average, EA3 PGS = educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test. The descriptive statistics are for all the Add Health respondents - for those with the EA3 PGS score available and without it.

2.5 Analytical strategy & methods

My analysis will consist of two parts. In the first part I will explain my four measures of parental involvement with individual and family-level characteristics using both between-families (OLS) and within-family (sibling fixed effects) approaches. In the second part I will explain educational attainment with parental involvement, controlling for family and individual factors.

2.5.1 Explaining parental involvement

I will first use a full sample of respondents and run OLS regressions explaining each parental involvement measure separately. The first model for each measure will include individual characteristics:

$$ParInvolve_i = \alpha + \boldsymbol{\beta} \times \boldsymbol{I_i} + \epsilon_i \tag{2.1}$$

where $ParInvolve_i$ indicates one of the four parental involvement measures; $\boldsymbol{I}$ represents a vector of seven individual characteristics (age, gender, GPA, educational attainment polygenic score, PPVT, self-efficacy, and behaviour problem index) and 10 principal components.

The second model will add family-level factors to the first model:

$$ParInvolve_i = \alpha + \boldsymbol{\beta} \times \boldsymbol{I_i} + \boldsymbol{\gamma} \times \boldsymbol{F_i} + \epsilon_i \tag{2.2}$$

where $\boldsymbol{F_i}$ represents a vector of three family factors (parent's education, household income, and father's absence).

I will also compare the OLS estimates with the estimates produced using sibling fixed effects regressions. The goal is to see whether individual factors have similar association sizes and directions using both approaches: controlling only for measured (OLS) and all unmeasured family-level factors that vary across families (sibling fixed effects) which could be correlated with individual factors and parental involvement. This model takes the following form:

$$ParInvolve_i = \alpha + \boldsymbol{\beta} \times \boldsymbol{I_i} + \sum_{k=2}^{N} \mu_k \times S_k(i) + \epsilon_i \tag{2.3}$$

where $S_k(i)$ is 1 if sibling i is in family k and 0 otherwise, with $k = 1$ family excluded as the reference.

2.5.2 Explaining educational attainment

The goal of the second part of the analysis is explaining differences in educational attainment by using parental involvement measures while controlling for individual and family factors. I will sequentially add variables into the OLS regressions and, thus, have three models for each parental involvement measure:

$$EduYears_i = \alpha + \beta \times ParInvolve_i + \epsilon_i \tag{2.4}$$

$$EduYears_i = \alpha + \beta \times ParInvolve_i + \boldsymbol{\gamma} \times \boldsymbol{I_i} + \epsilon_i \tag{2.5}$$

$$EduYears_i = \alpha + \beta \times ParInvolve_i + \boldsymbol{\gamma} \times \boldsymbol{I_i} + \boldsymbol{\theta} \times \boldsymbol{F_i} + \epsilon_i \tag{2.6}$$

where $EduYears_i$ denotes attained years of education; $ParInvolve_i$ indicates one of the four parental involvement measures; $\boldsymbol{I}$ represents a vector of seven individual characteristics (age, gen-

der, GPA, educational attainment polygenic score, PPVT, self-efficacy, and behaviour problem index) and 10 principal components; $\boldsymbol{F_i}$ is a vector of three family factors (parent's education, household income, and father's absence). Since there are many different model specifications even using the OLS (i.e. different combinations of both individual- family-level control variables), I run 512 different linear combinations for each parental involvement measure. The specification curves are reported in Figure 2.A.13.

Some of the diagnostic plots of the Gauss-Markov assumptions for using the OLS when modelling years of education can be found in the Appendix (Figures 2.A.9, 2.A.10, 2.A.11, 2.A.12). Overall, the fitted values slightly overpredict the educational years for those on the lower end of the educational attainment and underpredict for those at the higher end but otherwise the residuals average 0. There is little evidence for heteroskedasticity.

I will also use sibling fixed effects to estimate the association between parental involvement measures and years of education. I will first include only parental involvement as an explanatory measure and then control for individual characteristics that may vary between siblings:

$$EduYears_i = \alpha + \beta \times ParInvolve_i + \sum_{k=2}^{N} \mu_k \times S_k(i) + \epsilon_i \tag{2.7}$$

$$EduYears_i = \alpha + \beta \times ParInvolve_i + \boldsymbol{\gamma} \times \boldsymbol{I_i} + \sum_{k=2}^{N} \mu_k \times S_k(i) + \epsilon_i \tag{2.8}$$

This modelling approach does not allow us to rule out endogeneity — parenting and child's characteristics simultaneously influencing each other. Instead, the analyses performed in this chapter are intended to inform future research about potential biases in parenting research. Better research designs in the form of more extensive panel studies, natural experiments, and RCTs, are reviewed in the discussion section of this chapter.

2.6 Results

2.6.1 Explaining parental involvement

Results from the first stage of the modelling can be seen in Table 2.3. The estimated associations between different dimensions of parental involvement and my explanatory variables closely follow hypothesised patterns (the summary of whether I can confirm my hypotheses can be found in Table 2.5). As expected with Hypothesis 2.1, age is negatively associated with all the parental involvement measures. GPA and PPVT are very strong predictors of all parental involvement measures: they are positively associated with school-based activities and academic socialisation and negatively associated with parental control (Hypothesis 2.3 and 2.5). PPVT is a surprisingly good predictor of parental control: a 1 standard deviation increase in PPVT corresponds to a 0.125 standard deviation decrease in parental control. Self-efficacy seems to matter only for academic socialisation (Hypothesis 2.6). However, some results do not support my hypotheses. Controlling for other characteristics, girls report lower levels of their parents' school-based activities and parental control (Hypothesis 2.2). Gender does not seem to have any association with one's academic socialisation with either mother or father. In addition to being negatively associated with school-based activities and academic socialisation, behaviour problems are also associated with lower levels of parental control (Hypothesis 2.7). Finally, the educational attainment polygenic score is only modestly associated with school-based activities and academic socialisation and not at all related to parental control (Hypothesis 2.4).

Models 2, 4, 6, 8 in Table 2.3 also include family-level characteristics. First, it is important to note that after controlling for these characteristics the estimated associations with individual factors decrease, demonstrating that family-level and individual-level factors are connected and should be studied together to avoid confounding-induced bias. As expected, higher parental education and income are strongly positively associated with school-based activities and academic socialisation (Hypotheses 1.1 and 1.2). Father absence is negatively associated with school-based activities and positively associated with academic socialisation with mother (Hypothesis 1.3). Contrary

to my hypotheses, net of measured individual and other family characteristics, family income and education do not explain the differences in parental control (Hypotheses 1.1 and 1.2). The strongest predictor of parental control is father's absence: having an absent father decreases parental control by a 0.139 standard deviation (Hypothesis 1.3).

Table 2.3: Estimated coefficients from the OLS regressions associating parental involvement with individual- and family-level characteristics.

	School-based activities		Academic socialisation				Parental control	
			Mother		Father			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.760***	0.631***	0.280*	0.166	0.238†	0.136	3.354***	3.397***
	(0.192)	(0.188)	(0.124)	(0.124)	(0.134)	(0.134)	(0.110)	(0.111)
Age	-0.042***	-0.045***	-0.023**	-0.023**	-0.016†	-0.017*	-0.214***	-0.215***
	(0.012)	(0.012)	(0.008)	(0.007)	(0.009)	(0.008)	(0.007)	(0.007)
Female	-0.116**	-0.106**	0.028	0.022	-0.029	-0.025	-0.053*	-0.045†
	(0.041)	(0.040)	(0.026)	(0.026)	(0.029)	(0.028)	(0.023)	(0.023)
GPA	0.160***	0.114***	0.120***	0.104***	0.139***	0.115***	-0.041	-0.043**
	(0.023)	(0.022)	(0.015)	(0.015)	(0.017)	(0.017)	(0.013)	(0.013)
EA3 PGS	0.056**	0.027	0.043**	0.030*	0.040**	0.025†	-0.011	-0.009
	(0.021)	(0.020)	(0.014)	(0.014)	(0.015)	(0.015)	(0.012)	(0.012)
PPVT	0.163***	0.105***	0.112***	0.080***	0.088***	0.045*	-0.125***	-0.123***
	(0.026)	(0.026)	(0.017)	(0.018)	(0.019)	(0.019)	(0.015)	(0.016)
Self-efficacy	-0.022	-0.009	0.063***	0.064***	0.063***	0.068***	-0.011	-0.011
	(0.021)	(0.020)	(0.013)	(0.013)	(0.014)	(0.014)	(0.012)	(0.012)
Behaviour problems	-0.090***	-0.082***	-0.050***	-0.052***	-0.096***	-0.089***	-0.091***	-0.087***
	(0.022)	(0.022)	(0.015)	(0.015)	(0.016)	(0.016)	(0.013)	(0.013)
High mother's education		0.392***		0.207***		0.011		-0.012
		(0.042)		(0.028)		(0.032)		(0.025)
Income		0.124***		0.048**		0.066***		-0.027†
		(0.031)		(0.015)		(0.016)		(0.014)
Father absent		-0.096†		0.089**				-0.159***
		(0.049)		(0.032)				(0.029)
High father's education						0.239***		
						(0.033)		
N	2,495	2,495	5,636	5,636	4,646	4,646	5,882	5,882
R^2	0.088	0.139	0.063	0.076	0.076	0.096	0.178	0.182

Notes: GPA = Grade Point Average, EA3 PGS = educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test. SEs in parentheses. †p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Data source: the Add Health.

The second stage in this part of the analysis involves running sibling fixed effects and comparing these results with the results produced using the OLS approach. Before we do this, we should first check whether there is any variation in parental involvement between siblings located in the same household. For this purpose I report interclass correlations (ICC) for each of the four dimensions of parental involvement (Table 2.A.10 in the Appendix). First, I look at the full sibling sample and find that school-based activities for each sibling are very correlated (ICC = 0.84). It means that parents use an almost identical approach to school-based activities for all their children. In other words, they

do not seem to respond much to their children's individual characteristics and adjust the intensity of their school-based activities accordingly. In contrast, academic socialisation and parental control are only moderately correlated with ICC at around 0.25 for each dimension. Nevertheless, instead of genuine differences in parental choices regarding different parenting dimensions, these diverging ICCs can reflect the fact that school-based activities are reported by parents themselves with the parental questionnaire while other dimensions are reported by adolescents. The differences in ICCs could also reflect birth order effects: parents might become more lenient and less inclined to actively academically socialise with a younger child while maintaining similar levels of school-based activities, either out of existing expectations from school officials, previous commitments, or simply a habit. I also look at whether the ICCs differ by family SES (mother's and father's education, household income, and father's presence). However, I do not see any differences by SES: ICCs for both lower- and higher-SES siblings are fairly similar.

Since there is quite a lot of variation between siblings in their reported parental involvement, we can proceed to the analysis of sibling fixed effects (Table 2.4). In this part of the analysis I investigate which individual characteristics of a child are related to differential responses by parents in a single family. Within families, age, which in the sibling context can also be interpreted as birth order, does not matter for school-based parental involvement except for a strong negative association with parental control: every additional year for a child is associated with a 0.162 standard deviation decrease in parental control (Hypothesis 2.1). In contrast to the OLS results, Table 2.4 shows that females are more likely rather than less likely to be controlled by parents (Hypothesis 2.2). Another surprise is that within families higher GPA elicits fewer school-based parental activities (Hypothesis 2.3). The result, however, is not statistically significant. Somewhat unexpectedly, self-efficacy is related to academic socialisation with parents (Hypothesis 2.6). However, it is hard to tell which direction the association goes: whether more self-efficacious children evoke more academic socialisation with parents or whether more involvement from parents lead to more self-efficacious children. The negative association between behaviour problems and parental control also calls into question the direction of this association (Hypothesis 2.7). However, here, the more plausible mechanisms is that less parental control leads to more behaviour problems.

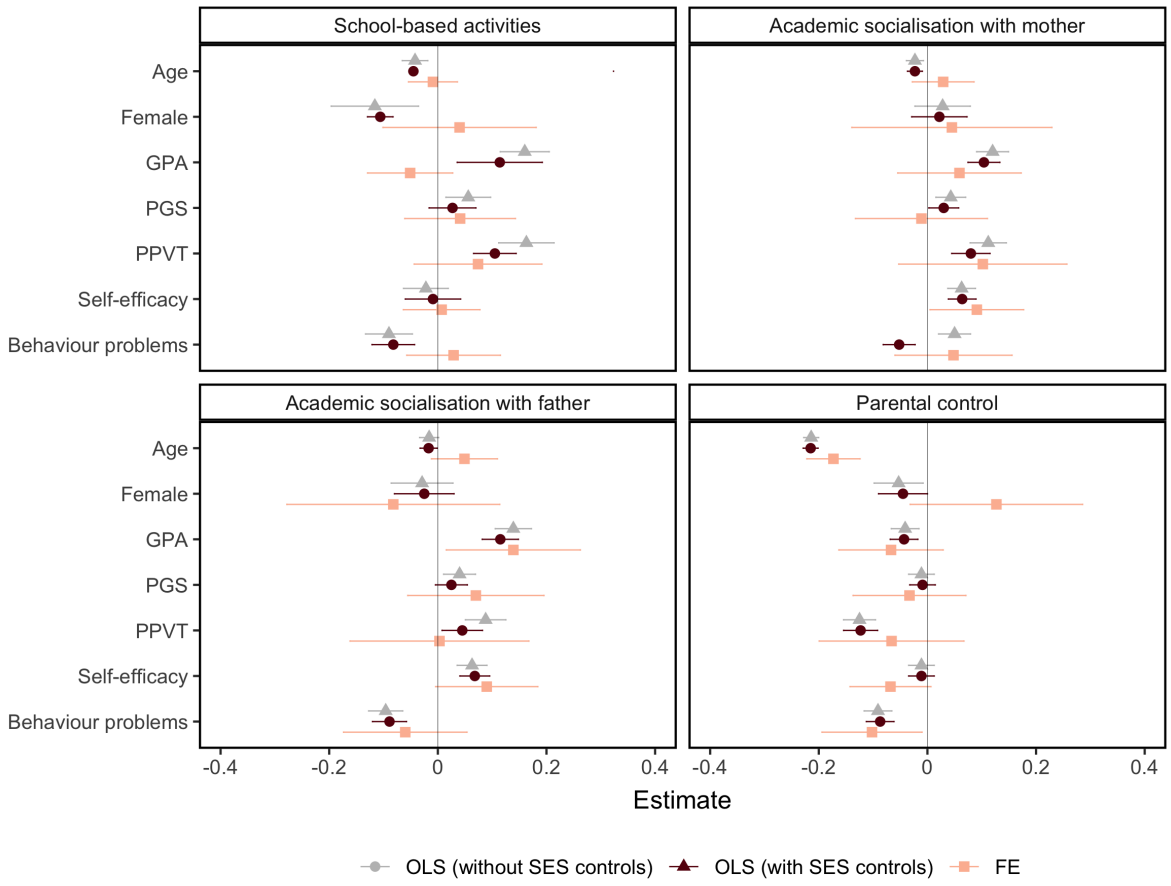
Table 2.4: Estimated coefficients from the OLS regressions associating parental involvement with individual-level characteristics using sibling fixed effects.

	School-based activities	Academic socialisation		Parental control
		Mother	Father	
	(1)	(2)	(3)	(4)
Constant	2.134***	-1.487	1.834	4.071***
	(0.555)	(0.950)	(1.455)	(0.823)
Age	-0.009	0.029	0.049	-0.173***
	(0.023)	(0.029)	(0.031)	(0.025)
Female	0.040	0.045	-0.082	0.127
	(0.072)	(0.094)	(0.100)	(0.081)
GPA	-0.051	0.059	0.139*	-0.067
	(0.040)	(0.058)	(0.063)	(0.049)
EA3 PGS	0.041	-0.011	0.070	-0.033
	(0.052)	(0.062)	(0.064)	(0.053)
PPVT	0.074	0.102	0.003	-0.066
	(0.060)	(0.079)	(0.084)	(0.068)
Self-efficacy	0.007	0.091*	0.090†	-0.068†
	(0.036)	(0.044)	(0.048)	(0.038)
Behaviour problems	0.029	0.048	-0.060	-0.102*
	(0.044)	(0.055)	(0.058)	(0.047)
N	508	1,088	942	1,123

Notes: GPA = Grade Point Average, EA3 PGS = Educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test. SEs in parentheses. †p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Data source: the Add Health.

Figure 2.3 provides a comparison of the association sizes between individual characteristics and each parental involvement measure in three different models: the OLS model not controlling for family-level factors, the OLS model controlling for family-level factors, and family fixed effects. As we can see from the figure, comparisons should be made cautiously as the point estimates from the family fixed effects part of my analysis have very wide error bars. What is visible, however, is that there is a substantial confounding of individual factors by family-level factors and the inclusion of them reduces the associations between those individual factors and the measures of interest. Family fixed effects, by controlling for all unmeasured and unmeasured family characteristics, further reduces those point estimates. Overall, most estimates of family fixed effects resemble OLS results, if not in size than at least in direction.

Figure 2.3: Comparing estimated coefficients of the associations between parental involvement measures with individual- and family-level variables using the full available sample (OLS) and the sibling sample (fixed effects).



Notes: GPA = Grade Point Average, PGS = educational attainment polygenic score, PPVT - Peabody Picture Vocabulary Test. OLS = Ordinary Least Squares, SES = socio-economic status, FE = fixed effects.
SES controls: parental education, income, father's absence.
Data source: the Add Health.

Table 2.5 summarises whether my previously stated hypotheses have been supported or refuted by the analysis.

Table 2.5: Summary of how the results correspond to the hypotheses.

		A	B	C	D
		School-based activities	Academic socialisation (mother)	Academic socialisation (father)	Parental control
H1.1	Education	+	+	+	.
H1.2	Income	+	+	+	.
H1.3	Father's absence	-	+	NA	-
H2.1	Age	-	-	-	-
H2.2	Female	-	.	.	-/+
H2.3	GPA	+	+	+	-
H2.4	EA3 PGS	+	+	+	.
H2.5	PPVT	+	+	+	-
H2.6	Self-efficacy	.	+	+	.
H2.7	Behaviour problems	-	-	-	-

Notes: Black "+" represents a supported hypothesis of positive association; black "-" represents a supported hypothesis of negative association; red "+" represents that instead of an expected negative association I find evidence for a positive association; red "-" represents that instead of an expected positive association I find evidence for a negative association; red "-/+" represents that the evidence for whether the association is negative or positive is mixed; red "." represents that I find no association between two measures.

GPA = Grade Point Average, EA3 PGS = educational attainment polygenic score, PPVT - Peabody Picture Vocabulary Test.

2.6.2 Explaining educational attainment

In the second part of the analysis I use the individual and family characteristics, explored in the previous section, to explain differences in educational attainment. The first stage of the analysis will model between-family differences in the data. Results from the OLS regressions can be seen in Table 2.6. Without any individual and family-level controls, all parental involvement measures are significant predictors of educational attainment: a 1 standard standard deviation increase in school-based activities, academic socialisation with mother or academic socialisation with father corresponds to an almost 0.6 year increase in educational attainment. These measures also explain 6-7% of the variance in years of education. However, the relationship between parental involvement and parental control is negative: a 1 standard deviation increase in parental control is associated with a 0.155 year reduction. The next step is to include individual and family-level factors to see whether any of those variables are actually confounding the relationship between parental involvement and educational attainment. After doing so we can see that the association estimates

are reduced by more than a half. Finally, after adding family-level characteristics to my models, the estimates of the associations between parenting and educational attainment are further reduced. Almost any other measure included in my models is now better at explaining the differences in years of education than the parental involvement measures. I also ran OLS models with the specifications in-between the small models (Models 1, 4, 7, 10) and the larger ones (Models 3, 6, 9, 12). The specifications curves can be seen in Figure 2.A.13 in the Appendix. The overall distribution of the estimates contains the preferred point estimate reported in Table 2.6. A more productive line of research would probably be testing the associations in different samples or using different operationalisations of key parenting variables.

Table 2.6: Estimated coefficients from the OLS regressions associating years of education with parental involvement and control variables.

	Explaining educational attainment											
	School-based activities			Academic socialisation						Parental control		
				Mother			Father					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	14.154***	11.911***	11.770***	14.255***	12.360***	12.195***	14.345***	12.432***	12.008***	14.218***	12.666***	12.464***
	(0.042)	(0.341)	(0.336)	(0.028)	(0.225)	(0.221)	(0.031)	(0.247)	(0.240)	(0.029)	(0.241)	(0.236)
Parental involvement	0.563***	0.262***	0.171***	0.544***	0.245***	0.198***	0.576***	0.250***	0.179***	-0.155***	-0.057*	-0.054*
	(0.041)	(0.036)	(0.036)	(0.028)	(0.024)	(0.024)	(0.032)	(0.027)	(0.027)	(0.030)	(0.026)	(0.025)
Age		0.119***	0.113***		0.094***	0.089***		0.091***	0.087***		0.072***	0.069***
		(0.022)	(0.021)		(0.014)	(0.014)		(0.016)	(0.015)		(0.015)	0.015
Female		0.322***	0.315***		0.281***	0.282***		0.318***	0.317***		0.293***	0.292***
		(0.072)	(0.071)		(0.048)	(0.047)		(0.053)	(0.051)		(0.047)	(0.046)
GPA		0.773***	0.715***		0.764***	0.689***		0.791***	0.706***		0.783***	0.703***
		(0.040)	(0.040)		(0.027)	(0.027)		(0.030)	(0.029)		(0.027)	(0.026)
EA3 PGS		0.266***	0.225***		0.285***	0.236***		0.277***	0.220***		0.292***	0.239***
		(0.037)	(0.037)		(0.024)	(0.024)		(0.027)	(0.026)		(0.024)	(0.024)
PPVT		0.515***	0.431***		0.552***	0.432***		0.540***	0.382***		0.574***	0.447***
		(0.046)	(0.046)		(0.031)	(0.031)		(0.035)	(0.034)		(0.031)	(0.031)
Self-efficacy		0.022	0.042		0.048*	0.060**		0.046†	0.066***		0.058*	0.065**
		(0.036)	(0.036)		(0.024)	(0.023)		(0.026)	(0.025)		(0.023)	(0.023)
Behaviour problems		-0.161***	-0.159***		-0.211***	-0.198***		-0.204***	-0.194***		-0.234***	-0.220***
		(0.040)	(0.039)		(0.027)	(0.023)		(0.030)	(0.029)		(0.026)	(0.026)
High mother's education			0.585***			0.644***			0.456***			0.671***
			(0.079)			(0.049)			(0.057)			(0.048)
Income			0.270***			0.259***			0.214***			0.271***
			(0.054)			(0.033)			(0.034)			(0.031)
Father absent			-0.049			-0.205***						-0.187**
			(0.087)			(0.058)						(0.057)
High father's education									0.572***			
									(0.059)			
N	2,495	2,495	2,495	5,636	5,636	5,636	4,646	4,646	4,646	5,882	5,882	5,882
R^2	0.071	0.356	0.385	0.061	0.369	0.408	0.067	0.373	0.422	0.004	0.354	0.396

Notes: GPA = Grade Point Average, EA3 PGS = Educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test. SEs in parentheses. †p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Models 1-3 use school-based activities as the parental involvement measure, Models 4-6 - academic socialisation with mother, Models 7-9 - academic socialisation with father, Models 10-12 - parental control.
Data source: the Add Health.

I also compare individual within families (Table 2.7). Although lower in sizes, the associations between individual characteristics and educational attainment follow the pattern observed in the OLS part of the analysis. The estimates for parental involvement, however, change significantly from the OLS part to the sibling fixed effects part. Instead of being positively associated with years of education, between siblings higher levels of school-based activities predict less educational attainment. The inclusion of individual-level controls, however, reduce the negative association by half from a 0.351 year decrease for a 1 standard deviation increase in school-based activities to a 0.210 year decrease. Nevertheless, I should note that these point estimates are not statistically significant. Academic socialisation with mother displays a positive association with years of education and the association is only moderately reduced after including individual characteristics in contrast to academic socialisation with father, where the association decreases from a 0.206 year to a 0.030 year indicating a high level of confounding by these characteristics. Finally, parental control within families in adolescence is not associated with later educational outcomes.

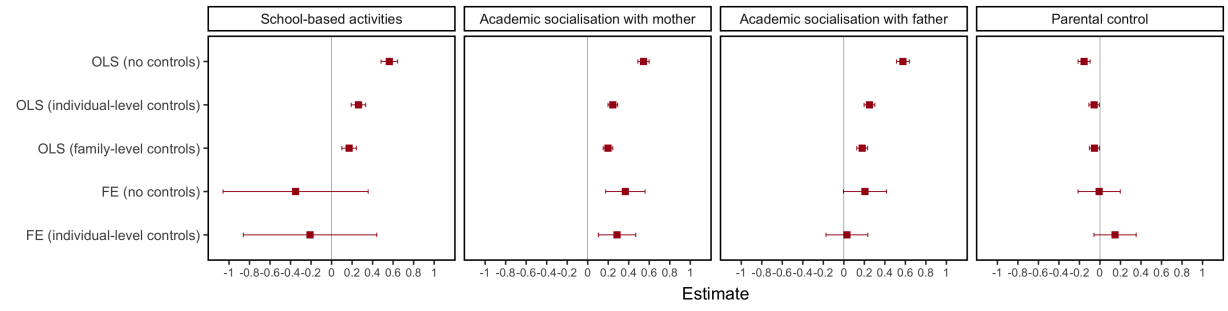
Table 2.7: Estimated coefficients from the OLS regressions associating years of education with parental involvement and control variables using sibling fixed effects.

	Explaining educational attainment							
	School-based activities		Academic socialisation				Parental control	
			Mother		Father			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	16.610***	11.011***	16.262***	13.199***	13.740***	12.147***	16.016***	12.315***
	(1.786)	(2.317)	(1.601)	(1.707)	(1.596)	(2.679)	(1.642)	(1.767)
Parental involvement	-0.351	-0.210	0.367***	0.286**	0.206†	0.030	-0.008	0.147
	(0.361)	(0.332)	(0.098)	(0.093)	(0.107)	(0.104)	(0.105)	(0.105)
Age		0.197*		0.124*		0.135*		0.152**
		(0.091)		(0.053)		(0.058)		(0.056)
Female		0.457		0.367*		0.439*		0.389*
		(0.284)		(0.169)		(0.184)		(0.168)
GPA		0.403*		0.480***		0.512***		0.486***
		(0.159)		(0.102)		(0.115)		(0.102)
EA3 PGS		0.295		0.230*		0.233		0.219*
		(0.205)		(0.111)		(0.119)		(0.111)
PPVT		0.529*		0.255†		0.216		0.347*
		(0.250)		(0.145)		(0.161)		(0.142)
Self-efficacy		0.118		0.075		0.174†		0.103
		(0.142)		(0.080)		(0.089)		(0.079)
Behaviour problems		-0.183		-0.203*		-0.136		-0.184†
		(0.174)		(0.099)		(0.106)		(0.098)
N	508	508	1,088	1,088	942	942	1,123	1,123

Notes: GPA = Grade Point Average, EA3 PGS = Educational attainment polygenic score, PPVT = Peabody Picture Vocabulary Test. SEs in parentheses. †p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Models 1-2 use school-based activities as the parental involvement measure, Models 3-4 - academic socialisation with mother, Models 5-6 — academic socialisation with father, Models 7-8 — parental control.
Data source: the Add Health.

Figure 2.4 shows how different the effect estimates are for each parental involvement measures in different models. For school-based activities in the OLS model the positive association with educational attainment decreases as more controls are added but it stays positive. In the family fixed effects model the association turns negative and is reduced when individual-level controls are included. Academic socialisation with mother overall has a positive relationship with years of education but more controls, including unmeasured family characteristics, lead to a significant reduction in the association size. It is also important to mention that the OLS model which controls for both individual and family characteristics produces a very similar estimate to the one from the family fixed effects model which conditions on sibling characteristics. The pattern is also similar to academic socialisation with father. What is notable in this case is how significantly the inclusion of controls reduces the association size. Parental control estimates move from being slightly negative to close to zero. All in all, family fixed effects produce the lowest association estimates. The upper bounds (or lower bounds in the case of parental control) of their confidence intervals roughly correspond to the estimates produced with OLS controlling for individual characteristics.

Figure 2.4: Comparing estimated coefficients of the associations between years of education and parental involvement measures using the full available sample (OLS) and the sibling sample (fixed effects).



Notes: OLS = Ordinary Least Squares, SES = socio-economic status, FE = fixed effects.
Individual-level controls: age, gender, GPA, the educational attainment polygenic score (EA3 PGS), Peabody Picture Vocabulary Test (PPVT), self-efficacy, behaviour problems.
Family-level controls: parental education, income, father's absence.
Data source: the Add Health.

2.7 Conclusion/Discussion

This paper aimed at understanding individual and family factors which explain the intensity of parental involvement and whether parental involvement truly matters for child's educational outcomes. Both educational policies and the media have been championing a very involved type of parenting without knowing the extent to which it makes a difference. While there is some experimental evidence on the importance of parental involvement in early childhood, relatively little is known about parenting during another important developmental period — adolescence. Although my study does not leverage any experimental design, it contributes to the existing literature by carefully investigating often disregarded characteristics which might confound the relationship between parental involvement and educational attainment. Such characteristics are found both at individual (age, gender, academic performance, cognitive ability, genetic propensity to education, self-efficacy, and behaviour) and family levels (parental education, household income, father's absence). I also explore the multidimensionality of parental involvement by distinguishing between its three theoretically inspired facets: school-based activities, academic socialisation, and parental control.

First, I find that parental involvement, instead of entirely being a choice as parenting books often advertise, is influenced by structural, cultural, and other social forces. In other words, the agency in the context of parenting is limited. Parental education and its accompanying cultural capital is one of the strongest explanatory factors of school-based activities and academic socialisation: the school-based activities' score is higher by a 0.392 standard deviation for mothers with more education than a high school diploma compared to those with high school diploma or less education. Income also makes a difference in this context, even net of other family factors, probably by relaxing time constraints. My analysis also shows that another structural impediment to parental involvement is father's absence, which is very prevalent in the US. Being a single parent reduces one's ability to regularly attend school events or effectively monitor and control child's behaviour. I also find that parental involvement is not only driven by cultural logic and structural realities but is also responsive to child's characteristics. To put it differently, any universal formula of

"correct" parenting needs to be adjusted and renegotiated to meet the challenges posed by child's personality and habits. Parents seem to become less involved as their child gets older and if the child is female, controlling for other child's characteristics. Academic performance, measured as GPA, and cognitive ability, measured as the Peabody Picture Vocabulary Test, have independently strong positive associations with school-based activities and academic socialisation and negative associations with parental control. Parents of high achieving children are likely to invest more to cultivate their talents in order to realise their full potential and reap later benefits. On the other hand, these strong associations might be contaminated by previous parental involvement and the direction of causality might go in the opposite direction or even in a self-perpetuating cycle: high levels of parental involvement earlier in life might have helped with child's performance which, in turn, elicited more parental involvement. The genetic propensity to education, measured with the educational attainment polygenic score, however, is not related to parental involvement, once other variables are controlled for. The associations between parental involvement measures and the polygenic score are significantly reduced after conditioning on family-level factors, demonstrating how higher-SES children are likely to have higher polygenic scores as well. My analysis also shows that one of the non-cognitive skills — self-efficacy — is also connected with academic socialisation. This mechanism, however, can again be self-reinforcing. Confident and resourceful children can evoke more attention from their parents in the form of communication about school work, grades, and future plans. At the same time, children whose parents have previously encouraged them through effective socialisation might become more self-efficacious. Behaviour problems are negatively associated with all my dimensions of parental involvement. Although behaviour problems, in theory, might provide parents with incentives to be more involved, negative associations between the problems and parental involvement might reflect not the current reaction to those problems but rather a serial correlation of parental involvement. For example, low levels of parental control in earlier age might have allowed for behaviour problems to occur in the first place. In addition to carefully controlling for family-level factors in between-family models, I also compare individuals within families using sibling fixed effects to control for unmeasured family characteristics. Although generally smaller in association sizes, estimates from the sibling fixed effects models follow the pattern observed in the OLS models. Nevertheless, some surprises also emerge. For instance, within families parents seem to compensate for academic performance differences by investing more in

school-based activities of a child who is performing worse. A study of intra-household allocation of parental involvement in adolescence using a longitudinal design would be a productive research direction for the future.

Second, I find that parental involvement in adolescence, contrary to what is suggested by intensive parenting proponents, has limited instrumental value for educational attainment. After controlling for the individual and family-level characteristics explored in the first part of the analysis, parental involvement has limited explanatory power for one's years of education compared to almost any other factor included in my models. Before conditioning on other variables, all parental involvement dimensions, except for parental control which has a very limited explanatory power, can explain a sizeable 5% of the variance in educational attainment. However, the R^2 of this size is rather small compared to the R^2 of around 36% in the models which include all other individual and family characteristics. After including other variables, which as my results show confound the relationship between parental involvement and educational attainment, the estimated association sizes of my four parenting measures are reduced by almost two thirds: for example, the estimated association decreases from a 0.486 year increase to a 0.188 year increase in educational attainment for a 1 standard deviation increase in academic socialisation with mother from the model with no controls to the one with all controls included. It is likely that an inclusion of many other child and family characteristics, not available in this dataset, would reduce the estimates even further, making the results less biased. To illustrate, in my models even after including many controls parental control is negative associated with educational attainment. It is hard to justify why more vigilant parenting would cause lower levels of educational attainment. One potential explanation is of children becoming more rebellious and subversive the more their parents try to control them. A more plausible explanation, however, is that I haven't accounted for some existing child's characteristics which both drive more parental monitoring and are related to worse academic outcomes.

Several limitations of this paper should also be acknowledged. I could not take into account every individual and family characteristic because of data limitations. Having a richer set of questions on cognitive and non-cognitive skills, for example, would have been helpful. Therefore, my attempt

to control for as many potential confounders in the association between parental involvement and educational attainment is limited at best. My conclusions should be seen as a critical evaluation of the intensive parenting doctrine and an inspiration for future research rather than definitive causal evidence. Although I attempted to control for unmeasured family factors using sibling fixed effects, the results from this set of analyses should be evaluated carefully. In spite of the Add Health dataset being well-known for its reasonably sized sibling sample, missing values in my dependent variables and the requirement for the genetic analysis that siblings were either fraternal or dizygotic twins reduced the sibling pool substantially. Consequently, the confidence intervals on the estimates produced using sibling fixed effects are very wide and, hence, results are often hard to interpret. Furthermore, as already stressed in the description of the results, reverse causality between the variables of interest is possible. Overcoming such issues to truly establish a causal link between parental involvement in adolescence and educational outcomes would require richer observational data, evidence from natural experiments or conducting an actual experiment.

Having a more extensive panel data spanning both early childhood and adolescence would be an improvement. Very detailed year-by-year measurements of individual (grades, cognitive and non-cognitive skills, behaviour on children's side; changes in jobs, education, relationships and friends, residence location, financial situation on parents' side) and contextual factors (such as changes in schooling quality) would potentially allow to identify when changes in parenting occur and the factors these changes are driven by. If coupled with sibling data, this approach could help explore exact times when within-family parental involvement diverges and to have greater confidence in the determinants of such divergence. Subsequent changes in children's educational outcomes, given changes in parenting, could then be studied more directly. However, I would still not be able to control for all time-varying child and family characteristics, thus leading to biased estimates. Another statistical approach of dealing with endogeneity would be to identify a natural experiment which would be associated with the treatment (parenting) and with the outcome (education) only through the treatment. Unfortunately, good instrumental variables are extremely hard to find in the context of parenting. There could be some random variation in school-based activities that parents are encouraged or required to attend driven by appointment of new school principals or school district superintendants because of retirement of incumbents. However, such naturally occurring

changes in school administrators would most likely have effect not only on how much school exposure parents receive but also on school quality more generally, thus violating the assumption of exclusion restriction.

Given the limitations of these statistical tools, the most practical and promising approach would be to design a randomised control trial (RCT) where parents would be experimentally assigned different intensities or types of parental involvement. Most existing interventions or RCTs, however, focus on "replacing" parents with, for instance, mentors or child development specialist, rather than equipping parents with the skills and tools to parent more effectively. Several notable exceptions exist in the case of prenatal and early-childhood development: the Nurse-Family Partnership visits young pregnant women and teaches them about best health practices and parenting (Robling et al., 2016); Baby College in the Harlem Children's Zone teaches young parents how to reason and effectively discipline children; Parents and Children Together (PACT) intervention sent out text reminders and information on the benefits of reading to children (Mayer et al., 2015); a brief parenting and life skills intervention taught teenage mothers about medical care and self-esteem (Cox et al., 2019); Triple P Online Brief organised a low-intensity parenting class for parents whose children displayed signs of early disruptive behaviour (Baker et al., 2017). Even if similar RCTs exist for parents of adolescents, the targeted outcomes are not academic but rather those relating to substance abuse (Allen et al., 2016), depression (Poole et al., 2018), or family conflict (Dittman et al., 2020). An RCT designed to address the aspects of parenting discussed in this paper — school-based activities, academic socialisation, and parental control — could take a form of parent-support groups, communication with teachers through other means than face-to-face meetings, or, following Mayer et al. (2015), automated reminders or pieces of advice sent via text messages. Several concerns about such RCTs should also be mentioned. First, RCTs in general can be criticised as not entirely ethical since they potentially withhold benefits from a control group. Nevertheless, it is debatable whether it is unethical to randomise when the effectiveness of a program is truly unknown (as is the case for parenting interventions in adolescence). Moreover, instead of using a control group with no treatment, all groups could receive a treatment but with varying intensity. Some critics would criticise such interventions in adolescence on the grounds of equity-efficiency trade-off as less efficient than interventions in early childhood and less than those directly

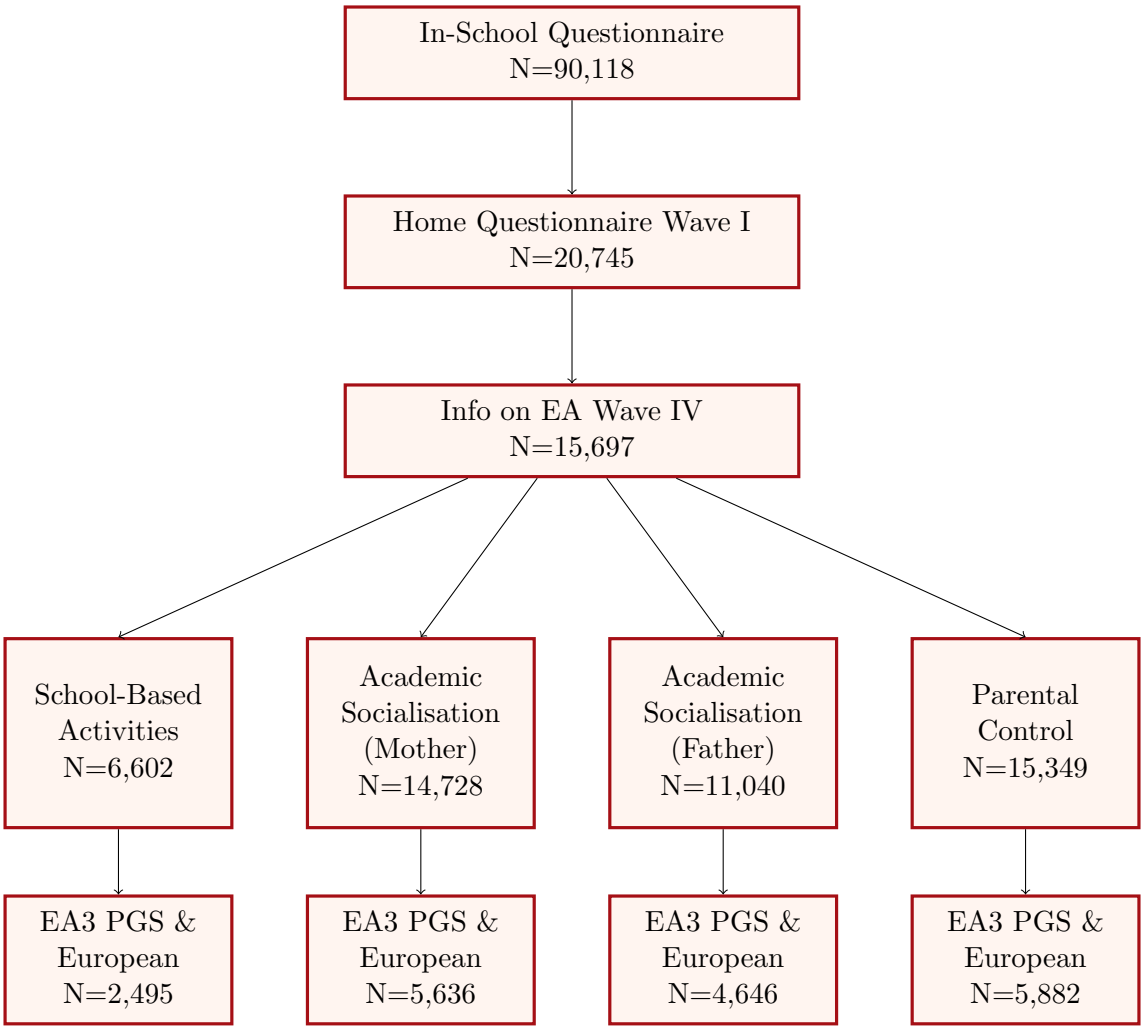
providing educational support rather than targeting parenting (Heckman, 2013). Nevertheless, I believe it is crucial to study adolescence even if it will not educationally benefit everyone: parental support and encouragement as well as parent-teacher dialogue might be essential for the least academically successful and those on the edge of dropping out. Finally, critics could also emphasise the fact that such interventions borrow notions of parenting from the dominant middle-class culture (Lareau, 2003) and stigmatise a large proportion of American parents as unfit to raise their children (West, 2007). However, a transparent relationship between parents and teachers characterised by an effective exchange of information, parental awareness of what is happening in a child's academic and personal life as well as mutual communication of future intentions is valued across cultures.

The operationalisation of the variables also requires some caveats. Although an improvement to what has been previously used in the literature, my dimensions are still rather crude and do not encompass the richness of adolescent-parent relationships. For instance, academic socialisation is much more than just being disappointed if one's child does not go to college. It is about implicit and explicit day-to-day role-modelling and inculcation of values which the disappointment measure only proxies for. Furthermore, only the items for the measure of school-based activities are directly reported by parents — other dimensions are constructed using answers provided by adolescents, who might not give the most objective account of their parents' involvement. Also, while in this paper I focus on parental involvement mostly in terms of time investments, other dimensions of parenting in adolescence, if such data were available, would be useful to explore. For example, information on monetary investments in extracurricular activities, private tutors, or college admission preparations could illuminate how parents delegate their involvement or further reinforce it with external help. Finally, while here I evaluate the efficacy of parental involvement through an entirely educational lens, it is worth investigating whether parental involvement matters more for other important child outcomes such as social adjustment, well-being, or later marital success. Also, the outcome of years of education might obscure the importance of parental involvement for the type of education one receives, i.e. prestige of university or college major. The crude measure of educational attainment captures the quantity of education rather than the quality of it, the latter measure potentially being highly dependent on parental inputs.

Given that most people are either givers or recipients of parenting and that the definition of effective parenting has become narrower over the years, the importance of studying this topic cannot be exaggerated. Although the current consensus is that higher levels of involvement are always better, my results do not support this view. Parental involvement is shaped by one's sociocultural context and is a result of a parent-child dialogue. Once these heterogeneous contexts are taken into account, parental involvement does not seem to make a big difference for child's eventual educational outcomes. In other words, a middle-class formula of intensive involvement, supervision, and concerted-cultivation might not be successfully disentangled from all other middle-class factors and might not translate into the same academic results in a lower-SES milieu. To know whether parental involvement in adolescence has a true causal effect for educational outcomes, further research, hopefully exploiting an experimental design, will be necessary.

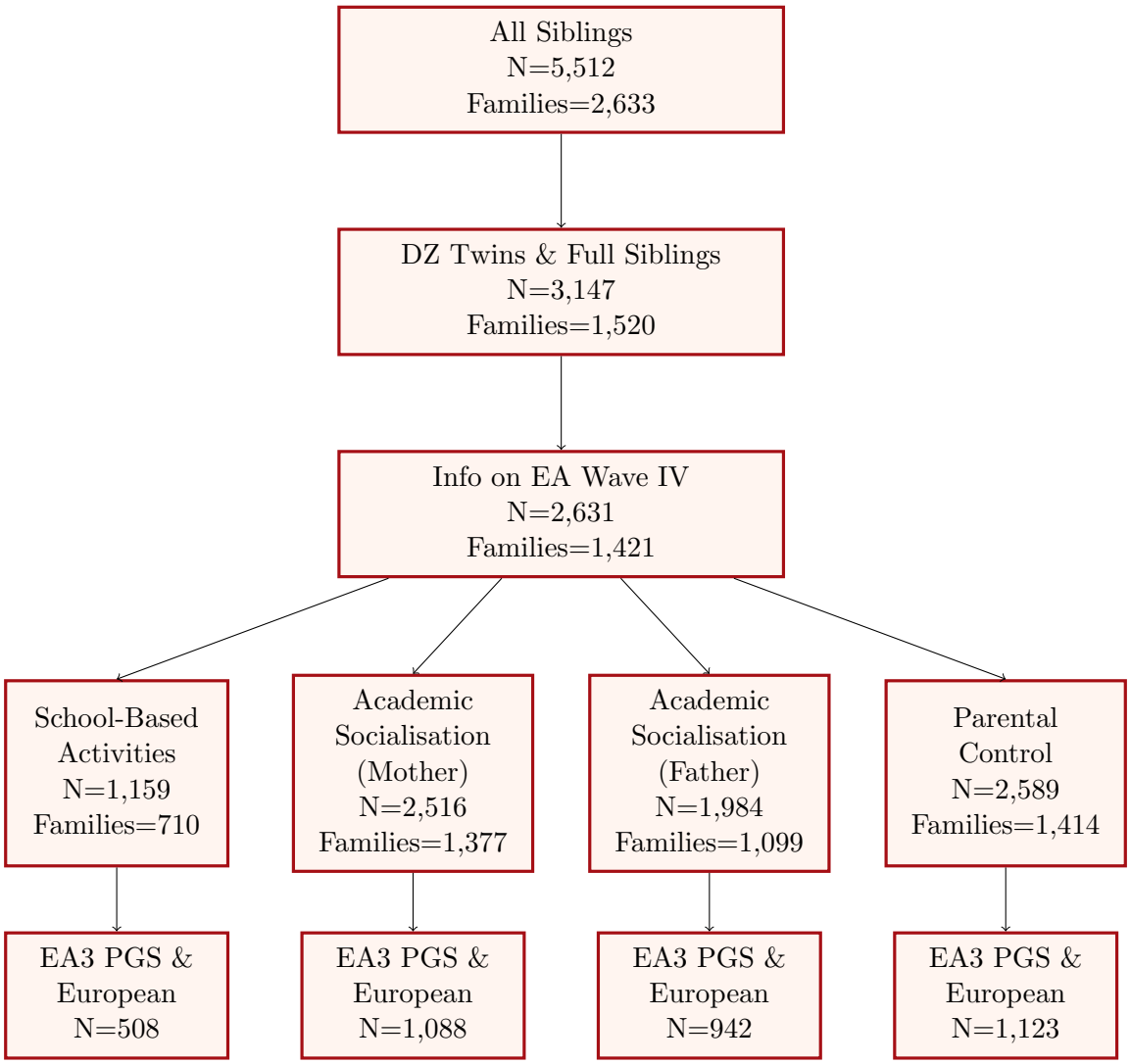
2.A Appendix

Figure 2.A.1: Reduction of the full sample in the Add Health.



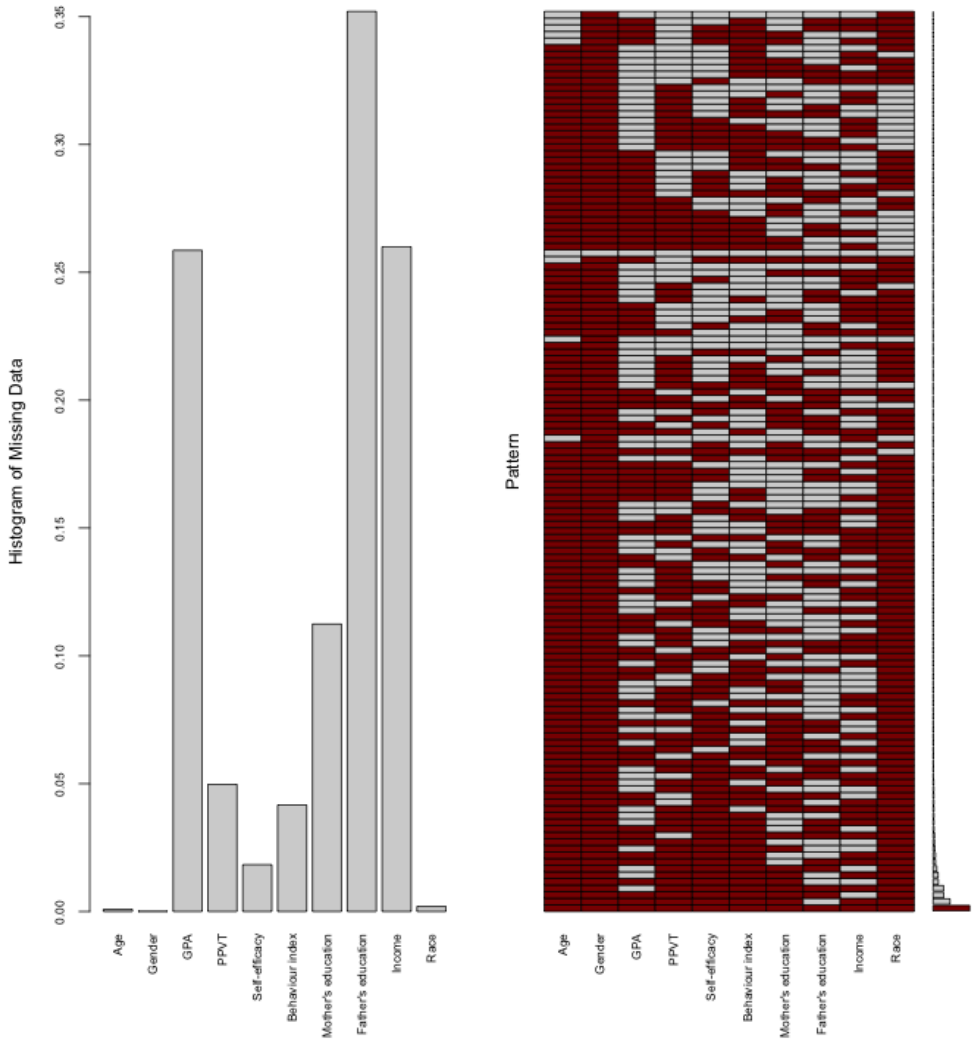
Notes: EA = educational attainment. EA3 PGS = educational attainment polygenic score.

Figure 2.A.2: Reduction of the sibling sample in the Add Health.



Notes: EA = educational attainment. EA3 PGS = educational attainment polygenic score.

Figure 2.A.3: Pattern of missing data in the Add Health.



Notes: On the left, proportions of missing data in each variable which were later imputed are displayed. On the right, combinations of different variables and their missingness (grey) or non-missingness (dark green) are shown.

Theoretical parenting dimensions	Composite measures	Questions in the Add Health
School-based activities		
	Parent-teacher association	Are you a member of parent-teacher associations?
	Talking with teachers	Have you talked with your child's teachers about (his/her) school work this school year, either informally or in a regularly scheduled parent-teacher conference?
	Volunteering at school	During this school year, have you participated in school fund-raising or done volunteer work for your child's school such as supervising lunch, chaperoning a field trip, etc.?
Academic socialisation		
	Disappointment if no college	How disappointed would your parent be if you did not graduate from college?
	Disappointment if no high school	How disappointed would your parent be if you did not graduate from high school?
	Talking about grades	In the past week, have you talked about your school work or grades with your parent?
	Talking about other school things	In the past week, have you talked about other things you're doing in school with your parent?
Parental control		
	Time bed weeknights	Do your parents let you make your own decisions about what time you go to bed on weeknights?
	Time home on weekends	Do your parents let you make your own decisions about the time you must be home on weekend nights?
	Approval of friends	Do your parents let you make your own decisions about the people you hang around with?
	TV (time)	Do your parents let you make your own decisions about how much television you watch?
	TV (programmes)	Do your parents let you make your own decisions about which television programs you watch?

Table 2.A.1: Measures of parental involvement.

	Age	Female	PGS	GPA	PVT	Self-efficacy	Behaviour problems
Age	1.000						
	(0.000)						
Female	-0.031	1.000					
	(0.018)	(0.000)					
EA3 PGS	0.005	-0.018	1.000				
	(0.024)	(0.024)	(0.000)				
GPA	-0.036	0.155	0.224	1.000			
	(0.020)	(0.020)	(0.026)	(0.000)			
PVT	-0.025	-0.040	0.207	0.293	1.000		
	(0.018)	(0.036)	(0.023)	(0.019)	(0.000)		
Self-efficacy	0.081	0.026	0.016	0.188	0.105	1.000	
	(0.018)	(0.044)	(0.024)	(0.020)	(0.018)	(0.000)	
Behaviour problems	0.159	-0.161	-0.058	-0.340	-0.024	-0.206	1.000
	(0.018)	(0.018)	(0.025)	(0.018)	(0.019)	(0.017)	(0.000)

Table 2.A.2: Correlations between individual-level explanatory factors.

Dimensions	Items	Full sample		Sibling sample	
		Proportion (%)	N	Proportion (%)	N
School-based activities					
	Parent-teacher association	0.32	17,497	0.32	2,719
	Talking with teachers	0.70	8,783	0.70	1,418
	Volunteering at school	0.32	8,773	0.36	1,416
Academic socialisation					
Mother					
	Disappointment: no college	0.69	19,372	0.68	2,992
	Disappointment: no high school	0.93	19,380	0.92	2,987
	Communication: grades	0.63	19,459	0.64	2,998
	Communication: school work	0.54	19,459	0.55	2,998
Father					
	Disappointment: no college	0.70	14,386	0.69	2,353
	Disappointment: no high school	0.92	14,378	0.92	2,351
	Communication: grades	0.53	14,429	0.54	2,355
	Communication: school work	0.45	14,429	0.48	2,355
Parental control					
	Time home on weekends	0.65	20,260	0.68	3,089
	Time bed weeknights	0.34	20,286	0.35	3,092
	Choice of friends	0.16	20,271	0.16	3,087
	TV time	0.18	20,285	0.20	3,091
	TV programmes	0.22	20,285	0.25	3,091

Table 2.A.3: Descriptive statistics of the items of parental involvement in the Add Health.

Table 2.A.4: Correlations between school-based activities in the Add Health.

	PTA	Volunteer	Talk with teachers
PTA	1.000		
	(0.000)		
Volunteer	0.456	1.000	
	(0.023)	(0.000)	
Talk with teachers	0.248	0.317	1.000
	(0.026)	(0.025)	(0.000)

Notes: PTA = Parent-teacher association.

Table 2.A.5: Correlations between items of academic socialisation with mother in the Add Health.

	Disappointed if no college	Disappointed if no HS	Talk about grades	Talk about school work
Disappointed if no college	1.000			
	(0.000)			
Disappointed if no HS	0.389	1.000		
	(0.016)	(0.000)		
Talk about grades	0.176	0.147	1.000	
	(0.018)	(0.018)	(0.000)	
Talk about school work	0.155	0.125	0.534	1.000
	(0.018)	(0.018)	(0.014)	(0.000)

Table 2.A.6: Correlations between items of academic socialisation with father in the Add Health.

	Disappointed if no college	Disappointed if no HS	Talk about grades	Talk about school work
Disappointed if no college	1.000			
	(0.000)			
Disappointed if no HS	0.403	1.000		
	(0.017)	(0.000)		
Talk about grades	0.195	0.174	1.000	
	(0.020)	(0.020)	(0.000)	
Talk about school work	0.169	0.141	0.578	1.000
	(0.020)	(0.020)	(0.031)	(0.000)

Notes: HS = high school.

Table 2.A.7: Correlations between items of parental control in the Add Health.

	Time weekends	Time weeknights	Friends	TV time	TV programmes
Time weekends	1.000				
	(0.000)				
Time weeknights	0.267	1.000			
	(0.017)	(0.000)			
Friends	0.235	0.194	1.000		
	(0.018)	(0.018)	(0.000)		
TV time	0.218	0.361	0.257	1.000	
	(0.018)	(0.017)	(0.017)	(0.000)	
TV programmes	0.233	0.338	0.242	0.452	1.000
	(0.017)	(0.017)	(0.017)	(0.015)	(0.000)

Table 2.A.8: Correlations between composite measures of parental involvement constructing using the Add Health.

	School-based activities	Socialisation with mother	Socialisation with father	Parental control
School-based activities	1.000			
	(0.000)			
Socialisation with mother	0.105	1.000		
	(0.027)	(0.000)		
Socialisation with father	0.108	0.669	1.000	
	(0.031)	(0.011)	(0.000)	
Parental control	-0.003	0.055	0.055	1.000
	(0.027)	(0.018)	(0.021)	(0.000)

Table 2.A.9: Sibling correlations in the Add Health sibling sample.

	Age	Female	PGS	GPA	PVT	Self-efficacy	Behaviour problems
Full sample	0.213	0.130	0.589	0.378	0.565	0.103	0.334
	(0.023)	(0.024)	(0.024)	(0.054)	(0.017)	(0.024)	(0.022)
High mother's education	0.190	0.103	0.649	0.382	0.510	0.117	0.306
	(0.035)	(0.036)	(0.032)	(0.079)	(0.056)	(0.037)	(0.034)
Low mother's education	0.226	0.177	0.534	0.339	0.562	0.083	0.376
	(0.033)	(0.034)	(0.037)	(0.041)	(0.024)	(0.036)	(0.031)

Family SES characteristic	School-based activities	Academic socialisation		Parental control
		Mother	Father	
Full sample	0.84	0.26	0.25	0.26
	(0.01)	(0.02)	(0.03)	(0.02)
Mother's education				
Above HS	0.85	0.25	0.26	0.26
	(0.02)	(0.04)	(0.04)	(0.03)
HS or below	0.79	0.23	0.214	0.30
	(0.02)	(0.03)	(0.04)	(0.03)
Father's education				
Above HS	0.89	0.23	0.24	0.25
	(0.01)	(0.04)	(0.04)	(0.04)
HS or below	0.76	0.21	0.19	0.25
	(0.03)	(0.04)	(0.04)	(0.04)
Household income				
Above median	0.84	0.21	0.23	0.21
	(0.02)	(0.04)	(0.04)	(0.04)
Below median	0.80	0.26	0.22	0.29
	(0.02)	(0.04)	(0.05)	(0.04)
Father's presence				
Present	0.85	0.24	0.25	0.24
	(0.01)	(0.03)	(0.03)	(0.03)
Absent	0.78	0.32		0.33
	(0.03)	(0.05)		(0.04)

Table 2.A.10: Interclass correlations (ICC) of parental involvement measures between siblings by socio-economic status in the Add Health.

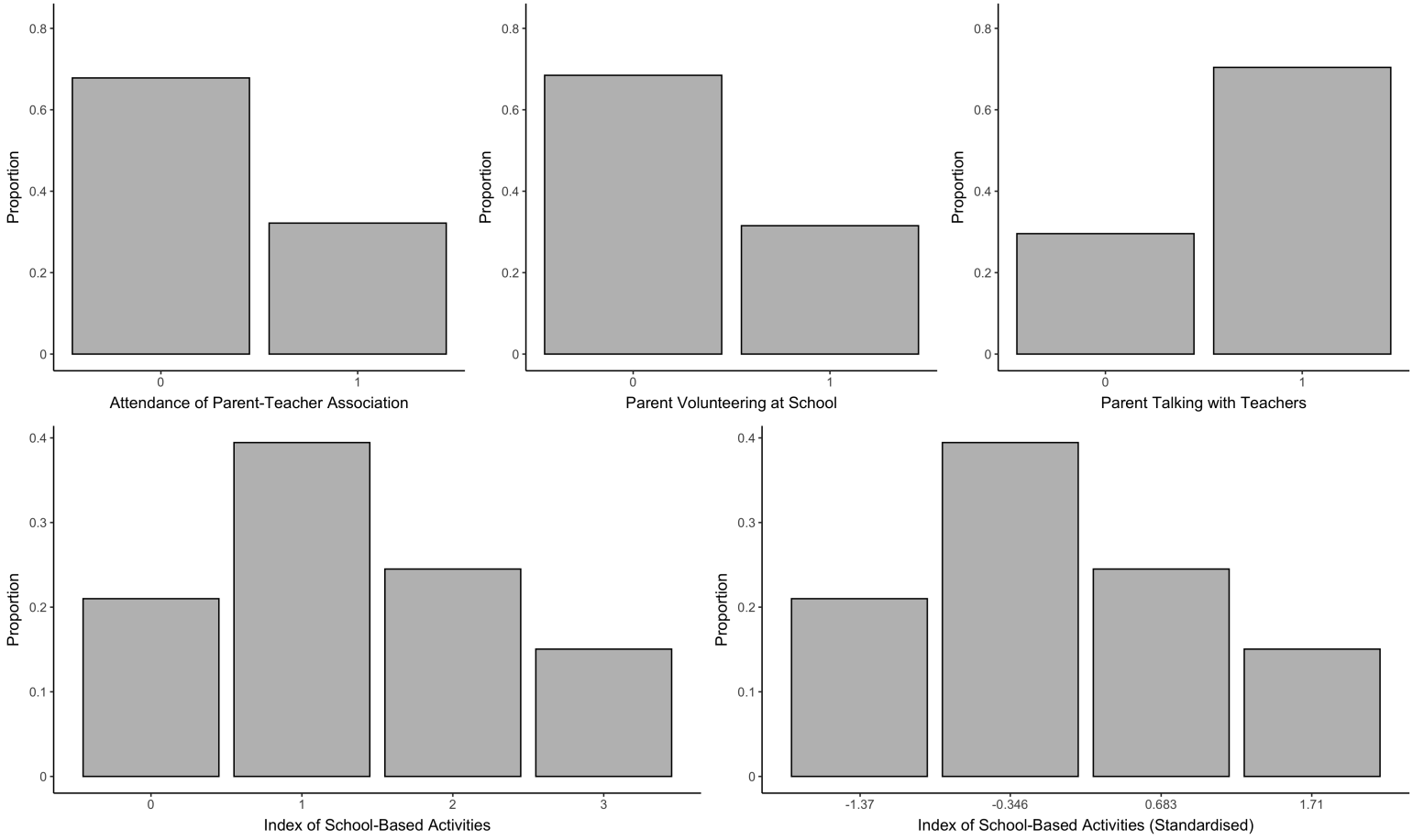


Figure 2.A.4: Distributions of items of school-based activities in the Add Health.

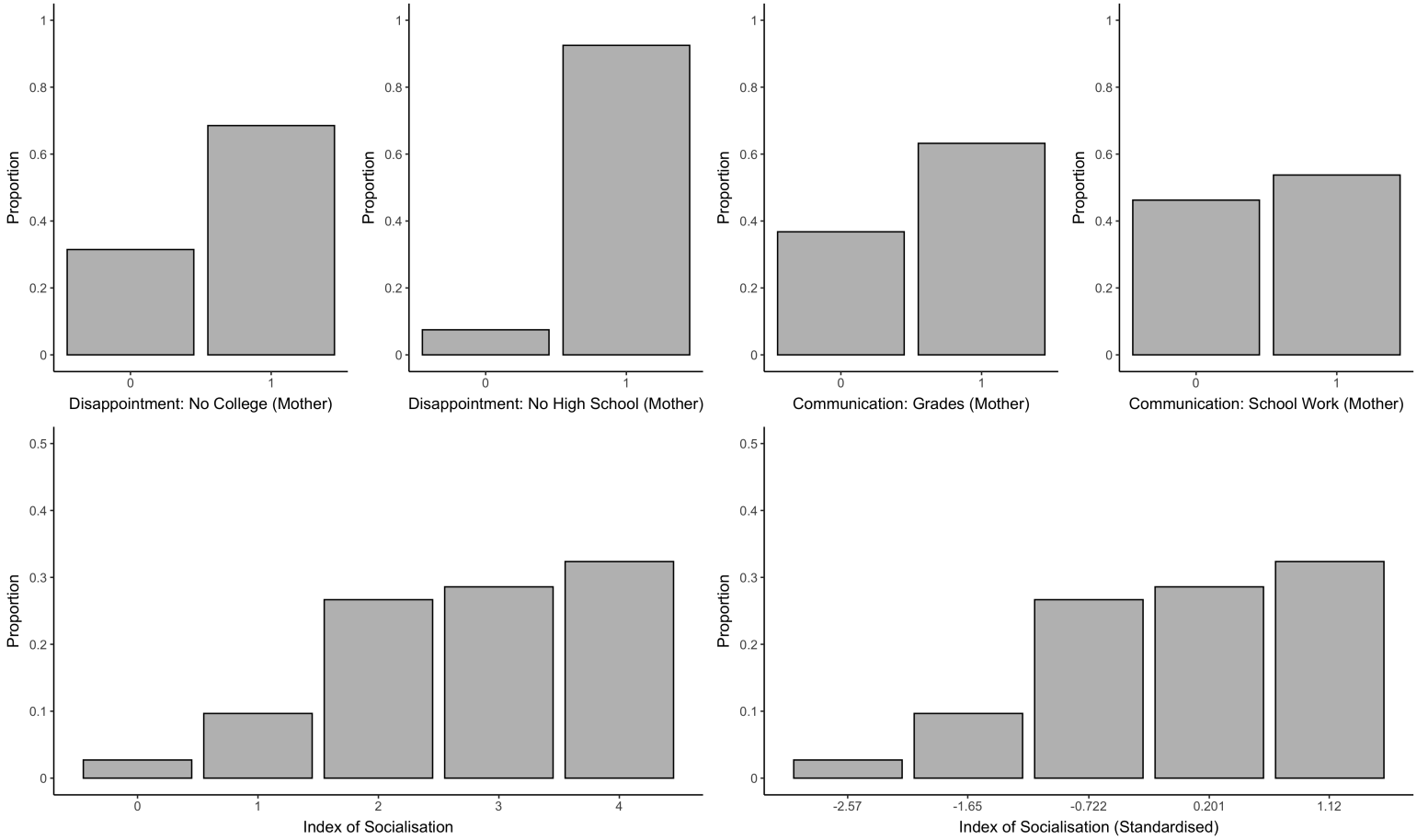


Figure 2.A.5: Distributions of items of academic socialisation with mother in the Add Health.

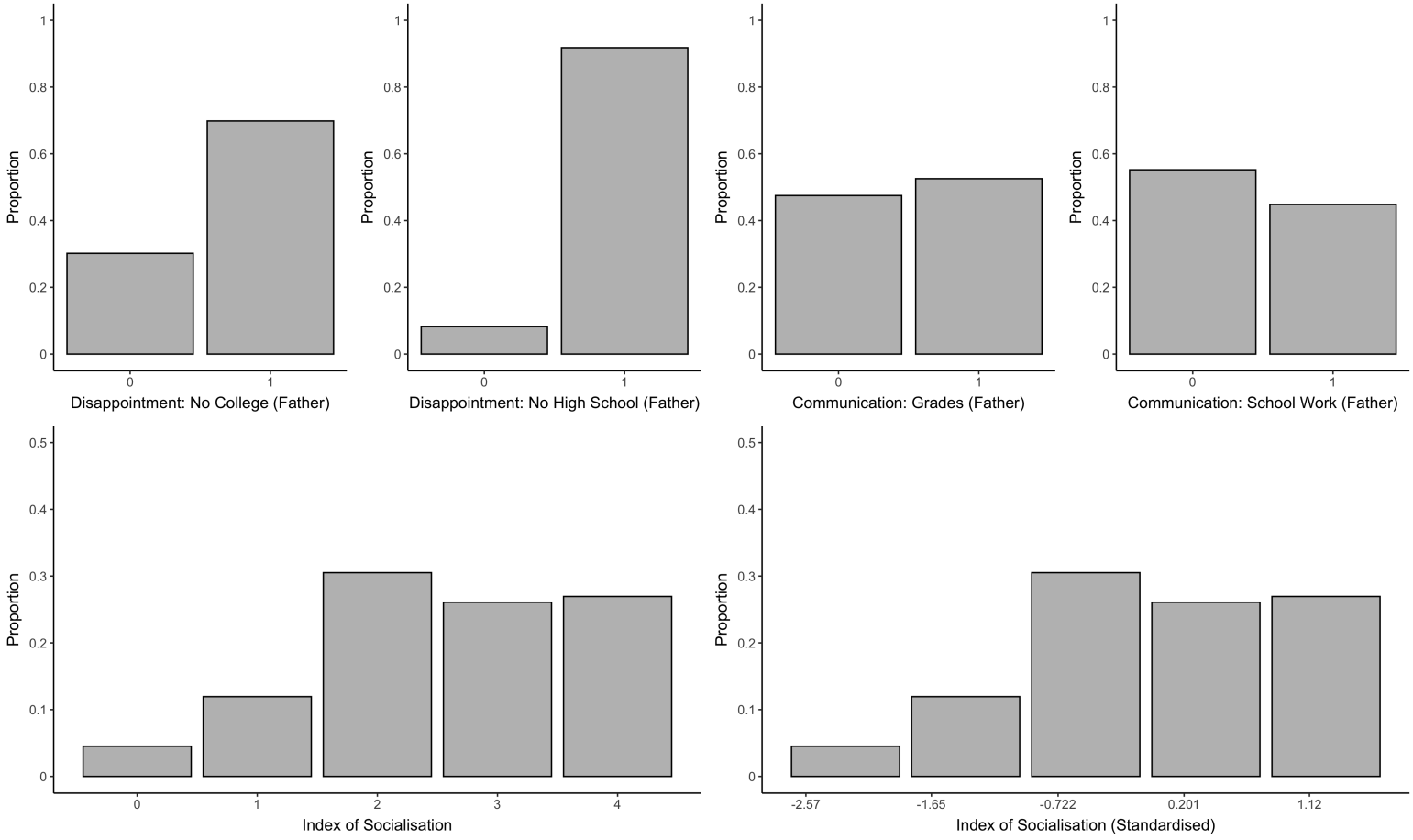


Figure 2.A.6: Distributions of items of academic socialisation with father in the Add Health.

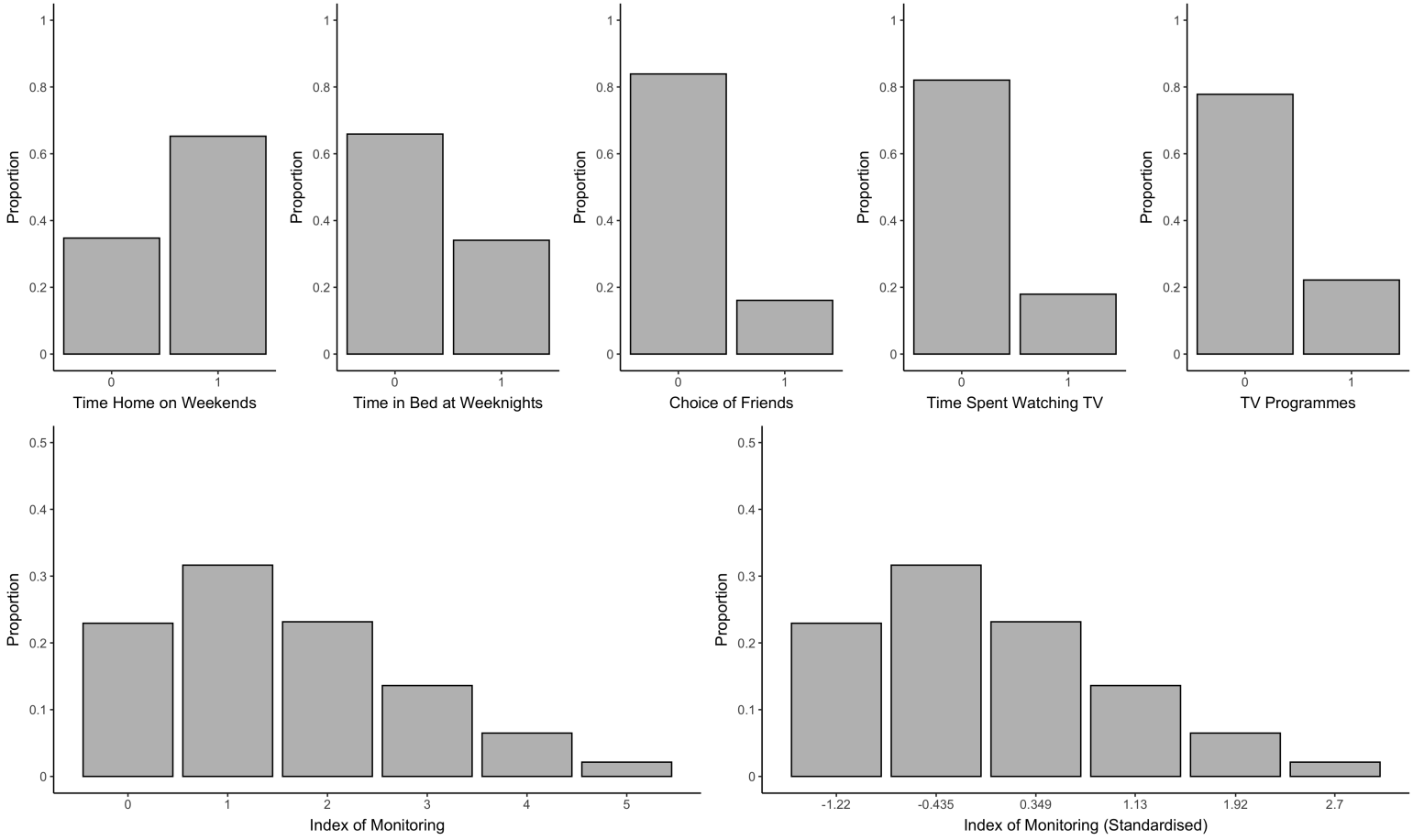


Figure 2.A.7: Distributions of items of parental control in the Add Health.

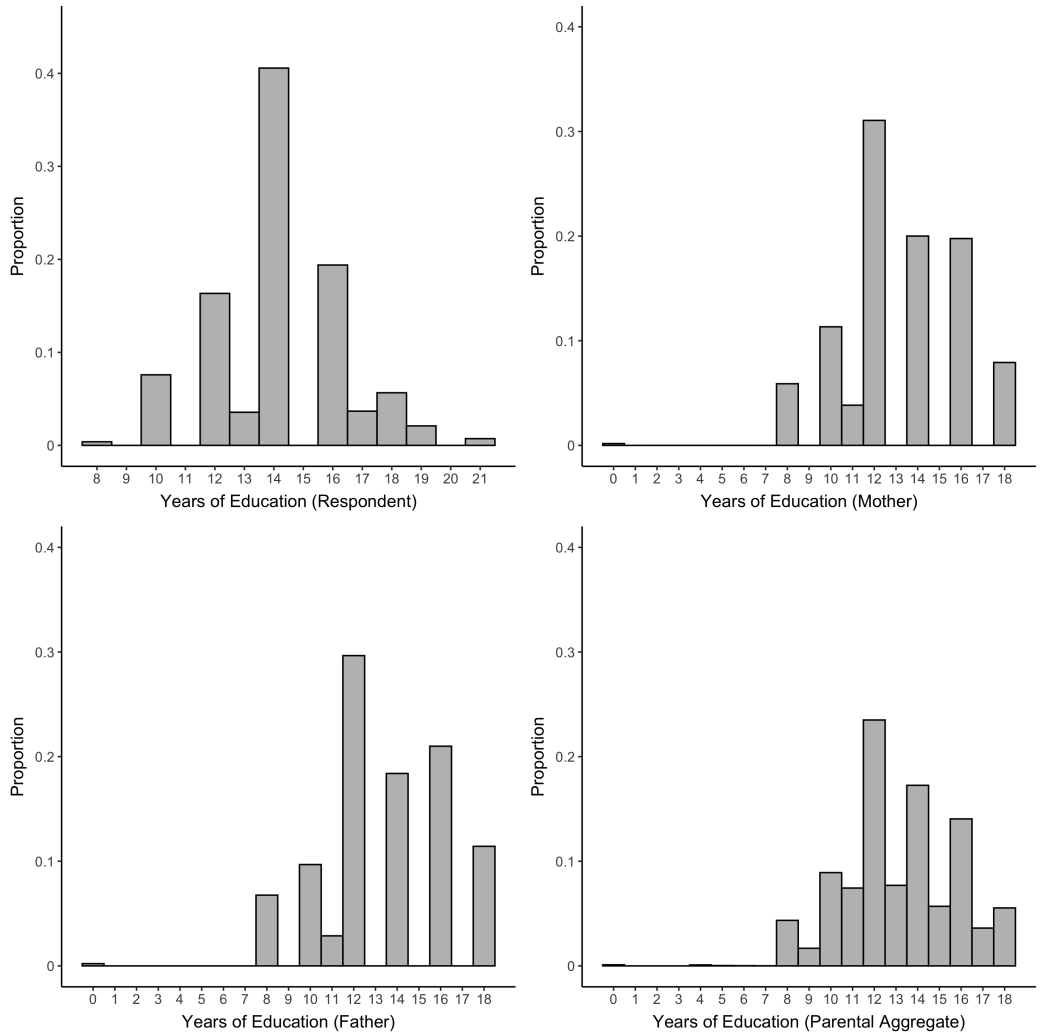


Figure 2.A.8: Distribution of years of education (for respondents, for mothers, and for fathers) in the Add Health.

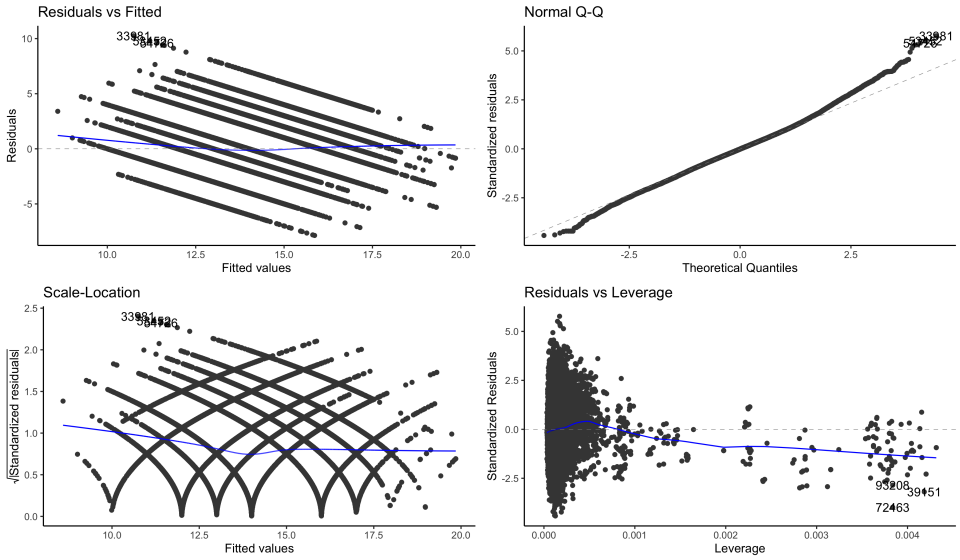


Figure 2.A.9: Diagnostic plots for the model associating years of education with school-based activities using the Add Health.

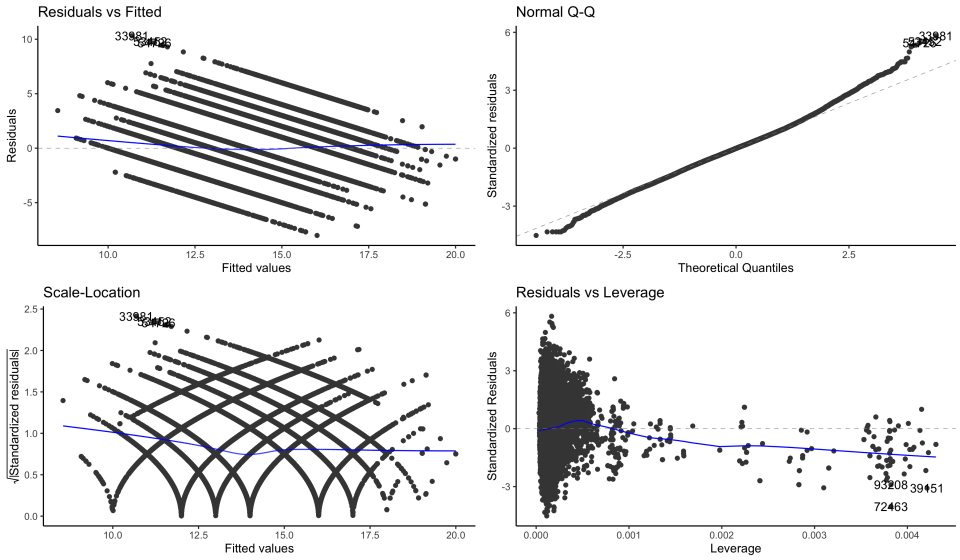


Figure 2.A.10: Diagnostic plots for the model associating years of education with academic socialisation with mother using the Add Health.

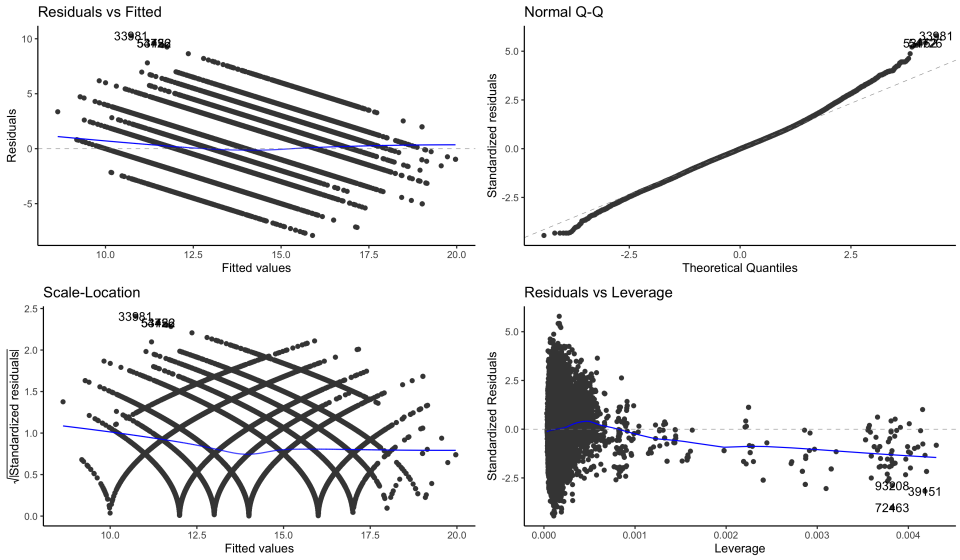


Figure 2.A.11: Diagnostic plots for the model associating years of education with academic socialisation with father using the Add Health.

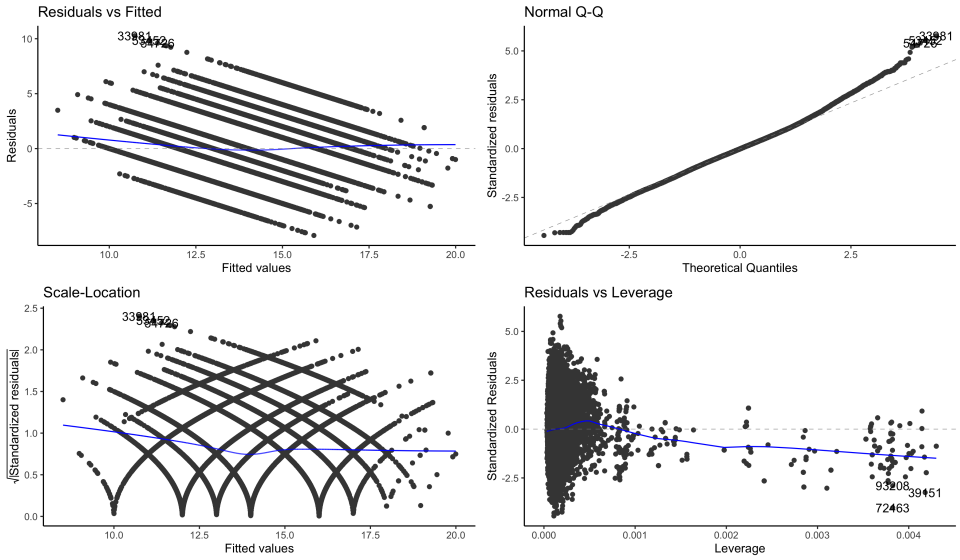
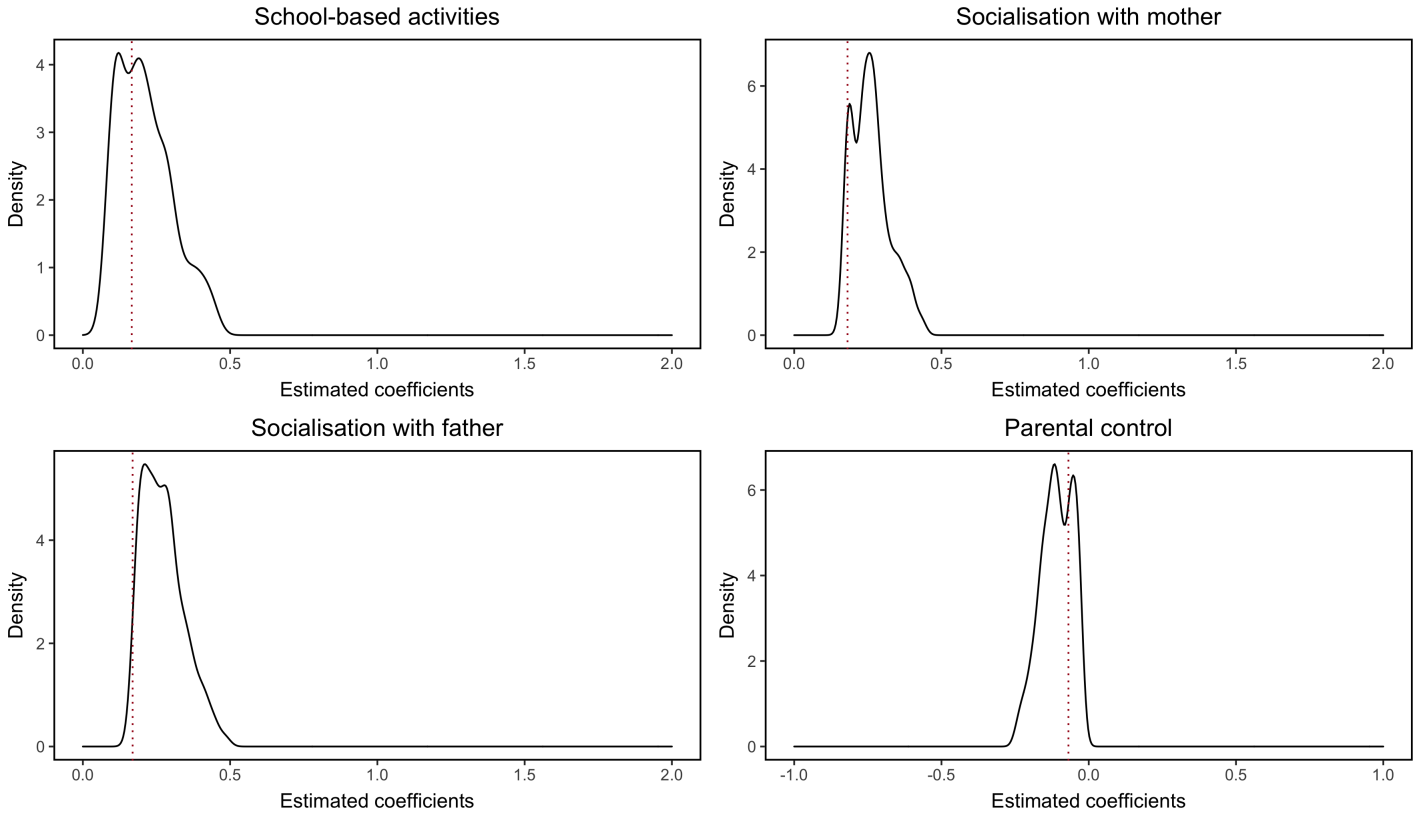


Figure 2.A.12: Diagnostic plots for the model associating years of education with parental control using the Add Health.



Kernel density graph of estimates from 512 models. The red vertical dotted line indicates the preferred estimate as reported in Table 2.6

Figure 2.A.13: Modeling distribution of the association between each dimension of parental involvement and years of education estimated using the Add Health.

Chapter 3. Genetic Predisposition to Education: The Role of Cognitive and Non-Cognitive Skills

Genetic Predisposition to Education: The Role of Cognitive and Non-Cognitive Skills

Abstract

Recent studies have shown that, in addition to being socially transmitted, education advantages can also be transferred via genetic channels. However, the exact mediating mechanisms of such transmission have remained elusive. It has been shown that genetic predisposition to education to large extent operates through cognitive ability. However, there is little research investigating the mediating role of non-cognitive skills. In this paper, I aim at filling this gap by asking whether non-cognitive skills also mediate the association between the educational attainment polygenic score, which intends to capture one's genetic propensity to education, and educational attainment. Moreover, since social contexts can facilitate or impede the realisation of one's genetic potential (Herd et al., 2019; Branigan et al., 2013), it is important to explore social heterogeneity in the mediation process. First, I run a mediation analysis on an imputed sample of 5,414 individuals and find that IQ and openness are the strongest mediators. Second, I take into account the role of social environment and investigate potential differences in the mediating effects by gender as well as father's education, which serves as a proxy for socio-economic status (SES). I find that almost all cognitive and non-cognitive skills measured in this paper are better at explaining differences in educational attainment for men than for women and for individuals from a higher-SES background than for those with lower SES. However, a differential explanatory pattern emerges for the polygenic score: it is a much better predictor of cognitive skills for men than for women and a much better predictor of non-cognitive skills for women than for men. This analysis provides new evidence on the interplay of genetic and social mechanisms operating in the context of educational attainment.

Keywords: educational attainment, polygenic score, non-cognitive skills, cognitive skills, Wisconsin Longitudinal Study

3.1 Introduction

Educational attainment is one of the most important life outcomes studied by social scientists. Social stratification researchers have shown that a higher number of years in education predicts higher wages and occupational prestige, life satisfaction, and health (Hout, 2015). Parents with higher levels of education create culturally enriched high-literacy environments (Cairns et al., 1998; Teachman, 1987) where they can spend time and resources reading (de Graaf et al., 2000; Teachman, 1987), imparting content and strategic academic knowledge (Pfeffer, 2008), acting as role models (Haveman and Wolfe, 1995; Parsons et al., 1982), investing in education and making school choices for their children (Chevalier et al., 2003). Less is known about another mechanism of educational persistence — genetic predisposition to educational attainment. Similarly to the environmental ones, genetic channels help to transfer skills, attitudes, and preferences from parents to children (Black et al., 2019). Twin and adoptee studies have shown that education is a phenotype with a high heritability of 35-77% (Branigan et al., 2013).

The evidence from twin studies has recently been strengthened with the help of molecular genetic tools, one such tool being a polygenic score which measures one's propensity for education. Such educational attainment polygenic score has been successfully used in sociological literature to establish the role of genetics in education (de Zeeuw et al., 2014; Krapohl and Plomin, 2016; Herd et al., 2019). For example, Belsky et al. (2016) find that even after controlling for one's socio-economic status the educational attainment polygenic score could predict different life outcomes, including completed education. Similarly, Domingue et al. (2015) report that siblings with the higher polygenic score are expected to complete more years of education. However, while existing research has established direct associations between the educational attainment polygenic score and educational achievement measures, the pathways through which this association operates are less clear. It is reasonable to expect cognitive skills to be one of such pathways: educational attainment polygenic score has been shown to be highly predictive of cognitive skills and cognitive skills, in turn, are important for learning and memory consolidation (Lee et al., 2018). However, the literature has so far mostly ignored another set of skills. It is now widely acknowledged that non-cognitive skills

such as work habits, personality traits, and approaches to learning also greatly matter for one's academic success, making those skills good candidates to play an important role in the association between the polygenic score and years of education (Heckman, 2006; Duckworth and Seligman, 2005; Borghans et al., 2008).

In this paper, I will first look at whether and to what extent different non-cognitive skills mediate the association between the educational attainment polygenic score, which measures genetic predisposition to education, and years of education. I will compare the mediating roles between different non-cognitive skills and between non-cognitive skills and cognitive skills, measured as IQ. My goal is not to judge which skills — cognitive or non-cognitive — are more important. The concept of non-cognitive skills is fundamentally different from cognitive skills: non-cognitive skills are not a cohesive unit for which a factor can be extracted (as is more or less the case for IQ). Rather, it is everything that is left after cognitive skills are taken into consideration — it is the residuals (Demange et al., 2021; Heckman and Rubinstein, 2001). Therefore, cognitive skills are not competitors or something to be juxtaposed to non-cognitive skills but rather they serve as a reference to the effect sizes that can be expected.

Moreover, because social contexts can also affect how much of one's genetic potential is actually realised (Herd et al., 2019; Branigan et al., 2013), the heterogeneity in mediation by gender and socio-economic status will also be explored. There has not been a lot of research focusing on such differences in the context of non-cognitive skills or heritability of education. In other words, while average mediation affects are already interesting, differences by gender and socio-economic status can provide further nuance about these underexplored mechanisms.

This paper aims at filling in important gaps in the existing literature and making several contributions. It opens up the black box of the educational attainment polygenic score and investigates the mediating mechanisms in its association with educational attainment. It also contributes to the ongoing debate on the genetic determinants of various non-cognitive skills. Furthermore, it explores the heterogeneity of mediating effects by one's social characteristics such as gender and socio-economic status. Overall, this paper constitutes one of the first attempts to employ a comparative, systematic, and theory-driven approach in studying the roles of cognitive and non-cognitive

skills in the context of genetic predisposition to education.

3.2 Theoretical framework

3.2.1 Genetic predisposition to educational attainment

Social scientists have traditionally focused on the social transmission of education advantages. Better educated parents are more likely, compared to their less educated counterparts, to invest in their children's education through financial means (Hout, 2015; Chevalier et al., 2003), quality time inputs (Teachman, 1987; Lareau, 2002; Pfeffer, 2008), or active socialisation and role-modeling (Haveman and Wolfe, 1995; Black et al., 2019; Parsons et al., 1982). In addition to these social mechanisms, however, educational advantages can be also transmitted genetically.The existing research shows that educational attainment (years of education) is highly heritable (Polderman et al., 2015). Heritability is a proportion of variation of a trait in a population that is attributable to genetic differences (Mills et al., 2020b). Twin and adoptee studies demonstrate that heritability of educational attainment can range from 35 to 77%, with the exact estimate contingent on a specific country, time period, and gender (Branigan et al., 2013). With the advent of molecular genetic tools, the biological architecture of the genetic predisposition to education can be investigated in more detail. The educational attainment polygenic score constructed from the latest educational attainment GWAS can explain 11-13% of the variance in years of schooling (Lee et al., 2018). As a consequence, the question is no longer whether education is heritable but the extent to which it is heritable (McGue et al., 2007). The exact mechanisms of its heritability, however, are not well understood since the association between the genetic predisposition to educational attainment, which can be measured with the polygenic score, and years of education is not direct but rather mediated by intermediate phenotypes, known as endophenotypes. Researchers have suggested that the most plausible candidates to be such endophenotypes are cognitive and non-cognitive skills (Benjamin et al., 2012; Okbay et al., 2016; Belsky et al., 2016).

3.2.2 Potential mediators: cognitive skills and non-cognitive Skills

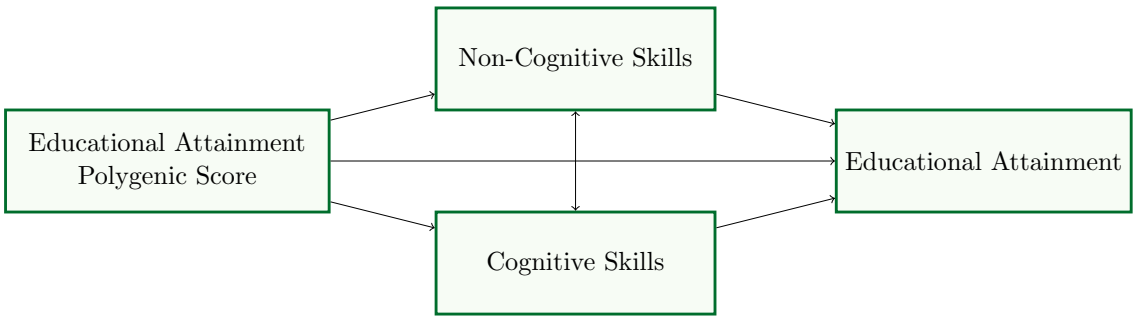


Figure 3.1: Theoretical causal graph of mediating mechanisms.

In order to play a role in explaining the genetic predisposition to educational attainment, the mediating factors need to be both heritable themselves and predictive of educational attainment (Figure 3.1). With regards to the determinants of educational achievement and attainment, historically the focus has been on cognitive skills, also known as the ability to solve abstract problems (Brunello and Schlotter, 2011). However, later it was understood that the early literature on the human capital formation completely ignored another important predictor of education and other life outcomes — non-cognitive skills, which denote personality traits and behavioural characteristics (Heckman and Rubinstein, 2001). Both cognitive and non-cognitive skills can be transmitted intergenerationally through family investments, socialisation, and imitation (Carneiro et al., 2006). Moreover, the genetic transmission of these skills is also at play: cognitive and non-cognitive skills are heritable (Borghans et al., 2008). Because these two sets of skills have a heritable component and are predictive of one's education, they are two plausible mediators in the association between the genetic propensity to education and years of education. The consensus on their relative heritabilities, however, has not been reached (Bowles et al., 2005). Although it is often difficult to conceptually and empirically distinguish cognitive and non-cognitive skills, in this paper they will be treated as separate constructs. Consequently, definitions, associations with educational attainment, and heritability estimates for cognitive and non-cognitive skills will be presented separately.

Cognitive skills

Cognitive skills are considered to broadly define the ability to understand complex and abstract ideas (Jencks, 1972b; Brunello and Schlotter, 2011), interpret informational symbols (Gintis, 1971), or adapt to unfamiliar environments through reasoning (Neisser et al., 1996). It is an umbrella term which often includes IQ, cognitive competence, general cognitive ability, or general mental ability (Jensen, 1998). According to Bartholomew (2004), intelligence, once a very popular concept, is too imprecise to be used for scientific purposes, making "cognitive skills" a preferred term in the context. Cognitive skills are often operationalised as either IQ (an intelligence quotient) or g — a general intelligence factor. IQ is an index computed from scores obtained on a set of test items (Bartholomew, 2004).

Cognitive skills have been shown to be important for many life outcomes, including occupational and educational trajectories (Fergusson et al., 2005; Batty et al., 2007). In general, greater cognitive ability makes knowledge acquisition much easier and faster, which, consequently, allows to perform better at academic tasks (Borghans et al., 2016). Better educational performance then, in turn, both indirectly and directly affects ultimate educational attainment. With regards to the indirect effect, higher performance brings a certain level of comfort to students and allows them to more fully enjoy school life. It can then result in higher intrinsic motivation and increased effort, which then translates into a higher desired level of education. Higher academic performance can also act directly by simply allowing students to meet the standard required to attain higher levels of education (Borghans et al., 2016). Empirical literature confirms the importance of cognitive skills in the educational context. Finn et al. (2014) demonstrate that IQ scores from the Wechsler Intelligence Scale for Children could predict fourth- and eight-grade test scores. Similarly, Neisser et al. (1996) find that IQ scores are predictive not only of childhood educational achievement but also of educational attainment in adulthood. Even among individuals with high levels of cognitive skills IQ is still a good predictor of educational levels (Gensowski, 2018).

It is well established in the academic literature that cognitive skills are heritable. Twin design estimates put family heritability of g at around 30-50% in childhood and 60-70% in adulthood

(Deary et al., 2009; Devlin et al., 1997; Krapohl and Plomin, 2016). The educational attainment polygenic score, which in this paper will encapsulate one's genetic predisposition to education, can explain 7-10% (Lee et al., 2018) of the variance in cognitive skills. Overall, since cognitive skills are both predictive of educational outcomes and genetically predisposed, it is reasonable to claim that they indeed play a role in the inheritance of educational attainment (Hypothesis 1). However, the role of the skills which are not cognitive should also be investigated.

Non-cognitive skills

The term of non-cognitive skills was introduced by Bowles and Gintis (1976) and more widely adopted by Jencks (1979). The ever-growing interest in non-cognitive skills, which are also called "soft" (Heckman et al., 2006), "character" (Gutman and Schoon, 2013), or "non-intellective" (Richardson et al., 2012), comes from the realisation that individual's success depends on a variety of factors beyond cognitive abilities (Groves, 2005). Although not consistently measured and defined (Pearce et al., 2016), the concept of non-cognitive skills generally encompasses personality traits (Heckman and Kautz, 2012a), social behaviour (Gutman and Schoon, 2013), work habits (Richardson et al., 2012), or patterns of thoughts and feelings (Borghans et al., 2008). Although theoretically very promising, research into non-cognitive skills is often complicated by measurement and data availability issues. Non-cognitive skills are usually self-reported, introducing biases and measurement error (Borghans et al., 2008; Groves, 2005). Furthermore, even if a given dataset contains some information on non-cognitive skills, the number of questions devoted to this subject is very limited (Gutman and Schoon, 2013).

This paper will focus on five specific measures of non-cognitive skills which have been demonstrated to be important for educational outcomes: conscientiousness, openness, perseverance, self-efficacy, and self-esteem. A summary of definitions of these specific non-cognitive skills can be found in Table 3.1.

Non-cognitive Skill	Definitions
Conscientiousness	diligence and thoroughness (McCrae and John, 1992); prudence (Hogan, 1986); will to achieve (Digman and Takemoto-Chock, 1981); compliance with rules and standards (Borghans et al., 2008)
Openness	sensitivity to art and beauty (McCrae and Costa, 1990); creativity (Rogers, 1961); intellectual interests (Rockeach, 1960); aesthetic sensitivity (Coan, 1974); need for intellectual stimulation and variety (Borghans et al., 2008); curiosity and broadmindedness (Smith-Whooley et al., 2019)
Perseverance	persistence in doing something despite difficulty (von Culin et al., 2014); insistence on mastering a skill (Gutman and Schoon, 2013)
Self-efficacy	beliefs about the degree of control (Bandura, 1982); beliefs that achievement is dependent on one's initiative
Self-esteem	person's appraisal of one's personal value (Leary and Baumeister, 2000); degree to which an individual perceives herself positively (Baumeister, 1993)

Table 3.1: Summary of the definitions of non-cognitive skills.

Non-cognitive skills are important for educational outcomes. For purely performance outcomes, such as standardised tests, IQ has been shown to be more important since it measures the speed at which you can analyse and manipulate novel information. However, for more holistic educational outcomes such as years of education, where consistent work ethic, following the rules, social and behavioral skills, and a passion for learning matter as much as grades, non-cognitive skills are essential. Research shows that self-esteem, self-efficacy, and perseverance, or grit, all matter for educational outcomes, including grades and college completion (Heckman, 2006; Andrisani and Nestel, 1976; Penke et al., 2007; Marsh and Craven, 1997; Laidra et al., 1981; Duckworth and Seligman, 2005; Wang and Eccles, 2012; Alan et al., 2019; Moffitt et al., 2011; Duncan et al., 2007; Mischel et al., 1989; Shoda et al., 1990). However, meta-analyses consistently show that among all non-cognitive skills, two concepts from the Big 5 framework stand out — conscientiousness and openness (Smithers et al., 2018; Poropat, 2009, 2014). Conscientiousness and good work habits developed in childhood are associated with later school success (Farkas, 2003; Cawley et al., 2001). For example, spending more than 7 hours per week on homework in primary school predicts completion of post-secondary education (Deke and Haimson, 2006). Poropat (2014) even shows

that conscientiousness measured at high school can be as predictive of later educational attainment as cognitive skills. Overall, conscientiousness has been consistently shown to be the best personality predictor for college grades and second best predictor for years of education overall (Meijers et al., 2006; Poropat, 2009; Borghans et al., 2016). Another Big 5 personality trait which can predict educational outcomes is openness (Meyer et al., 2019; von Stumm and Ackerman, 2013). More open, or curious, individuals tend to have more intrinsic interest in the process of learning and, hence, to seek higher levels of education (Almlund et al., 2011; Laidra et al., 2007).

Non-cognitive skills are highly heritable. The estimates for the family heritability of conscientiousness average 45% (Jang et al., 1996; Loehlin et al., 1998; Riemann et al., 1997; Tucker-Drob et al., 2016); openness — around 55% (Jang et al., 1996; Loehlin et al., 1998; Tucker-Drob et al., 2016). Only few existing studies have investigated heritability of perseverance, self-efficacy, or self-esteem estimate, and those estimate are at around 40% (Rimfeld et al., 2016; Krapohl et al., 2014; Greven et al., 2009; McGue et al., 2007).

The high family heritability estimates do not mean that the genetic transmission channels for these specific skills overlap with the channel for educational attainment. In other words, the role of non-cognitive skills in the transmission of educational attainment is well documented but the same cannot be said about the genetic part of this transmission — the one we are interested in in this paper. Preliminary evidence is inconclusive. Smith-Whooley et al. (2019) find the educational attainment polygenic score from the third educational attainment GWAS can predict 0.6% of the variance in conscientiousness and 2.1% of the variance in openness. Okbay et al. (2016) estimate that openness accounts for around 7% of the association between the educational attainment polygenic score and years of education. No previous studies which would explore GWAS heritability of perseverance, self-efficacy, or self-esteem exist at the moment. I aim to fill in this gap by systematically studying the role played by five non-cognitive skills in the genetic transmission of education. We can predict the genetic transmission to approximate the phenotypic one. In other words, we can expect non-cognitive skills, which play the most important role in explaining differences in educational attainment, to also play the most important role in mediating the genetic predisposition to educational attainment. Consequently, I hypothesise that conscientiousness and openness

will be the most important non-cognitive endophenotypes in the mediation process because they have been consistently shown to predict educational attainment the best of all non-cognitive skills (Hypothesis 2).

3.2.3 Heterogeneity by gender and SES

Investigating only average mediating effects might mask important heterogeneities in terms of gender or socio-economic status. As Herd et al. (2019) report, in the WLS, a cohort study for those born between 1939 and 1940, the association between genetics and education is much stronger for men than for women. Because of institutional, cultural and social constraints which women faced at that time in order to acquire higher levels of education, even those with high genetic propensities could not realise their full academic potential. In other words, the study of genetic transmission of educational privileges should follow a sociocontextual perspective and incorporate the particularities of different historical periods, cohorts, or social forces (Branigan et al., 2013; Conley and Domingue, 2016). As Figure 3.2 shows, the heterogeneity by gender can occur in two places in our mediation model. Gender could moderate the association between the educational attainment polygenic score and cognitive/non-cognitive skills. This might results from differential parenting or teacher behaviour directed towards boys and girls, which could have stifled the realisation of girls' cognitive and non-cognitive skills. Moreover, gender can also introduce differences in the relationship between skills and educational attainment. We can imagine that in a social environment where women's educational potential is impeded, even very high cognitive and non-cognitive skills might not translate into educational advantages (Herd et al., 2014; Heath et al., 1985; Turkheimer et al., 2003). Overall, we can expect cognitive and non-cognitive skills to play a more important role in explaining the genetic predisposition of educational attainment for men compared to women (Hypothesis 3).

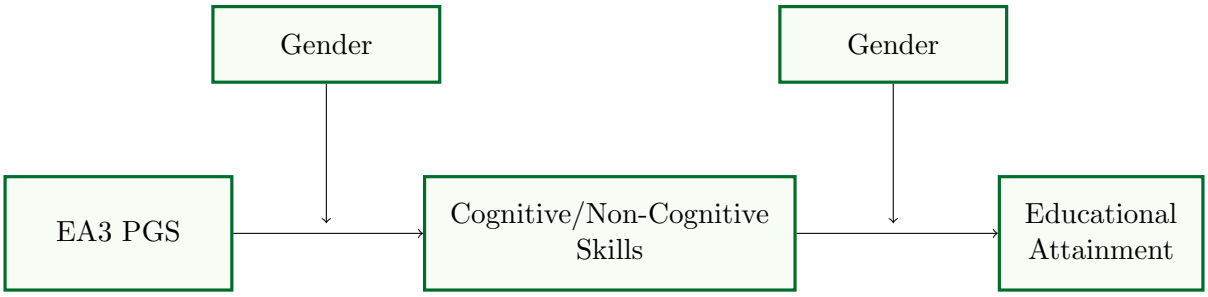


Figure 3.2: Theoretical causal graph of mediating mechanisms with the moderation by gender.

Genetic differences by socio-economic status (SES) are also likely. In this context, individual's social and genetic inheritances are correlated (Domingue et al., 2015; Krapohl and Plomin, 2016). As demonstrated in Figure 3.3, parental characteristics, such as education, occupation, or income, are likely to be positively associated with parents' and, in turn, child's genetic predisposition to education. Socio-economic status can also influence both cognitive/non-cognitive skills and educational attainment through socialisation and parental investments in child's development (Cunha and Heckman, 2008; Conti and Heckman, 2010). By looking at the mediation process for different levels of SES separately we can both take care of potential confounding by SES and investigate whether there are interesting heterogeneous patterns of the role of cognitive and non-cognitive skills. In general, we can expect higher SES individuals to have more opportunities to successfully turn their partially genetically influenced cognitive and non-cognitive skills into tangible educational advantages through more enriched learning environments at home and schools which reward such skills (Lareau, 2002; Heath et al., 1985; Rowe et al., 1999). Thus, it is probable that cognitive and non-cognitive skills play a more important role in explaining genetic transmission of education for individuals from more privileged socio-economic backgrounds (Hypothesis 4).

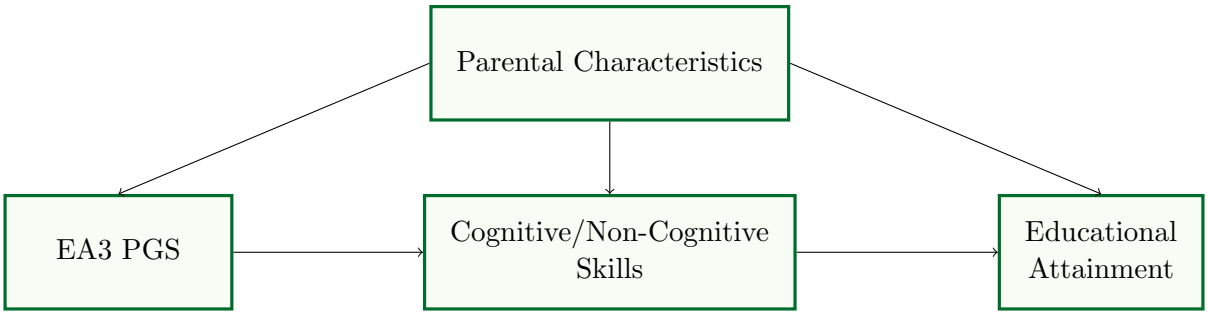


Figure 3.3: Theoretical causal graph of mediating mechanisms with the confounding by parental characteristics.

3.3 Data

3.3.1 Dataset

The stated hypotheses in this paper will be tested using the Wisconsin Longitudinal Study (WLS). This dataset is uniquely suited for our needs because it contains an extensive questionnaire on respondents' personality traits (non-cognitive skills), the Henmon-Nelson IQ test which measures cognitive skills and has molecular genetic data. The original survey was first administered in 1957 and collected information on a random sample of one third of all graduating seniors at Wisconsin's high schools who were born in 1939-1940. The response rate from the sampling frame was 100% (Herd et al., 2019). The dataset follows respondents from the first data collection in late adolescence in 1957 throughout their adulthood with additional data collections in 1964, 1975, 1992, 2004, and 2011. It has rich information on family background, adolescent characteristics, educational history, employment experiences, income and wealth, family formation, physical and mental health, psychological well-being, cognition, end of life planning, and mortality (Herd et al., 2014). Saliva samples for DNA sequencing were collected in 2006 to 2007 using mailed Oragene kits and in 2011 during individual interviews. ∼710,000 markers were genotyped using Omni-Express BeadChip at Johns Hopkins' Center for Inherited Disease Research (Herd et al., 2014). Despite its numerous strengths, the dataset poses challenges for the extrapolation of the results to wider populations because almost all respondents are non-Hispanic whites and because of the truncated measure of

educational attainment (Hauser and Palloni, 2011).

3.3.2 Sample size

The original WLS questionnaire in 1957 was administered to 10,317 high school graduates. Over the years, 8,729 of the graduates' siblings were interviewed, making a total sample size of 19,046 individuals. Since the focus of this paper is on education, only the graduates are included in the analyses (Figure 3.A.1). Genetic information and, consequently, the educational attainment polygenic score is available for 5,414 graduates. 5,384 of those graduates have reported their years of education (Figure 3.A.1). Because around 16-18% of information is missing on non-cognitive skills (the pattern of missingness can be seen in Figure 3.A.2 in the Appendix), a listwise deletion of observations with at least one missing non-cognitive entry would have resulted in a much reduced sample size of 4,028. Consequently, missing data was imputed for the respondents with genetic information, resulting in the analytical sample of 5,414 individuals. More details about the method of multiple imputation which was used in this paper can be found in the Appendix.

3.4 Measurement

3.4.1 Outcome variable — educational attainment

My dependent variable of educational attainment is measured as years of education. In 1975 study participants had to report how much schooling they had obtained since graduating from high school in 1957. The categorical responses to this question were recoded into a numerical form to facilitate the analysis and the interpretation of results. The category of *high school graduate or less, or less than one year of college* was recoded into 12 years of education; *one-year college certificate, one year of college* — 13 years; *two-year teaching certificate, associate degree, two-year diploma or certificate, two years of college* — 14 years; *college trained registered nurses, one or more years post two-year degree (but no higher degree) — 15 years; bachelor's degree — 16 years; master's degree*

with major field of continuing/vocational education, curriculum, education, elementary education — 17 years; *two-year master's* — 18 years; *LLD, LLB, JD, or one or more years post two-year-master's level degree* — 19 years; *PhD, MD, other doctorates not previously included, or one or more years post-professional degree* — 20 years. The distribution of the semi-continuous measure for educational attainment can be seen in Figure 3.A.4 in the Appendix. Since the sampling frame of the WLS included only high school graduates, 12 years is by definition the lowest amount of schooling the WLS participants could have achieved. Such selection on the dependent variable should be kept in mind when interpreting the results. 60% of the respondents report not going beyond 12 years of schooling in the analytical sample. Around 15% of the participants received a four-year bachelor's degree corresponding to 16 years of education. The measure ranges to 20 years with around 2% of the individuals receiving a PhD or an MD. The mean for years of education is 13.55 with a standard deviation of 2.23 (Table 3.1).

3.4.2 Explanatory variable — cognitive skills

Cognitive skills in the WLS are measured with the Henmon-Nelson IQ test. The measure of IQ score mapped from the raw Henmon-Nelson test score will be used in the analysis because it is the recommended IQ measure of the WLS administrators. The test was administered during respondents' junior year of high school in 1956 and was later linked with the WLS data. The test is a 30-minute multiple-choice assessment which asks 90 verbal and numerical questions (Henmon and Nelson, 1946). The Henmon-Nelson test was widely used in Wisconsin from the 1930s to the 1960s and aimed at identifying and encouraging gifted adolescents, who might not have otherwise considered pursuing further education, to attend college (Hauser and Palloni, 2011). The result on the test in our analytical sample ranges from 61 to 145. The distribution of the variable approximates a normal distribution (Figure 3.A.4 in the Appendix). The mean IQ is 102.22 with a standard deviation of 14.55 (Table 2). The measure is standardised to z-scores (mean = 0, SD = 1) before using it in the analysis to make the interpretation of estimates easier.

3.4.3 Explanatory variables — non-cognitive skills

Five variables measuring non-cognitive skills will be used in the analysis (Table 3.A.2). The WLS has information on some other dimensions of non-cognitive skills (for example, the Big 5 constructs of neuroticism, agreeableness, and extroversion). However, after a thorough review of literature, it was decided to focus on perseverance, self-efficacy, self-esteem and two strongest candidates to be the mediators — conscientiousness and openness. These concepts are either directly measured with the Big 5 questionnaire (conscientiousness and openness) or inferred from the more general personality questions. All five non-cognitive skills were measured in 1993, when the respondents had already entered advanced adulthood. One might raise a question whether non-cognitive skills measured in adulthood can be predictive of educational outcomes achieved earlier in life. However, there is strong evidence supporting the argument that non-cognitive skills display a high degree of longitudinal rank-order consistency over one's lifespan starting in late adolescence/early adulthood (Harris et al., 2016; Groves, 2005; Briley and Tucker-Drob, 2014a).

Conscientiousness

The composite measure of conscientiousness, directly provided by the WLS, is constructed using an abbreviated version of the Big Five Inventory from Berkeley Personality Lab (John et al., 1991). It is a summary score of six questions with a mean of answered items substituted for missing items. The six questions measuring conscientiousness are the following: *1) To what extent do you agree that you see yourself as someone who does a thorough job?; 2) To what extent do you agree that you see yourself as someone who is a reliable worker?; 3) To what extent do you agree that you see yourself as someone who tends to be disorganized?; 4) To what extent do you agree that you see yourself as someone who is lazy at times?; 5) To what extent do you agree that you see yourself as someone who does things efficiently?; 6) To what extent do you agree that you see yourself as someone who is easily distracted?* The scale for each question ranges from 1 to 6, the score of 6 meaning the most conscientiousness. In our analytical sample the summary score ranges from 8 to 36 and has a mean of 29.20 with a standard deviation 4.13 (Table 3.1). The summary score

for conscientiousness found in the WLS has a Cronbach's alpha of 0.66 and has been extensively used in previous research and shown as having good reliability and forming a unidimensional scale (Hauser et al., 2003; Mueller and Plug, 2006; Flynn and Smith, 2007). All item loadings are above 0.6. In the analysis the standardised version of the conscientiousness score will be used with a mean of 0 and a standard deviation of 1 (as is the case for all other non-cognitive skill measures).

Openness

The measure of openness can also be found in the Big 5 Questionnaire. The summary variable is constructed as a sum of values on the following six items: *1) To what extent do you agree that you see yourself as someone who prefers the conventional, traditional?; 2) To what extent do you agree that you see yourself as someone who prefers work that is routine and simple?; 3) To what extent do you agree that you see yourself as someone who values artistic, aesthetic experiences?; 4) To what extent do you agree that you see yourself as someone who has an active imagination?; 5) To what extent do you agree that you see yourself as someone who wants things to be simple and clear-cut?; 6) To what extent do you agree that you see yourself as someone who is sophisticated in art, music, or literature?*. The coding system for this variable is the same as for conscientiousness. In our sample the lowest score for openness is 6 and the highest is 36. Cronbach's alpha for this measure is 0.64. The mean is 22.90 and the median is 22. The standard deviation is 4.81 (Table 3.1). Factor analysis demonstrates that an optimal number of factors for the six items is two and not one (as chosen by the administrators of the WLS). Items 1, 2, 5 and items 3, 4, 6 load onto two different factors. In the main body of the analysis I will use the overall measure provided by the WLS because of its strong theoretical underpinnings (Ackerman and Heggestad, 1997). However, I will also run my analysis for two separate factors (measured as sums of the corresponding item values) as previous research shows that openness might have two neurally distinctive facets — intellectualism, encapsulating fantasy and aesthetic sensitivity(items 3, 4, 6), and adventurousness/unconventionality (items 1, 2, 5) (von Stumm et al., 2011; DeYoung et al., 2005).

Perseverance

The measure of perseverance is constructed with four items reported in 1993: *1) To what extent do you agree that even when things seem hopeless, you keep on fighting to reach your goals?; 2) To what extent do you agree that you stick to your goals and projects even in the face of great difficulties?; 3) To what extent do you agree that the harder a goal is to achieve, the more appeal it has to you?; 4) To what extent do you agree that you can be very stubborn in pursuing your goals?* The scale for each question ranges from 1 to 5, the score of 5 characterising the most persevering respondents. The summary score was not directly provided by the WLS (in contrast to conscientiousness and openness). To construct a composite indicator for perseverance, the scores on each item were added together. Cronbach's alpha for this measure is 0.74. The four items load on one factor with item loadings all above 0.65. The constructed summary score ranges from 4 to 20, has a mean of 14.80 and a standard deviation of 2.55 (Table 3.1).

Self-efficacy

Self-efficacy is measured using the environmental mastery dimension of the Ryff (1989) psychological well being scale, following (Carr, 2020). The scale has five items: *1) To what extent do you agree that you are good at juggling your time so that you can fit everything in that needs to get done?; 2) To what extent do you agree that you often feel overwhelmed by your responsibilities?; 3) To what extent do you agree that you are quite good at managing the many responsibilities of your daily life? 4) To what extent do you agree that you do not fit very well with the people and community around you? 5) To what extent do you agree that you have difficulty arranging your life in a way that is satisfying to you?* . The values for each item range from 1 to 6, the score of 6 showing the most self-efficacy. To construct a composite indicator for self-efficacy, the scores on each item were added together. Cronbach's alpha for this measure is 0.70. The five items load on one factor with item loadings all above 0.62. The mean of the summary score is 23.70 with a standard deviation of 3.99 (Table 3.1).

Self-esteem

Self-esteem is measured with three questions asked in 1993, which closely follow the original self-worth items of Rosenberg (1965): *1) To what extent do you agree that, in general, you feel confident and positive about yourself?; 2) To what extent do you agree that when you compare yourself to friends and acquaintances, it makes you feel good about who you are?; 3) To what extent do you agree that your attitude about yourself is probably not as positive as most people feel about themselves?.* The values for each self-esteem questions range from 1 to 6, the score of 6 describing respondents with the most self-esteem. To construct a composite indicator for self-esteem, the scores on each item were added together. Cronbach's alpha for this measure is 0.74. The five items load on one factor with item loadings all above 0.66. The variable has a mean of 14.90 and a standard deviation of 2.70 (Table 3.1).

Correlations between different skills can be found in Table 3.A.4 in the Appendix. IQ is moderately correlated with openness (r = 0.253) and only weakly correlated with other non-cognitive skills. Non-cognitive skills are moderately correlated to one another and those correlations range from around r = 0.2 to r = 0.5, with most correlations centered around r = 0.3. This shows that while these five non-cognitive measure have some fundamental similarity to one another and are distinct from the concept of cognitive skills, they, nevertheless, capture unique dimensions of one's personality.

3.4.4 Educational attainment polygenic score

The educational attainment polygenic score is a single aggregate measure of the multiple genetic effects correlated with educational attainment (Belsky and Israel, 2014; Ware et al., 2017; Vilhjamsson et al., 2015). The polygenic score for an individual i is equal to the sum of the allele counts a_{ij}(0,1 or 2) from each SNP j = 1,..M, multiplied by a weight w_j: $PGS_i = \sum_{j=1}^{m} w_j a_{ij}$ (Mills et al., 2018), where the weights are association coefficients produced in the educational attainment genome-wide association study (GWAS). In this paper the polygenic score from the latest

1.1 million-person educational attainment GWAS will be used (Lee et al., 2018). It's the third educational attainment GWAS and, as is convention, the educational attainment polygenic score in this paper will be abbreviated as the EA3 PGS. Individual's educational attainment polygenic score can then be interpreted as a single quantitative measure of one's genetic propensity to educational attainment (Mills et al., 2018). The polygenic score was constructed by the WLS using the LDPred approach. It is a Bayesian method which does not discard any redundant SNPs (those in linkage with other SNPs) but rather weights each SNP by its conditional effect given other SNPs. A posterior effect size of each SNP is inferred from a prior on its effect size and linkage disequilibrium information from an external panel (Vilhjamsson et al., 2015). In our analytical sample the EA3 PGS has a mean of -0.2 and a standard deviation of 0.14 (Table 3.1). To facilitate the analysis, the score was standardised to have a mean of 0 and a standard deviation of 1. The distribution of the standardised measure, which is very close to a standard normal distribution, can be seen in Figure 3.A.4 in the Appendix.

3.4.5 Covariates

Gender

The measure of gender in the WLS is self-reported. 52.3% of the respondents are male and 47.6% — females (Table 3.1).

Socio-economic status

Socio-economic status (SES) is measured as father's education. The categories and their corresponding numerical values in years of schooling are the following: *1) Did not attend high school — 8 years; Attended high school — 10 years; Graduated from high school — 12 years; Attended trade or business school - 14 years; Attended college — 14 years; Graduated from college — 16 years; Has master's or PhD — 19 years.* The categorical variable for father's education was recoded into a numerical one. It has a mean of 10.85 and a standard deviation of 2.89 (Table 3.1). In the analyses

exploring mediation heterogeneity by SES, a binary variable will be used: 0 if father has less than a high school diploma and 1 if father has a high school diploma or higher educational credentials. This particular dichotomisation, and not the more common one which divides individuals in those with no college degree and those with a college degree, was chosen because only 5% of the fathers in the sample have a bachelor's degree or a higher level of education (Figure 3.A.4 in the Appendix). Also, instead of including both mother's and father's education, only father's education is used for two reasons. First, the education levels of both parents are rather strongly correlated (r=0.5). Second, because of the historical context, there is little variation in mother's educational attainment since half of the mothers of individuals who graduated from high school in 1957 did not have more than 10 years of education.

3.4.6 Principal components

Although the WLS participants are predominantly of Northern European ancestry, the ancestral difference between, for example, individuals of Polish and British descents are, nevertheless, genetically discernible (Herd et al., 2019). Consequently, 10 genetic principal components will be included in the analysis to account for population stratification, also known as allele differences within specific ancestral groups, to avoid a potential risk of confounding by ancestry differences (Belsky et al., 2018). The WLS provides ready-made principal components.

Table 3.1: Descriptive statistics of the Wisconsin Longitudinal Study.

Variable	Mean	Median	Range	SD	N
Years of education	13.55	12.00	12.00 - 20.00	2.23	5,414
Father's years of education	10.85	10.00	8.00 - 19.00	2.89	5,414
Gender					5,414
Female	0.52				2,830
Male	0.48				2,584
EA3 PGS	-0.20	-0.21	-0.67 - 0.37	0.14	5,414
IQ	102.00	102.00	61.00 - 145.00	14.70	5,414
Conscientiousness	29.20	30	8.00 - 36.00	4.13	5,414
Openness (overall)	21.90	22.00	6.00 - 36.00	4.81	5,414
Openness (intellectualism)	12.50	13.00	3.00 - 18.00	3.05	5,414
Openness (adventurousness)	9.36	9.00	3.00 - 18.00	3.10	5,414
Perseverance	14.80	15	4.00 - 20.00	2.55	5,414
Self-efficacy	23.70	24.00	5.00 - 30.00	3.99	5,414
Self-esteem	14.90	15.00	3.00 - 18.00	2.70	5,414

Notes: EA3 PGS = educational attainment polygenic score.

3.5 Analytical strategy & methods

Mediation analyses are used to evaluate the relative magnitude of different mechanisms through which a given variable affects an outcome (VanderWeele, 2016). Two statistical approaches are generally used to study mediation: the product method and the difference method. In the product method, made popular by Baron and Kenny (1986), two regressions are specified. The first one regresses the outcome (Y) on the explanatory variable of interest (a), the mediator (m), and the covariates (c): $Y = \theta_0 + \theta_1 a + \theta_2 m + \theta_3' c$. The second one regresses the mediator (M) on the explanatory variable of interest (a) and the covariates (c): $M = \beta_0 + \beta_1 a + \beta_2' c$. Using this approach the direct effect is then interpreted as θ_1 and the indirect (mediated) effect is taken as the product $\beta_1 \theta_2$. The second mediation approach is known as the difference method. It first fits a regression where the outcome (Y) is regressed on the explanatory variable of interest (a) and covariates (c): $Y = \phi_0 + \phi_1 a + \phi_2' c$. Then it fits regression which also includes the mediator (m): $Y = \theta_0 + \theta_1 a + \theta_2 m + \theta_3' c$. The indirect (mediated) effect is interpreted as the difference between the effect of the explanatory variable of interest (a) before the mediator is added and its effect after the mediator included in the regression: $\phi_1 - \theta_1$.

To specify and run the mediation analyses corresponding to Hypothesis 1 (the mediation by cognitive skills) and Hypothesis 2 (the mediation by non-cognitive skills) in this paper I use the "lavaan" package in R, which is geared towards structural equation modeling. While other packages exist for the implementation of mediation analyses (such as the "mediation" package in R), only "lavaan" was fully integrated with the "mice" package, which allowed me to run the mediation analyses on the datasets imputed using multiple imputation and seamlessly pool the results. The mediation approach used in this package corresponds to the previously discussed product method. Because my dependent variable (years of education) is semi-continuous and my mediators are continuous variables and because I fit my models using ordinary liner squares, the alternative approach of the difference method would yield identical results (VanderWeele, 2016). Using "lavaan" entails specifying the mediation models. I run separate models for each mediator of interest. In essence, I specify two regressions for each mediator required for the the product method:

$$EduYears_i = \theta_0 + \theta_1 PGS_i + \theta_2 Skill_i + \theta_3' c_i \tag{3.1}$$

$$Skill_i = \beta_0 + \beta_1 PGS_i + \beta_2' c_i \tag{3.2}$$

where $EduYears_i$ measures years of education for an individual i; $Skill_i$ indicates one of the cognitive or non-cognitive skills; PGS_i stands for the educational attainment polygenic score; c_i is a vector of covariates which includes father's education and 10 principal components. The indirect (mediating) effects then correspond to the product $\beta_1\theta_1$.

The main limitation of this approach is the fact that I run the mediation model for each mediator separately. It is possible that mediators affect one another (for example, non-cognitive skills can influence test scores of cognitive skills) or the effect of one mediator can interact (moderate) the effect on the outcome of another mediator. These specifications are beyond the scope of this paper. Possibilities for future research on multiple mediation is discussed in further detail in the

conclusion/discussion section of this paper.

To investigate the heterogeneity by gender and SES, considered in the Hypothesis 3 and the Hypothesis 4, our sample will be divided into two groups of male/female and high SES (father has a high school diploma)/low SES (father does not have a high school diploma) respectively. The mediation analysis will be run on each particular subgroup and resulting estimates will be compared between groups.

3.6 Results

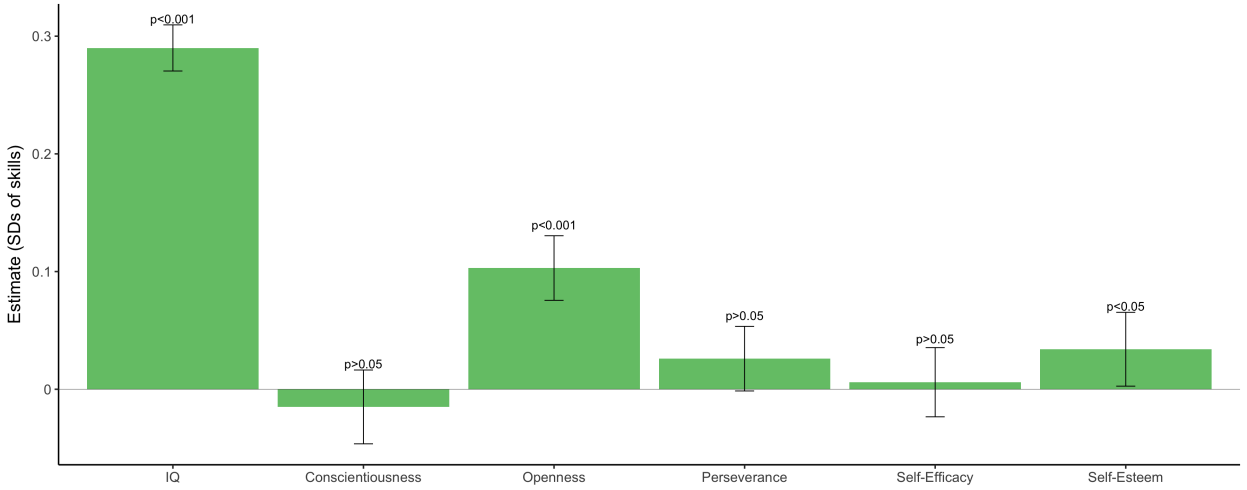
3.6.1 Mediating role of non-cognitive skills

In the first part of our analysis I am testing Hypothesis 1 and Hypothesis 2 which claims that IQ as well as cosncientiousness and openness are the most prominent mediators in the association between the educational attainment polygenic score and years of education. The results of this part are displayed in Table 3.1. The first section of the table reports direct effects of the educational attainment polygenic score on cognitive and non-cognitive skills as well as direct effects of cognitive and non-cognitive skills on our outcome of interest — years of education. Table 3.1 shows the estimates of these effects with associated standard errors (in parentheses) and p-values.

The direct effects of the polygenic score on cognitive and non-cognitive skills, reported in Table 3.1, are also summarised in Figure 3.1. The educational attainment polygenic score best explains IQ: a 1 standard deviation increase in the the polygenic score is associated with a 0.27-0.30 standard deviation increase in IQ. With regards to non-cognitive skills, the educational attainment polygenic score can only explain openness: a 1 standard deviation increase in the polygenic score corresponds to a 0.08-0.13 standard deviation increase in the measure of openness. The point estimates for the association with other non-cognitive skills range from a -0.02 standard deviations in the case of conscientiousness to a 0.03 standard deviations in the case of self-esteem. As can be seen from the error bars drawn in Figure 3.1, none of the estimates for those four non-cognitive skills can be

considered to be very reliable.

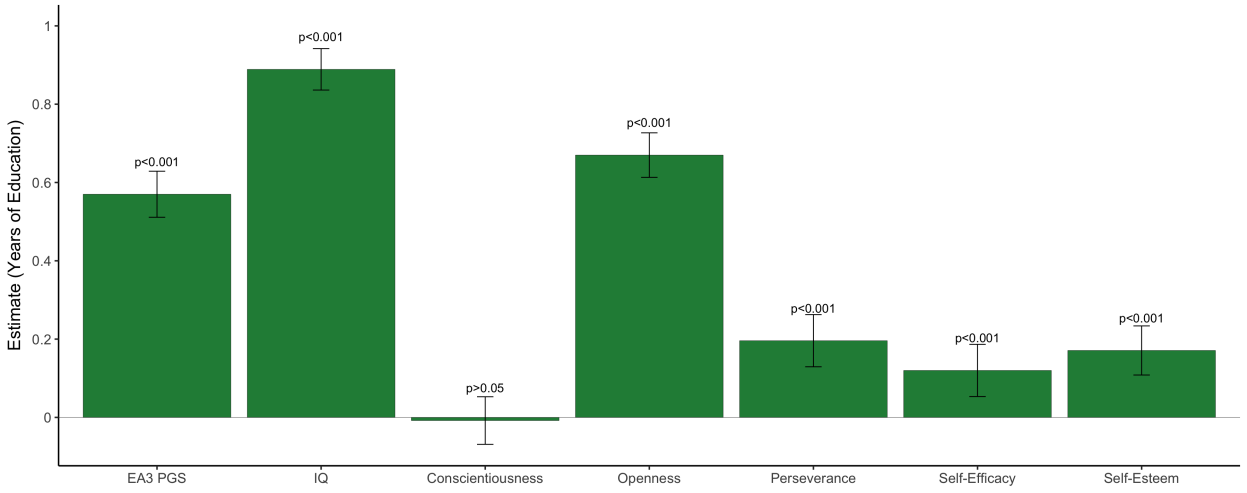
Figure 3.1: Estimated coefficients of the associations between the educational attainment polygenic score and cognitive/non-cognitive skills in the WLS.



Notes: Associations are measured as a 1 standard deviation increase in the educational attainment polygenic score (EA3 PGS) for a corresponding standard deviation change in cognitive/non-cognitive skills, with associated 95% confidence intervals represented as error bars and p-values written above the error bars.

The summary for the effects of cognitive and non-cognitive skills on years of education, which are also reported in Table 3.1, can be seen in Figure 3.2. In contrast to Figure 3.1, the x-axis in Figure 3.2 lists variables that are independent rather than dependent in this part of the analysis. When entered into the mediation models separately and compared to one another, the best predictors of educational attainment are the educational attainment polygenic score, IQ, and openness: a 1 standard deviation increase in the polygenic score is associated with a 0.51 - 0.63 year increase in educational attainment (before any mediators are added); a 1 standard increase in IQ — with a 0.83 - 0.94 year increase in educational attainment; a 1 standard deviation increase in openness — with a 0.61-0.73 year increase in educational attainment. Conscientiousness does not have any association with years of education while other three non-cognitive skills — perseverance, self-efficacy, and self-esteem — are only weakly associated.

Figure 3.2: Estimated coefficients of the associations between years of education and the educational attainment polygenic score/cognitive skills/non-cognitive skills in the WLS.



Notes: Associations are measured as a 1 standard deviation increase in the explanatory variables for a corresponding year change in years of education, with associated 95% confidence intervals represented as error bars and p-values written above the error bars.

Finally, the mediating effects of cognitive and non-cognitive skills are estimated. Those estimates can be found in the lower part of Table 3.1 under the rubric of indirect effects because these estimates represent proportions of the association between the educational attainment polygenic score and educational attainment which do not go directly but rather indirectly through the mediating factors. The mediating effect of IQ amounts to a 0.23 - 0.28 standard deviation, or around 44% of the entire association between the educational attainment polygenic score and educational attainment. Conscientiousness does not mediate the association at all while perseverance, self-efficacy, and self-esteem are weak mediators. Among non-cognitive skills only openness mediates the relationship between the educational attainment polygenic score and years of education and this mediation accounts for 12% of the whole association. Consequently, while Hypothesis 1 stating the important mediating role of cognitive skills is supported by our results, the evidence for Hypothesis 2 is more nuanced: among non-cognitive skills, openness is indeed the strongest mediator while conscientiousness plays no role in the mediation process. When comparing cognitive and non-cognitive skills, cognitive skills seem to be much more important in this context.

Figure 3.4 summarises the results from mediation models for each cognitive and non-cognitive skill.

It displays the direct effects from the educational attainment polygenic score to a particular skill, from a particular skill to years of education, and from the polygenic score to years of education. It is worth noting that the association between the polygenic score and education in this graph represents the direct association or, in other words, the association not of mediated by any of our cognitive and non-cognitive skill measures. Rather then reporting point estimates, Figure 3.4 shows 95% confidence intervals for each estimate. As shown in Figure 3.3, my mediation analysis accounts for around 60% of the association: 44% of it is mediated by IQ (cognitive skills), 12% — by openness, and 3% by the rest of non-cognitive skills investigated in this paper. The remaining 41% represent the proportion of the association mediated by some endophenotypes that are not measured in this paper.

I run the analyses not only on the general measure of openness provided by the WLS but on its two facets (intellectualism and adventurousness) which load onto two factors. The results can be found in Table 3.A.6. The mediating role of the intellectualism facet (5.60%) is much lower than that of the adventurousness facet (9.70%). Both of these estimates are lower than 12% estimated using the overall measure of openness which incorporates both facets using equal weights.

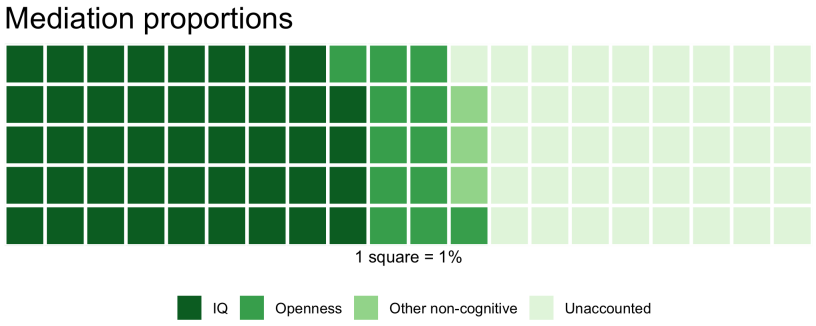


Figure 3.3: Proportions of the total association between the educational attainment polygenic score and years of education mediated by cognitive and non-cognitive skills in the WLS.

Figure 3.4: Direct and indirect effects of the educational attainment polygenic score on years of education, mediated by cognitive and non-cognitive skills in the WLS.



Notes: The association sizes are shown as confidence intervals in parentheses.

Table 3.1: Direct and indirect effects associated with the EA3 PGS, cognitive, non-cognitive skills, and educational attainment (years of education).

Direct effects/Total effects	b(SE)	p	
EA3 PGS →IQ	0.286***(0.010)	<0.001	
EA3 PGS →Conscientiousness	-0.015(0.016)	0.363	
EA3 PGS →Openness	0.103***(0.014)	<0.001	
EA3 PGS →Perseverance	0.026†(0.014)	0.090	
EA3 PGS →Self-efficacy	0.006(0.015)	0.680	
EA3 PGS →Self-esteem	0.034*(0.016)	0.024	
EA3 PGS →EA	0.573***(0.030)	<0.001	
IQ →EA	0.889***(0.027)	<0.001	
Conscientiousness →EA	-0.008(0.031)	0.787	
Openness →EA	0.671***(0.029)	<0.001	
Perseverance →EA	0.196***(0.034)	<0.001	
Self-efficacy →EA	0.120***(0.034)	<0.001	
Self-esteem →EA	0.171***(0.032)	<0.001	
Indirect effects	b(SE)	p	% mediated
EA3 PGS →IQ →EA	0.254***(0.008)	<0.001	44.30%
EA3 PGS →Conscientiousness →EA	0	0.461	0%
EA3 PGS →Openness →EA	0.069***(0.003)	<0.001	12.0%
EA3 PGS →Perseverance →EA	0.005†(0.006)	0.100	0.90%
EA3 PGS →Self-efficacy →EA	0.0007(0.004)	0.690	0.12%
EA3 PGS →Self-esteem →EA	0.006*(0.002)	<0.030	1.05%

Notes: EA3 PGS = educational attainment polygenic score. EA = educational attainment.
†p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Separate analyses were conducted for each mediator.
The EA3 PGS included in the analyses was regressed on the ten genetic ancestry principal components.
Data source: the Wisconsin Longitudinal Study.

3.6.2 Heterogeneity of mediation by gender

In the previous section I estimated average mediating effects of non-cognitive skills. Here I test whether those effects are different depending on one's gender (Hypothesis 3). These analyses can provide us with more nuance about the role of non-cognitive skills in explaining predisposition to educational attainment.

Investigating gender differences in this context is important for several reasons. As Herd et al. (2019) demonstrate, the educational attainment polygenic score is less predictive of educational attainment for women than for men in the WLS because relatively few women would get high levels of education in the late 1950s and 1960s. In other words, even women with great genetic potential could not realise it given the social constraints they faced. The educational polygenic score could also account for disparities in non-cognitive skills differently for men and women. Finally, the effect of non-cognitive skills on education might also differ by gender. I explore these possibilities in the following analyses where I divide our sample into two groups — men and women — and run our mediation models on these two groups separately. I hypothesise that cognitive and non-cognitive skills play a more important role in mediating the association between the educational attainment polygenic score and years of education for men than for women (Hypothesis 2).

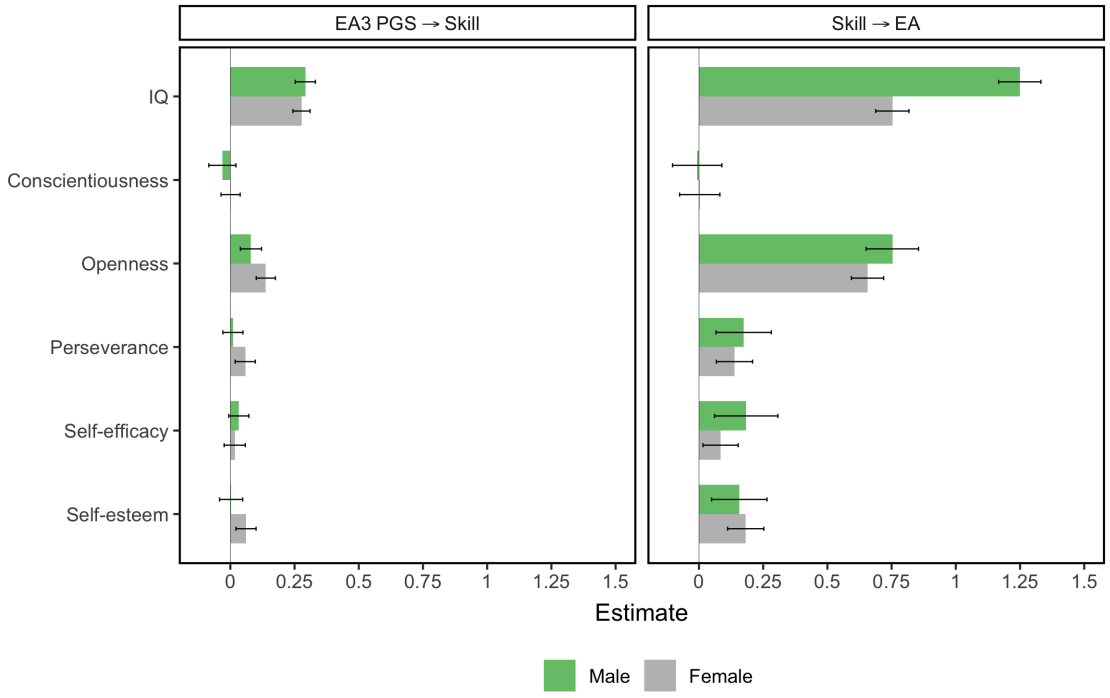
Table 3.A.7 in the Appendix shows full results for this set of analyses. We can also observe the key findings by looking at Figure 3.5. It demonstrates how the polygenic sore prediction of cognitive and non-cognitive skills differs by gender. It is better at explaining differences in IQ for men than for women; however, it is better at explaining non-cognitive skills for women than for men. For example, while for men a 1 standard deviation increase in the educational attainment polygenic score is associated with a a 0.08 standard deviation increase in openness, for women it is associated with a 0.138 standard deviation increase.

When we look at how cognitive and non-cognitive skills explain differences in years of education, gender differences also emerge (Figure 3.5). Non-cognitive skills are better predictors of education for men and women. This should not come as a surprise given that the variation in women's

education at that historical moment was much lower than for men.

The full mediation model by gender allows us to uncover an interesting fact (Table 3.A.7). As is the case for the aggregate model (Table 3.1), two main mediators emerge — IQ and openness. However, the mediated proportions of these two skills differ by gender. IQ mediates a higher proportion for men (50%) than for women (47%) whereas openness plays a much bigger role in explaining genetic predisposition for education among women (21%) than among men (8.3%). These differences in mediated proportions are statistically significant (more details can be found in Table 3.5 in the Appendix). As a consequence, our Hypothesis 3, which argued that both cognitive and non-cognitive skills will have a more major role for men than for women is not supported: while cognitive skills indeed better explain genetic transmission of education for men, non-cognitive skills are more important for women.

Figure 3.5: Estimated coefficients of the associations in the mediation model moderated by gender.



Notes: On the left, the estimates of the association between a 1 SD increase in the educational attainment polygenic score (EA3 PGS) and an increase in cognitive/non-cognitive skills in SDs; on the right, the estimates of the association between a 1 SD increase in cognitive/non-cognitive skills and an increase in educational attainment in years of education. Error bars represent 95% confidence intervals.
Data source: the Wisconsin Longitudinal Study.

3.6.3 Heterogeneity of mediation by socio-economic status

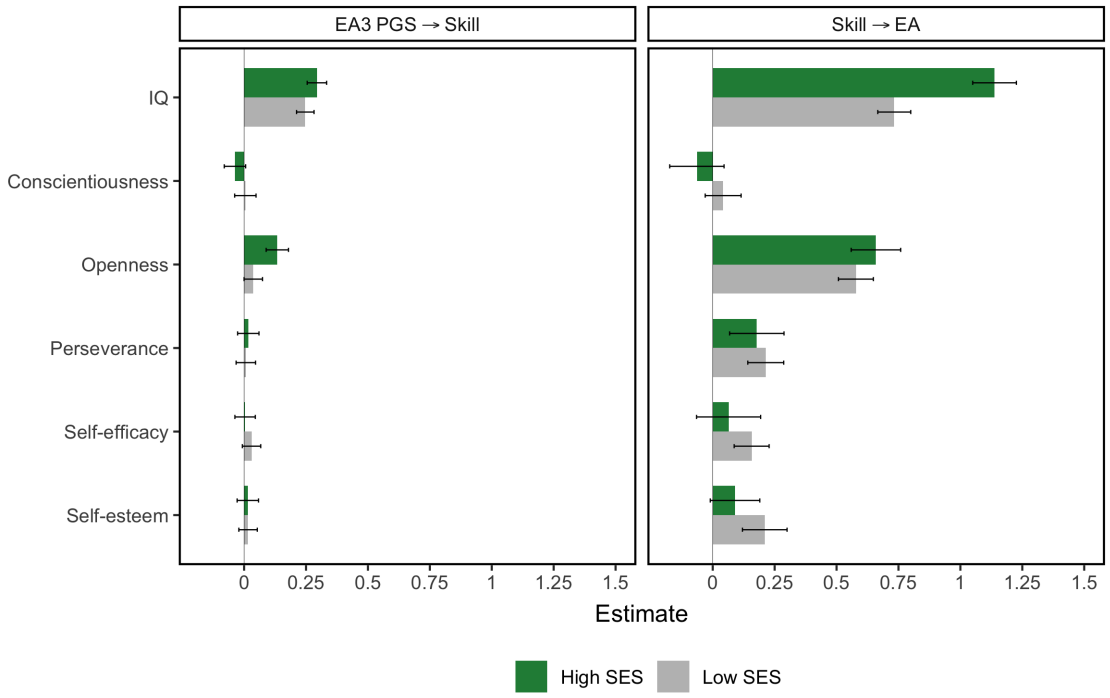
Finally, I run the analyses separately for respondents with high SES (with fathers who have at least a high school diploma) and with low SES (with father who do not have a high school diploma). I expect cognitive and non-cognitive skills to be stronger mediators for lower-SES individuals compared to higher-SES respondents (Hypothesis 4).

As can be seen in Figure 3.6, the educational attainment polygenic score is much more predictive of cognitive and non-cognitive skills for respondents from high SES backgrounds. Unlike in the context of gender, there is no difference in the SES pattern between cognitive and non-cognitive skills.

Cognitive and non-cognitive skills also explain our outcome of interest — educational attainment — better in higher-SES contexts (Figure 3.6). However, the disparity between two groups is higher for cognitive skills than for non-cognitive ones. The difference is very stark in the case of IQ: while for high SES individuals a 1 standard deviation increase in IQ predicts a 1.1 year increase in educational attainment, the same increase in IQ predicts only a 0.73 year increase for respondents with lower SES.

The full mediation model by SES can be seen in Table 3.A.8 in the Appendix. Cognitive and non-cognitive skills consistently play a more major mediating role for higher-SES individuals. For instance, IQ mediates 53.2% of the association between the polygenic score and educational attainment for high-SES respondents and 44.5% for low-SES respondents. Openness also mediates 14% compared to 8.6% for high- and low-SES individuals respectively. As predicted by our Hypothesis 4, both cognitive and non-cognitive skills explain the genetic predisposition to educational attainment better for higher-SES than for lower-SES individuals.

Figure 3.6: Estimated coefficients of the associations in the mediation model moderated by SES.



Notes: On the left, the estimates of the association between a 1 SD increase in the educational attainment polygenic score (EA3 PGS) and an increase in cognitive/non-cognitive skills in SDs; on the right, the estimates of the association between a 1 SD increase in cognitive/non-cognitive skills and an increase in educational attainment in years of education. Error bars represent 95% confidence intervals.
Data source: the Wisconsin Longitudinal Study.

3.7 Discussion/Conclusion

Educational attainment is a complex social process which is shaped by an array of structural and individual characteristics. Novel scientific methods have established that some of these characteristics are genetic. However, little attention so far has been paid to understanding the ways in which these genetic mechanisms operate. The pathways between one's genetic potential and realised educational levels are likely marked by many intermediate factors. In this paper I focused on cognitive and non-cognitive skills as potential factors to play a role in the association between one's genotype and years of education. I found that the mediating role of individual non-cognitive skills, namely conscientiousness, openness, perseverance, self-efficacy, and self-esteem, is much weaker than that of IQ. However, together these five non-cognitive skills could still explain around 15%

of the association of interest (compared to 44% by IQ). Openness is the most important mediator among non-cognitive skills: it predicts educational attainment and is predicted by the educational attainment polygenic score. In contrast to the literature claiming that conscientiousness is the best predictors of educational attainment among all non-cognitive skills, in our sample this measures was not at all associated with years of education. The finding of openness being an important mediator and its estimated magnitude is in line with previous research (Belsky et al., 2016; Smith-Whooley et al., 2019). Open individuals enjoy learning, are inquisitive, often thrive in academic environments and, thus, are more likely to, first, feel comfortable at school, then earn good grades, and, finally, have the academic results prerequisite for attending college (Borghans et al., 2008).

I also go beyond existing literature by investigating how the ways in which the association between genetics and education plays our are contingent on social contexts. Our logic to suspect such mediation differences is twofold. First, it draws on the literature showing that institutional, social, cultural, and financial barriers can constrain one's genetic potential (Herd et al., 2019; Turkheimer et al., 2003; Rowe et al., 1999; Scarr-Salapatek, 1971). Second, it takes into account research on gender differences in personality and non-cognitive skills (Maccoby and Jacklin, 1978). In this paper, I documented substantial differences in the genetic transmission process by gender and SES. For women in the WLS, non-cognitive skills are much better at explaining their genetic predisposition to educational attainment whereas for men the role of cognitive skills is much more pronounced. Why would the polygenic score, contrary to our expectations, be better at explaining non-cognitive skills for women than for men? It is possible that boys and girls, equally genetically predisposed to education, have been socialised differently — with boys encouraged to develop their cognitive potential and girls - the non-cognitive one. In the context of SES, however, both non-cognitive and cognitive skills seem to be better at explaining years of education for more privileged individuals compared to those from more modest backgrounds. Here the suppression mechanism seems to be at play: in environments with lower levels social, cultural, or economic capital, talents can be harder to manifest (Turkheimer et al., 2003). In the context of non-cognitive skills, this could mean a lack of opportunities to cultivate non-cognitive skills and, even if those skills are realised, barriers to translate such skills into tangible educational benefits. Overall, our results demonstrate the importance of going beyond IQ and of taking into account social contexts if we

want to truly grasp the mechanisms underlying one's propensity to education.

The research design of this paper, however, has inherent limitations and the presented results should be interpreted with several caveats in mind. The models included each cognitive or non-cognitive skill separately and did not explore their interconnections. For example, openness and IQ have been shown to have a similar genetic architecture and, as a consequence, the mediating role of openness might be driven by this genetic correlation. In general, the associations between our measures could be much more complicated with multiple interdependent endophenotypes acting together. Moreover, I did not explore the complementarity of cognitive and non-cognitive skills. Non-cognitive skills, even if not so strong as mediators, might be moderating the association between IQ and years of education (Cunha and Heckman, 2007).

The generally weak associations between the educational attainment polygenic score and non-cognitive skills found in this paper do not mean that non-cognitive skills are not transmitted genetically. Genetic determinants of educational attainment and IQ might simply be very different from those of non-cognitive skills. Well-powered GWAS of non-cognitive skills should be conducted in the future to clarify this assumption. Because non-cognitive skills are distinct from one another, genetic architecture of each specific skills should be studied. Genetic SEM could be used to investigate the concept of openness and how it genetically relates to IQ or the educational attainment polygenic score. The polygenic score should also not be understood as the ultimate genetic predictor of educational attainment. As we have seen, although powerful, the polygenic score prediction is still nowhere close to the one suggested by a very high family heritability of education. The educational attainment polygenic score might not incorporate rare variants which could potentially be associated with non-cognitive skills. The research designs also lacks essential information on parental genotypes and, hence, I could not test the effect of non-transmitted alleles which are related to years of education (Kong et al., 2018b).

The results also do not mean that non-cognitive skills other than openness do not matter for educational success. Whereas cognitive skills are predictive in a greater variety of tasks, non-cognitive skills are more predictive in more specific contexts (Borghans et al., 2016). There have been attempts to extract a factor similar to g for non-cognitive skills but even the simplest frameworks, such

as the Big 5, have multiple components. In general, non-cognitive skills have been demonstrated to be the most important at the ends of distributions: the effects of moving from the bottom to the top decile in personality are comparable to those of cognitive skills even if the variance explained by non-cognitive skills is lower (Borghans et al., 2008). Future papers should further investigate the extremes in the distribution of non-cognitive skills.

Any study of non-cognitive skills is fraught with measurement issues since these skills have inherently "noisy" measures (Groves, 2005). Self-reported personality traits and skills are known to be biased (Paulhus et al., 1998). Unfortunately, multi-respondent longitudinal data, which could help mitigate self-reporting biases, are not available in the WLS. Furthermore, a distinction between different non-cognitive skills is often arbitrary. For example, perseverance and conscientiousness are conceptually very similar and it is difficult to decide which measures found in the dataset better represent one concept than another (Rimfeld et al., 2016). The dichotomy of cognitive and non-cognitive skills is also somewhat artificial and their conceptual and empirical separation should be investigated (Borghans et al., 2008). To illustrate, conscientiousness is very important when taking IQ tests, whose results then reflect both cognitive and non-cognitive attributes (Segal, 2008). Openness and perseverance also influence the evolution of cognitive ability over the life cycle (Borghans et al., 2008). Moreover, many skills, such as emotional intelligence, cognitive style, intellectual engagment, or practical intelligence, are "quasi-cognitive" and, thus, are difficult to place in one of the binary baskets of cognitive or non-cognitive skills (Borghans et al., 2008).

The fact that non-cognitive skills are measured late in life in the WLS while cognitive skills are measured in high school is also worth mentioning. Our cross-sectional data cannot address the possibility of endogeneity: non-cognitive skills can be influenced by one's educational attainment as much as they themselves influence years of education (Heckman and Rubinstein, 2001). For example, a high level of self-efficacy later in life can be the product of high levels of schooling and subsequent prestigious employment. Although the assumption made in this paper about personality being more or less stable over time starting in adolescence has a lot of support in the literature (Roberts and DelVecchio, 2000; Briley and Tucker-Drob, 2014b), longitudinal designs should be used in the future to further investigate this claim.

Finally, several characteristics of the WLS need to be acknowledged as they limit the extent to which these results can be extrapolated to wider populations. First, the entire sample in the WLS is of European ancestry. More generally, genome-wide association studies do not pay enough attention to non-European individuals. As a consequence, polygenic scores produced using those studies are not as predictive in non-white samples (Mills and Rahal, 2019a). The results in this paper, thus, should be approached with caution and not generalised to non-European groups. Second, with regards to education, the WLS has a left-truncated sample: because of how respondents were sampled, nobody has less than a high-school diploma. Such decreased variation in the outcome values we observe might lead to an underestimate of the effect of the EA3 PGS or non-cognitive skills, which seem to matter most at the distributional ends (Heckman et al., 2006). The types of non-cognitive skills which matter for graduating from high school versus not graduating might also be different than those predicting going beyond high school once high school is completed (Carneiro et al., 2006). Third, since the WLS is a rather old cohort, the associations found here between the EA3 PGS and education or between non-cognitive skills and education might not be the same for younger samples. In other words, the mechanisms producing educational inequalities might change from one cohort to another and, thus, the magnitudes of relationships between measures might not be replicated using a different dataset. However, in the words of Domingue and Fletcher (2020, p. 11), the limitations of the WLS are "countered by the recognition that very few datasets contain the information we require for our analysis".

Overall, this paper makes several contributions to the existing literature. I utilise novel scientific tools to study a sociological topic of the transmission of educational advantages and go beyond what has previously been done. Instead of only looking at direct associations between one's genotype and educational attainment, I investigate the pathways underlying these associations and find that, in addition to cognitive skills, openness is also an important mediator in this context. By exploring heterogeneity by gender and SES I also highlight the importance of looking at the effects of one's genotype on social outcomes through a sociocontextual lens (Herd et al., 2019), because not only biology plays a role in social phenomena but social factors also shape one's biology.

3.A Appendix

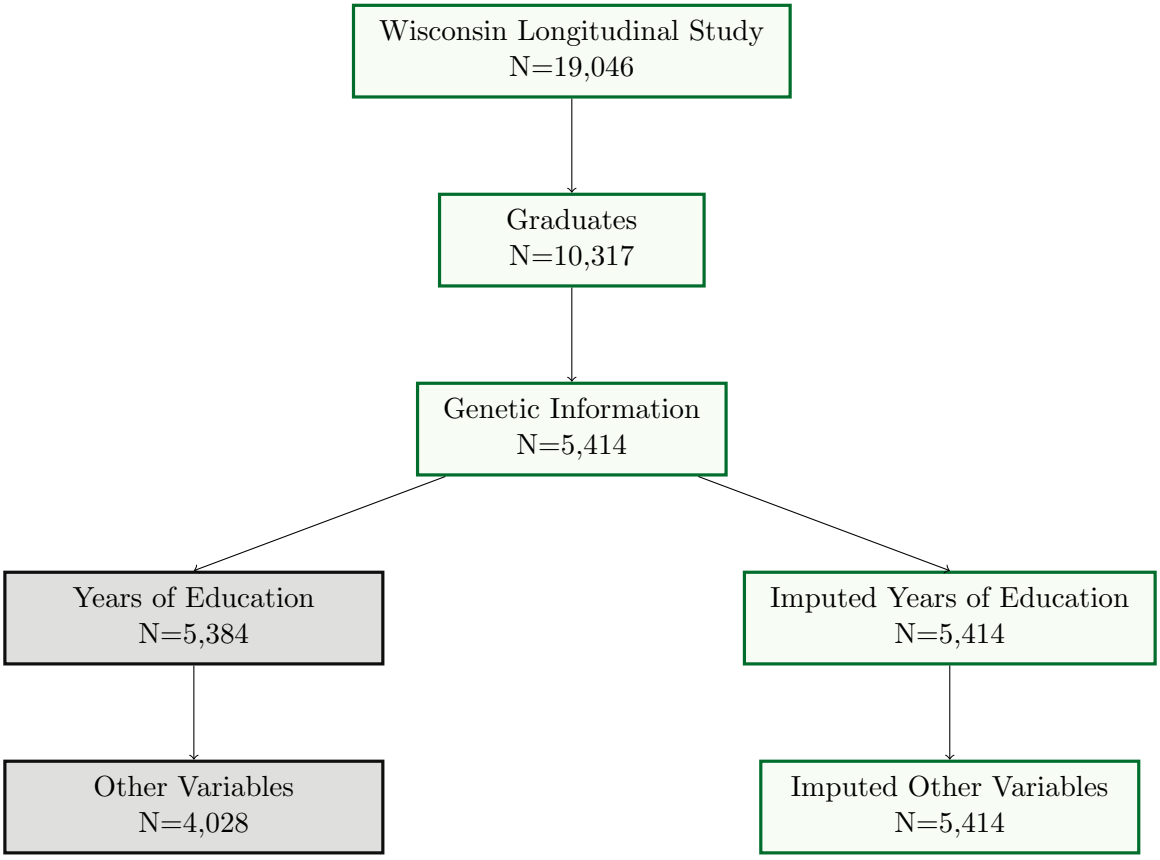
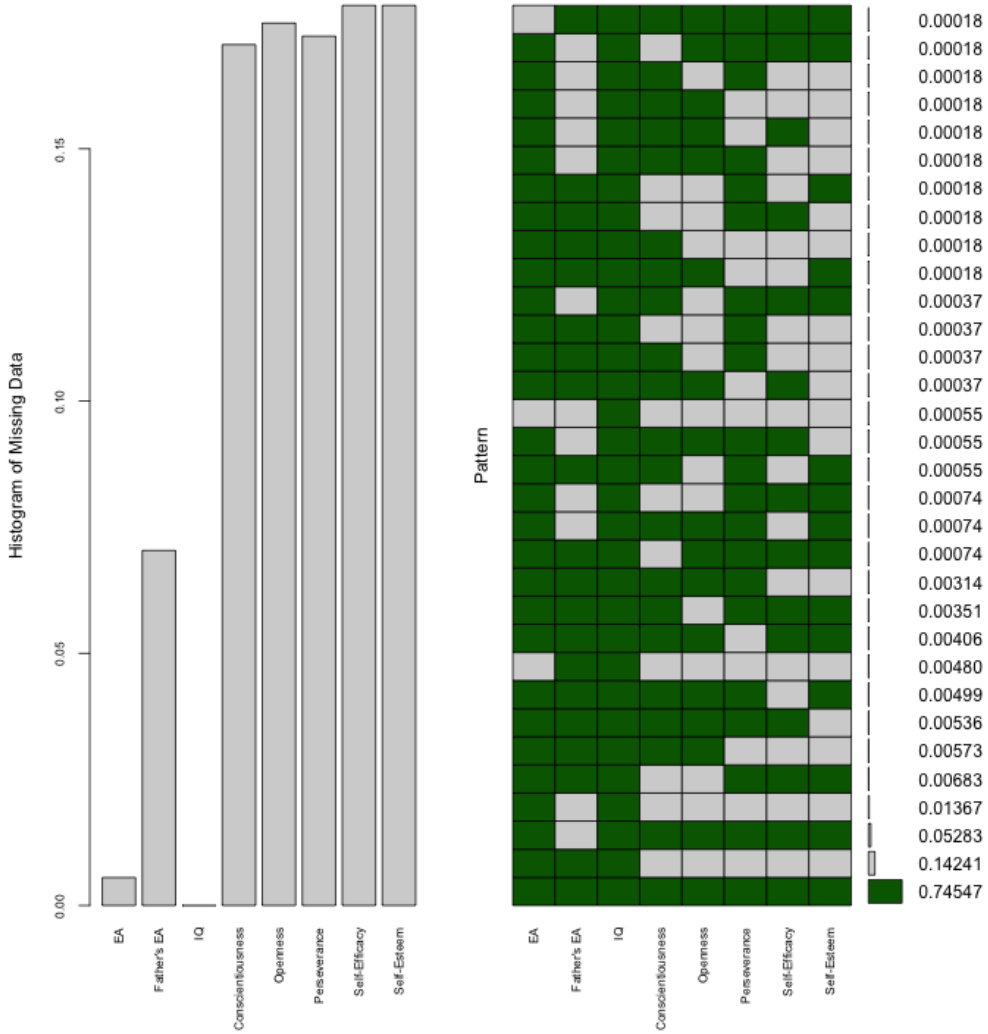


Figure 3.A.1: Sample reduction before (on the left) and after imputation (on the right, used in the analyses).

Figure 3.A.2: Pattern of missing data in the WLS.



Notes: On the left, proportions of missing data in each variable are displayed. On the right, combinations of different variables and their missingness (grey) or non-missingness (dark green) are shown. For example, the bottom entirely green row shows that around 75% of the respondents do not have any missing data or the row above it that around 14% respondents have missing data on father's education and all non-cognitive measures.

Description of imputation.

The method of multiple imputation was used to impute missing data. The pattern and the frequency of missing data for each variable can be found in Figure 3.A.2 in the Appendix. For most measures of non-cognitive skills around 16-18% of the data is missing. For other variables either no data is missing (for example, IQ), or the rate of missing data is very low (for example, educational attainment). In such situations multivariate imputation by chained equations (MICE), also known as fully conditional specification (FCS), is considered to be one of the best imputation methods (van Buuren and Groothuis-Oudshoorn, 2011).

The process of multiple imputation was implemented using the MICE package in R (van Buuren and Groothuis-Oudshoorn, 2011). MICE uses the Markov Chain Monte Carlo (MCMC) method which exploits the correlation structure of the data and performs imputation for each variable with missing data by iteratively regressing incomplete variables on other explanatory factors (van Buuren, 2018). "Cart", denoting classification and regression trees, was chosen as the imputation method because its non-parametric form allows for complex relationships between variables of different types (numeric and non-numeric) without requiring to specify many different interaction and non-linear effects (Burgette and Reiter, 2010). The number of imputations (the number of new datasets to be created) was specified as 30. While the default of 5 imputations might often be deemed sufficient on efficiency grounds (Rubin, 1987), having more imputations is considered to be better and 30 is often considered to be optimal (Bodner, 2008) The number of iterations for each imputation was set to 100: a rather high number of 100 ensures convergence yet it does not require excessive computational power and, consequently, does not take a long time to run (van Buuren, 2018). In addition to the main mediation results reported in this paper, in the Appendix I also include results produced by running the models on an unimputed sample (Table 3.A.5).

Assumptions of missing completely at random (MCAR) and missing at random (MAR) should be investigated to understand the potential causes of missingness in our data. In the case of MCAR, missingness should not be associated with the data: there should be no systematic relationship between a missing data point and values on other measures for the same observation. In the case of MAR, however, the propensity of missingness can be associated with other variables, or,

in other words, observed data. For example, if the missingness for the conscientiousness variable is conditional on the polygenic score (respondents with lower polygenic scores are more likely to have missing data on conscientiousness than those with higher polygenic scores), then the data can be said to be MAR and not MCAR. Figure 3.A.3 in the Appendix shows that, indeed, for all non-cognitive skills respondents with missing values on those skills had on average a slightly lower polygenic score. This conclusion is also confirmed in Table 3.A.1 in the Appendix which displays results from the t-tests comparing means of the polygenic score and years of education between respondents with missing and non-missing data on each measure of non-cognitive skills. Individuals with lower levels of educational attainment and the polygenic score are more likely to have missing data on non-cognitive measures. If we assume that our data is MAR, multiple imputation can be performed. However, we cannot test an assumption of data missing not at random (MNAR). This situation would amount to the missingness being associated not with observed data (as in the case of MAR) but with unobserved data or missing values. The cause of missingness in that case could never be known to us. For example, graduates with the lowest conscientiousness scores might not be answering questions about non-cognitive skills. The only way to deal with such situation would be to do follow-up interviews with the respondents who have missing data. Since the WLS did not conduct such follow-up interviews, for the sake of our imputation and further analyses, we will have to assume that our data is either MAR or MCAR.

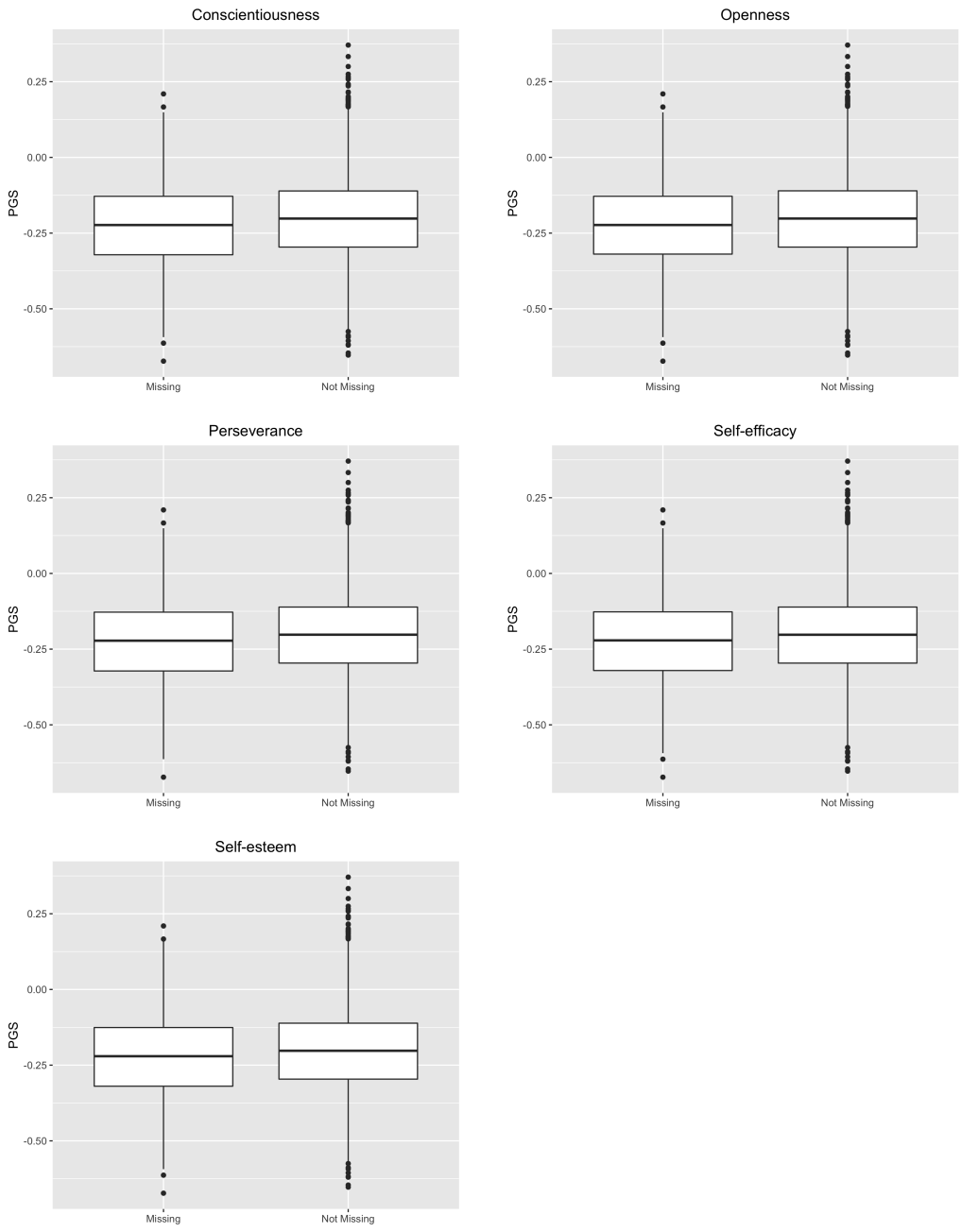


Figure 3.A.3: Boxplots comparing the polygenic score distributions between respondents with missing and non-missing data on each non-cognitive skill item in the WLS.

	Difference in EA3 PGS					*Difference in EA*				
	Mean	95% CI	t	df	p-value	Mean	95% CI	t	df	p-value
Conscientiousness	-0.022	-0.032 -0.012	-4.306	1318.5	1.789e-05	-0.469	-0.614 -0.323	-6.328	1413.3	3.320e-10
Openness	-0.022	-0.032 -0.012	-4.317	1376	1.692e-05	-0.465	-0.609 -0.321	-6.328	1462.7	3.302e-10
Perseverance	-0.021	-0.032 -0.012	-4.305	1318.5	1.784e-05	-0.420	-0.564 -0.275	-5.847	1333.3	3.408e-05
Self-efficacy	-0.023	-0.033 -0.012	-4.328	1272.5	1.622e-05	-0.440	-0.587 -0.292	-5.845	1354	6.343e-09
Self-Esteem	-0.021	-0.031 -0.011	-4.155	1323.1	3.464e-05	-0.419	-0.565 -0.272	-5.596	1401.5	2.641e-08

Table 3.A.1: T-test for the difference in the polygenic score and educational attainment between respondents with missing and non-missing data on each non-cognitive skill measure in the WLS.

Measure	Dataset question
Conscientiousness	*1. To what extent do you agree that you see yourself as someone who does a thorough job?* *2. To what extent do you agree that you see yourself as someone who is a reliable worker?* *3. To what extent do you agree that you see yourself as someone who tends to be disorganized?* *4. To what extent do you agree that you see yourself as someone who is lazy at times?* *5. To what extent do you agree that you see yourself as someone who does things efficiently?* *6. To what extent do you agree that you see yourself as someone who is easily distracted?*
Openness	*1. To what extent do you agree that you see yourself as someone who prefers the conventional, traditional?* *2. To what extent do you agree that you see yourself as someone who prefers work that is routine and simple?* *3. To what extent do you agree that you see yourself as someone who values artistic, aesthetic experiences?* *4. To what extent do you agree that you see yourself as someone who has an active imagination?* *5. To what extent do you agree that you see yourself as someone who wants things to be simple and clear-cut?* *6. To what extent do you agree that you see yourself as someone who is sophisticated in art, music, or literature?*
Perseverance	*1. To what extent do you agree that even when things seem hopeless, you keep on fighting to reach your goal* *2. To what extent do you agree that you stick to your goals and projects even in the face of great difficulties?* *3. To what extent do you agree that the harder a goal is to achieve, the more appeal it has to you?* *4. To what extent do you agree that you can be very stubborn in pursuing your goals?*
Self-efficacy	*1. To what extent do you agree that you are good at juggling your time so that you can fit everything in that needs to get done?* *2. To what extent do you agree that you often feel overwhelmed by your responsibilities?* *3. To what extent do you agree that you are quite good at managing the many responsibilities of your daily life?* *4. To what extent do you agree that you do not fit very well with the people and community around you?* *5. To what extent do you agree that you have difficulty arranging your life in a way that is satisfying to you?*
Self-esteem	*1. To what extent do you agree that, in general, you feel confident and positive about yourself?* *2. To what extent do you agree that when you compare yourself to friends and acquaintances, it makes you feel good about who you are?* *3. To what extent do you agree that your attitude about yourself is probably not as positive as most people feel about themselves?*

Table 3.A.2: WLS items for non-cognitive measures.

	Unimputed					*Imputed*				
	Mean	Median	Range	SD	N	Mean	Median	Range	SD	N
Education	13.55	12.00	12.00 - 20.00	2.23	5,384	13.55	12.00	12.00 - 20.00	2.23	5,414
Father's Education	10.87	10.00	8.00 - 19.00	2.90	5,033	10.85	10.00	8.00 - 19.00	2.89	5,414
Gender										5,414
Female	0.52				2,830	0.52				2,830
Male	0.48				2,584	0.48				2,584
IQ	0.12	0.10	-2.65 - 2.99	1.00	5,414	0	-0.01	-2.80 - 2.91	1.00	5,414
EA3 PGS	0	-0.01	-3.32 - 4.06	1.00	5,414	0	-0.01	-3.32 - 4.06	1.00	5,414
Conscientiousness	0	0.19	-4.66 - 1.64	1.00	4,490	0	0.11	-4.69 - 1.66	1.00	5,414
Openness	0	0.03	-3.30 - 2.94	1.00	4,467	0	0.03	-3.33 - 2.96	1.00	5,414
Perseverance	0	0.07	-4.24 - 2.03	1.00	4,461	0	0.06	-4.27 - 2.01	1.00	5,414
Self-Efficacy	0	0.06	-4.70 - 1.57	1.00	4,463	0	0.03	-4.66 - 1.52	1.00	5,414
Self-Esteem	0	0.03	-4.41 - 1.14	1.00	4,488	0	0.28	-4.44 - 1.08	1.00	5,414

Table 3.A.3: Comparing descriptive statistics of standardised variables between unimputed and imputed samples using the WLS.

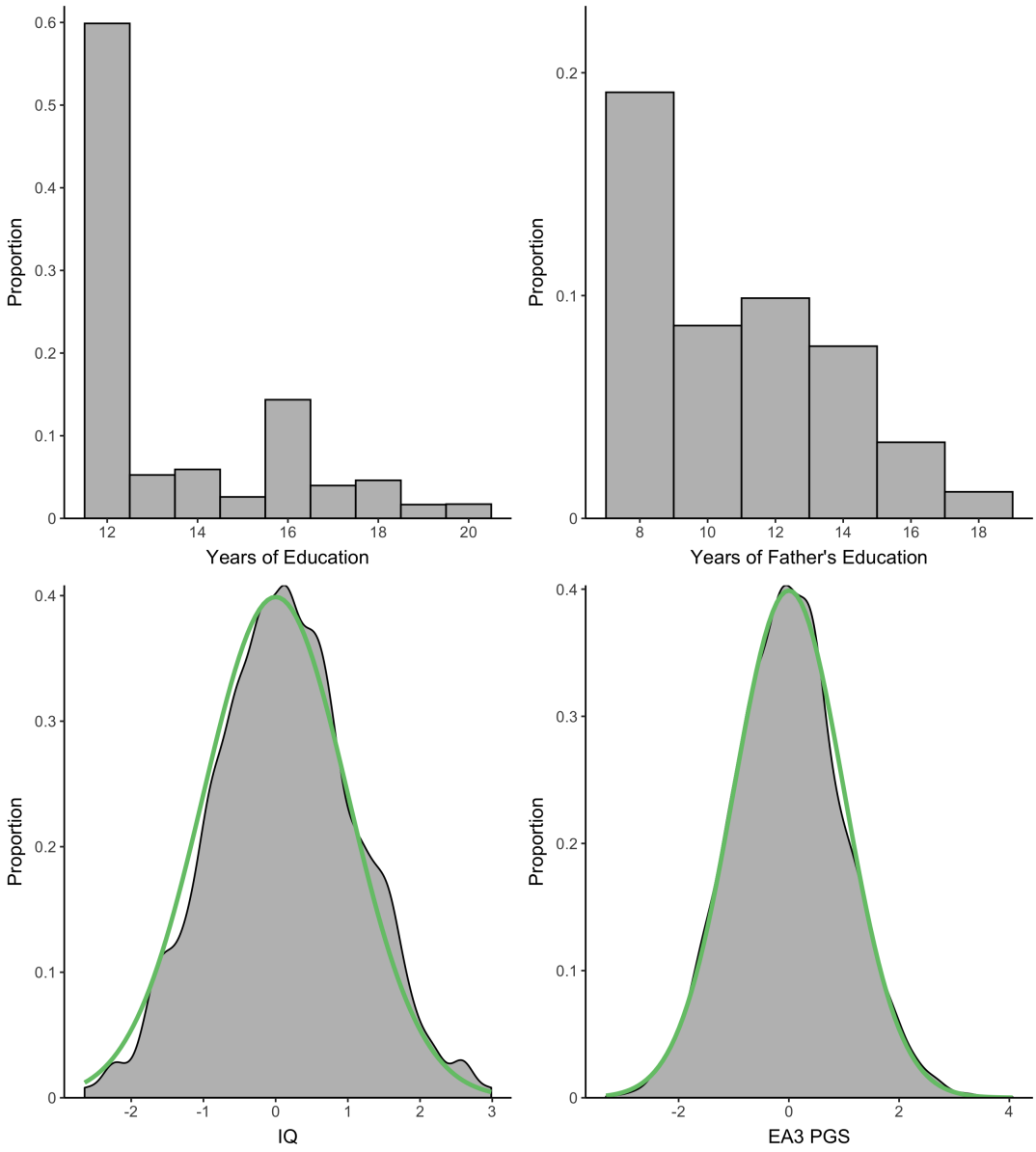


Figure 3.A.4: Histograms for respondent's education and father's education and kernel density plots for standardised IQ and standardised polygenic score in the WLS.

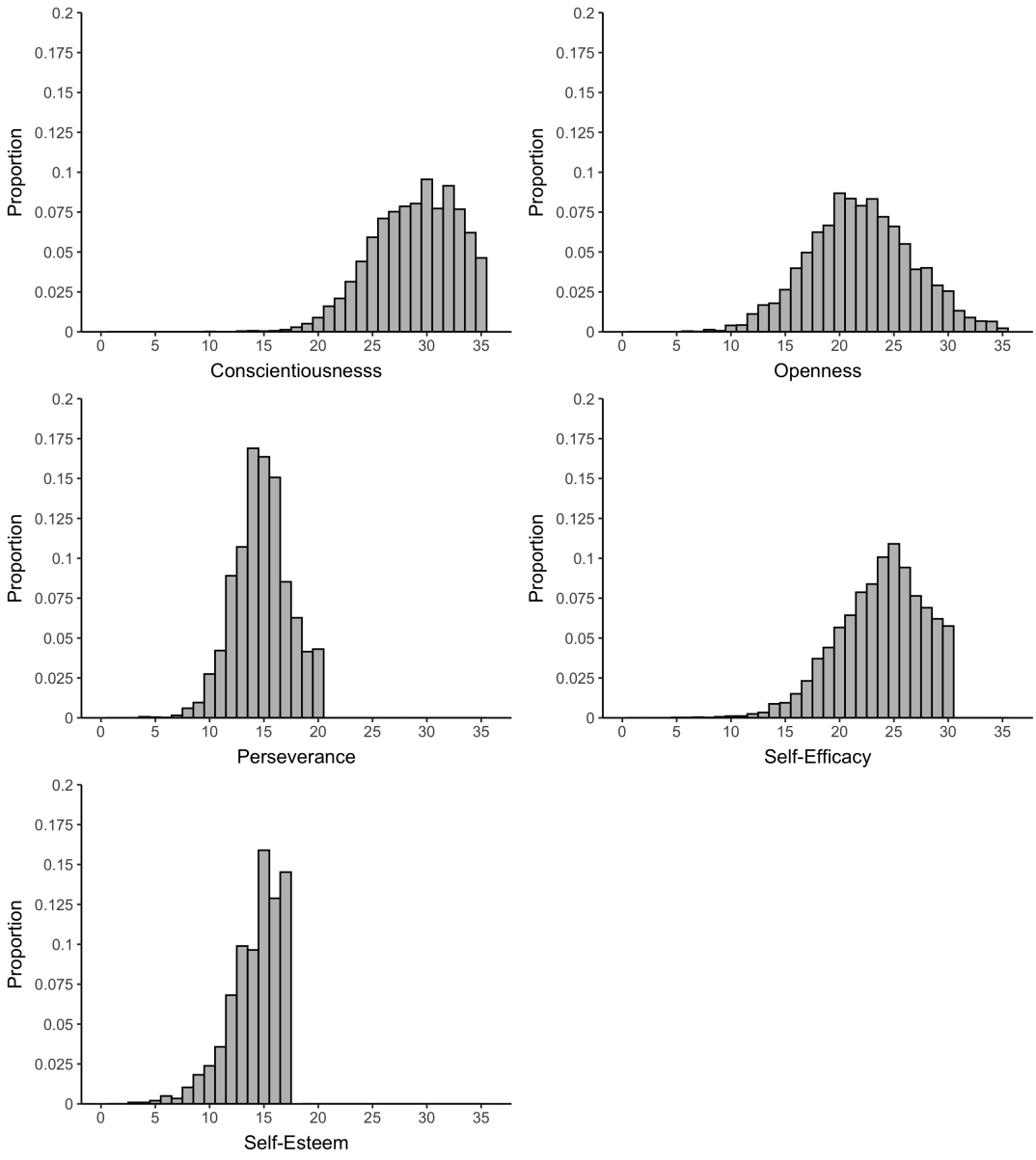
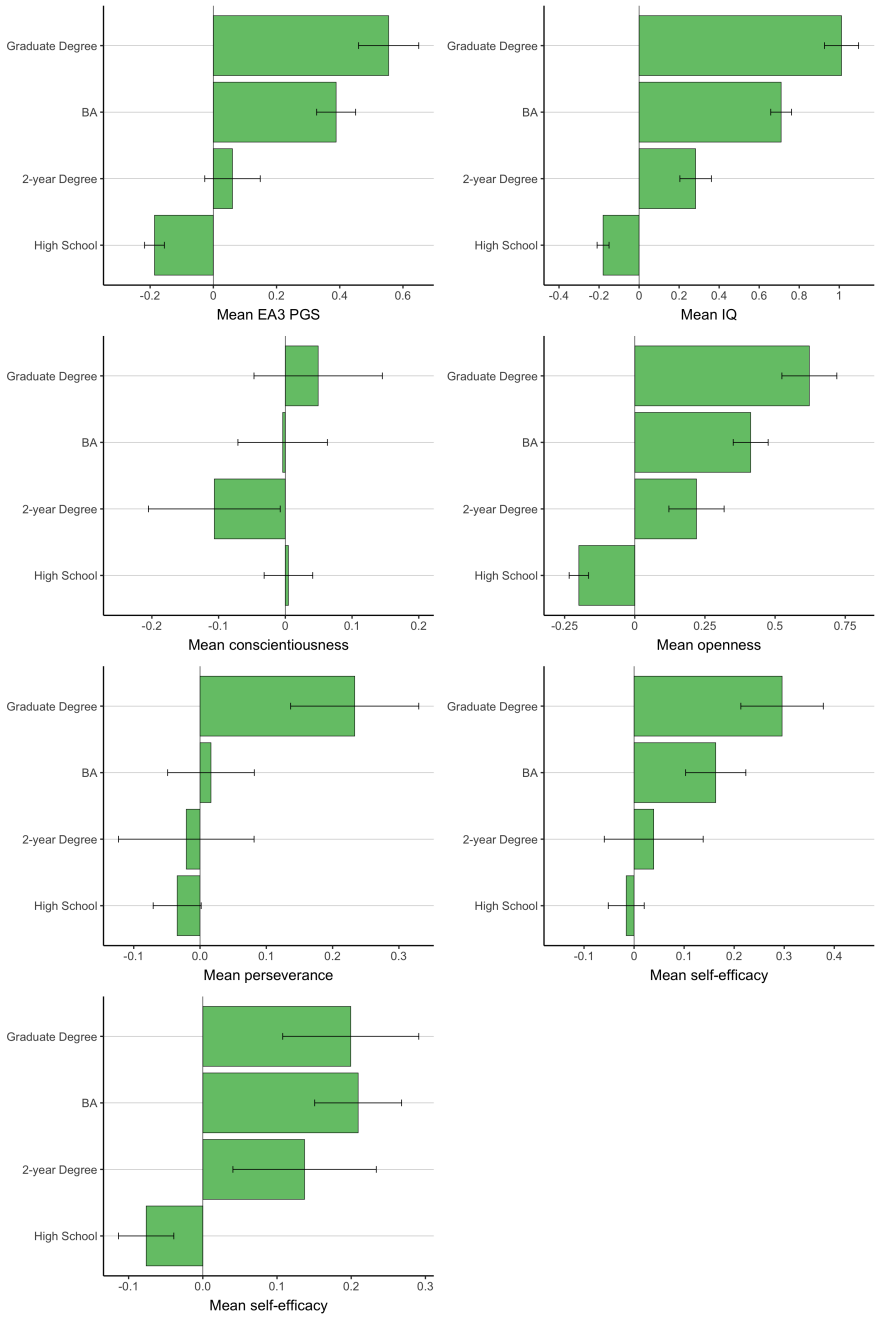


Figure 3.A.5: Histograms of unstandardised measures of non-cognitive skills in the WLS.

	IQ	Conscient-iousness	Openness (overall)	Openness (intellectualism)	Openness (adventurousness)	Perseverance	Self-efficacy	Self-esteem
IQ	1.000							
	(0.000)							
Conscientiousness	-0.006	1.000						
	(0.012)	(0.000)						
Openness (overall)	0.253	0.124	1.000					
	(0.011)	(0.012)	(0.000)					
Openness (intellectualism)	0.146	0.033	0.770	1.000				
	(0.012)	(0.012)	(0.005)	(0.000)				
Openness (adventurousness)	0.239	0.162	0.779	0.196	1.000			
	(0.012)	(0.012)	(0.005)	(0.012)	(0.000)			
Perseverance	0.025	0.295	0.297	0.249	0.218	1.000		
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.000)		
Self-efficacy	0.056	0.480	0.200	0.106	0.207	0.302	1.000	
	(0.012)	(0.009)	(0.012)	(0.012)	(0.012)	(0.011)	(0.000)	
Self-esteem	0.072	0.359	0.289	0.214	0.241	0.400	0.638	1.000
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.010)	(0.007)	(0.000)

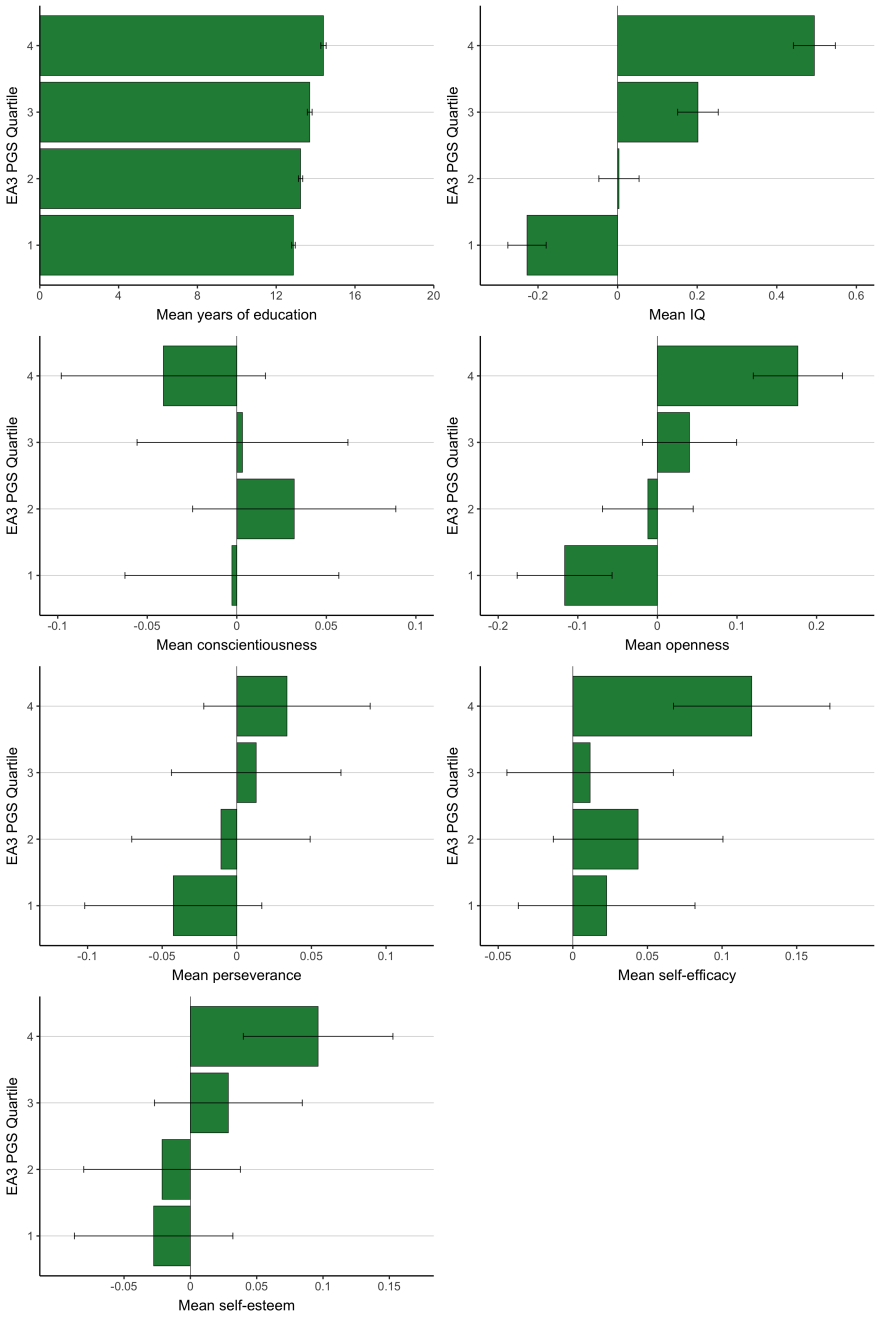
Table 3.A.4: Correlations between individual-level explanatory factors.

Figure 3.A.6: Associations between a categorical measure of educational attainment and the main explanatory variables in the WLS.



Notes: Educational attainment and cognitive/non-cognitive skills are positively related: the higher the educational category attained by an individual, the higher on average her cognitive and non-cognitive skills.

Figure 3.A.7: Associations between EA3 PGS quintiles and other variables in the WLS.



Notes: The educational attainment polygenic score is positively associated with years of education, IQ, openness, and self-esteem.

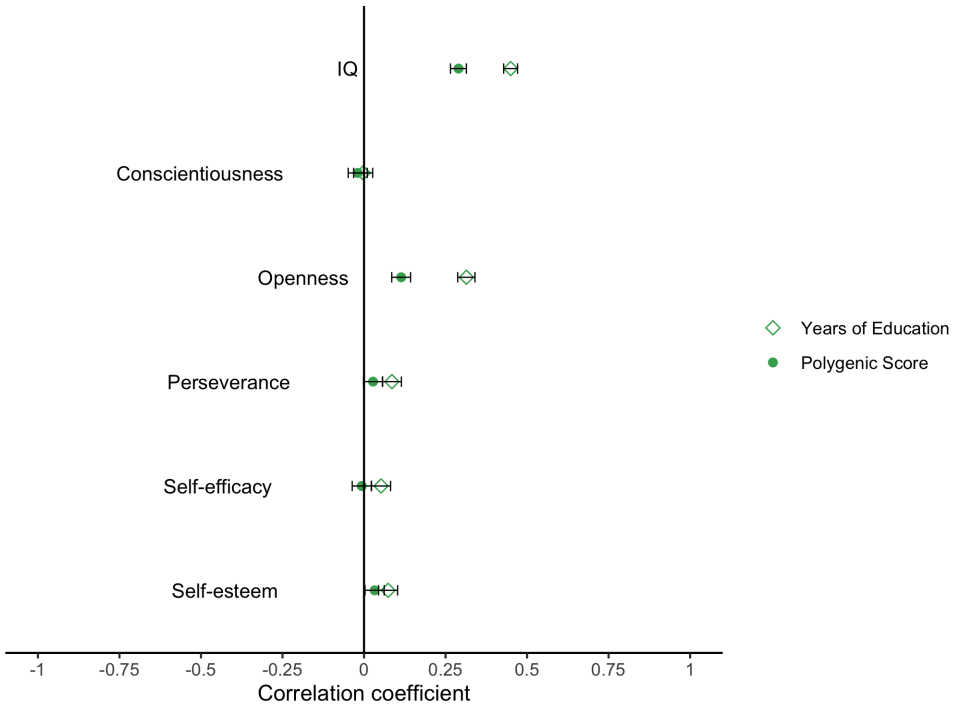


Figure 3.A.8: Pearson's correlations between cognitive/non-cognitive skills and the polygenic score/years of education in the WLS.

IQ is rather strongly correlated with both the polygenic score (r = 0.3) and educational attainment (r = 0.45). Among five non-cognitive skills, only openness displays moderate correlations with the polygenic score (r = 0.08) and educational attainment (r = 0.23).

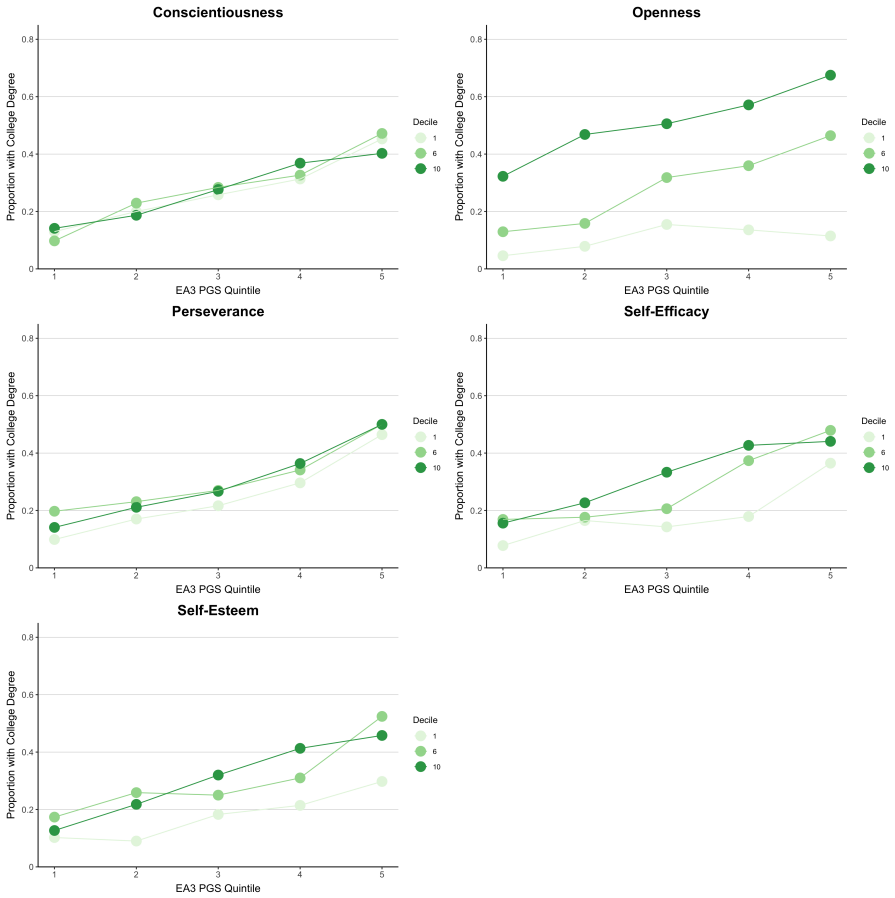


Figure 3.A.9: Proportions with a college degree, by EA3 PGS quintiles and 1st (0-10th percentile), 6th (50-60th percentile), and 10th deciles (90-100th percentile) in non-cognitive skills in the WLS.

The higher the polygenic score quintile, the higher the probability of receiving a college degree. However, some of the non-cognitive skills introduce further heterogeneity in this association. In the case of openness, for example, a high level of openness can compensate for a low level of the educational attainment polygenic score: respondents in the first quintile of the polygenic score who are in the highest decile of the openness measure, are almost as likely to graduate from college as those in the fourth quintile in the polygenic score but in the sixth decile in openness (Figure 3.A.9). Similar conclusions can be drawn for self-efficacy and self-esteem. Conscientiousness, and perseverance, however, do not follow this pattern and, as previously shown in Figure 3.A.8, do not seem to be associated with the educational attainment polygenic score or years of education.

Table 3.A.5: Estimated coefficients of direct and indirect effects: sensitivity analysis.

Direct effects/Total effects	b(SE)	p	
EA3 PGS →IQ	0.263***(0.015)	<0.001	
EA3 PGS →Conscientiousness	-0.018(0.016)	0.268	
EA3 PGS →Openness	0.100***(0.016)	<0.001	
EA3 PGS →Perseverance	0.029(0.016)	0.071	
EA3 PGS →Self-efficacy	0.016(0.015)	0.290	
EA3 PGS →Self-esteem	0.052***(0.016)	<0.001	
EA3 PGS →EA	0.578***(0.035)	<0.001	
IQ →EA	1.032***(0.034)	<0.001	
Conscientiousness →EA	-0.012(0.036)	0.739	
Openness →EA	0.722***(0.035)	<0.001	
Perseverance →EA	0.157***(0.035)	<0.001	
Self-efficacy →EA	0.148***(0.038)	<0.001	
Self-esteem →EA	0.182***(0.037)	<0.001	
Indirect effects	b(SE)	p	% mediated
EA3 PGS →IQ →EA	0.241***(0.015)	<0.001	41.70%
EA3 PGS →Conscientiousness →EA	0	0.750	0%
EA3 PGS →Openness →EA	0.063***(0.010)	<0.001	10.80%
EA3 PGS →Perseverance →EA	0.004(0.002)	0.114	0.69%
EA3 PGS →Self-efficacy →EA	0.004(0.004)	0.270	0.69%
EA3 PGS →Self-esteem →EA	0.009***(0.004)	<0.001	1.55%

Notes: EA3 PGS = Educational attainment polygenic score. EA = Educational Attainment.
†p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Separate analyses were conducted for each mediator.
The EA3 PGS included in the analyses was regressed on the ten genetic ancestry principal components.
Data source: the Wisconsin Longitudinal Study.

Sensitivity analysis was done on an unimputed sample of 4,036 respondents (those for whom all variables were not missing). The estimates produced using this sample are in general similar to those produced in the main analysis after multiple imputation.

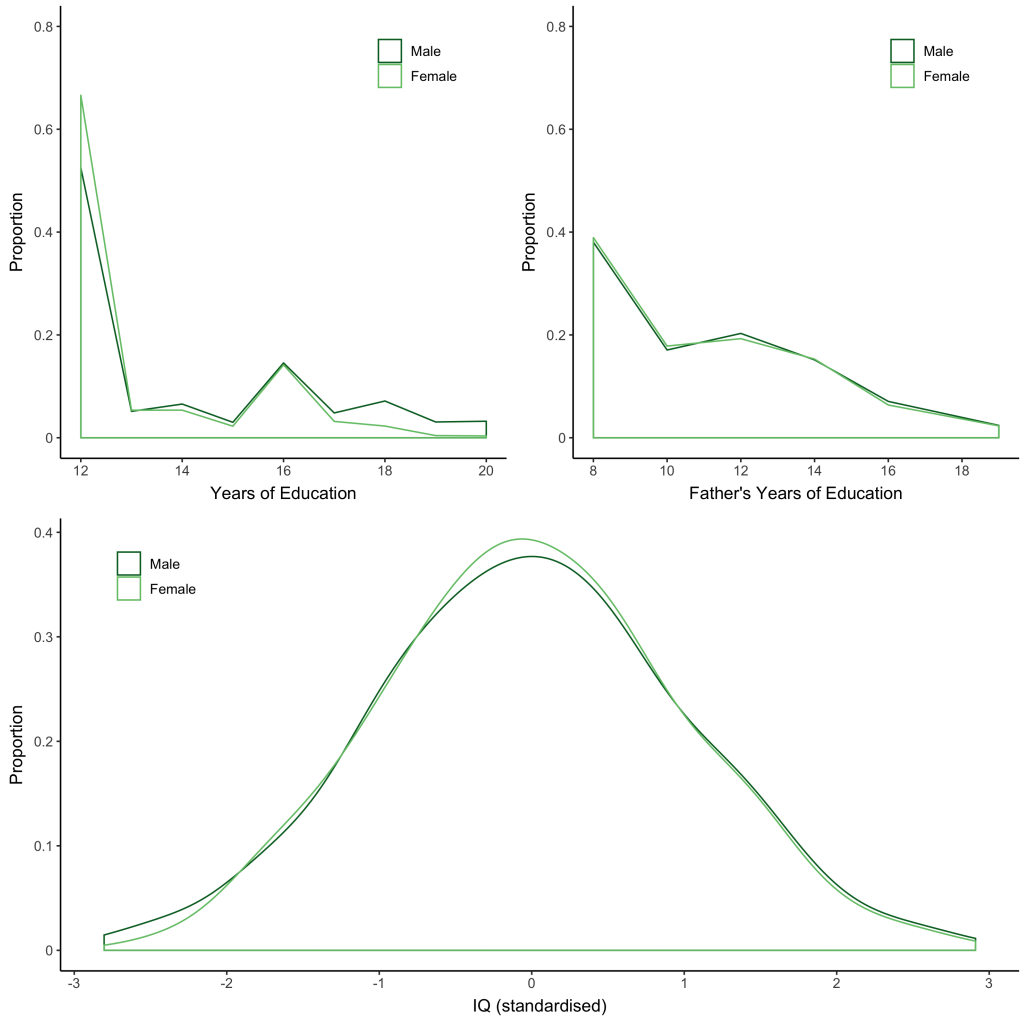


Figure 3.A.10: Distributions of years of education, father's years of education, and IQ by gender in the WLS.

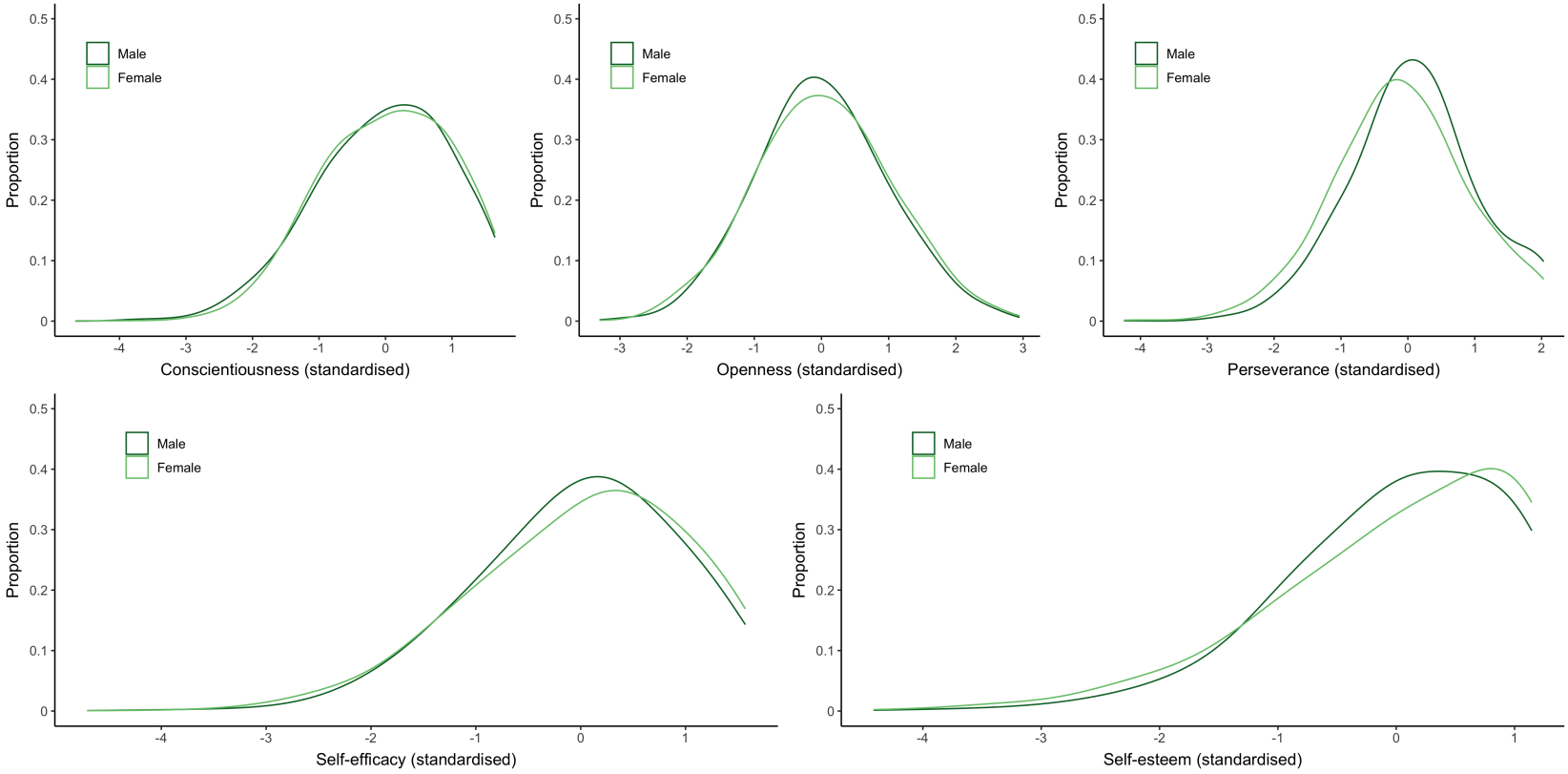


Figure 3.A.11: Distributions of non-cognitive skill measures by gender in the WLS.

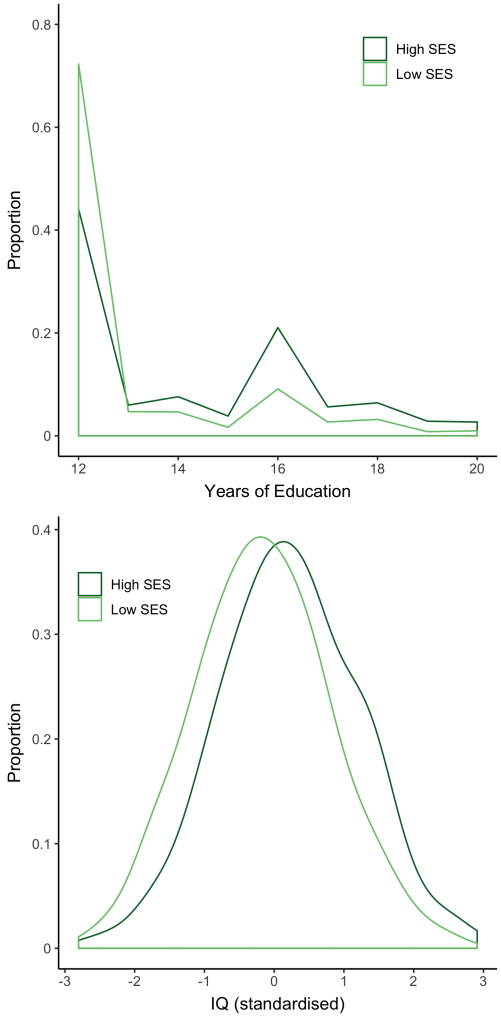


Figure 3.A.12: Distributions of years of education and IQ by SES in the WLS, where low SES means father's education below high school and high SES — high school and above.

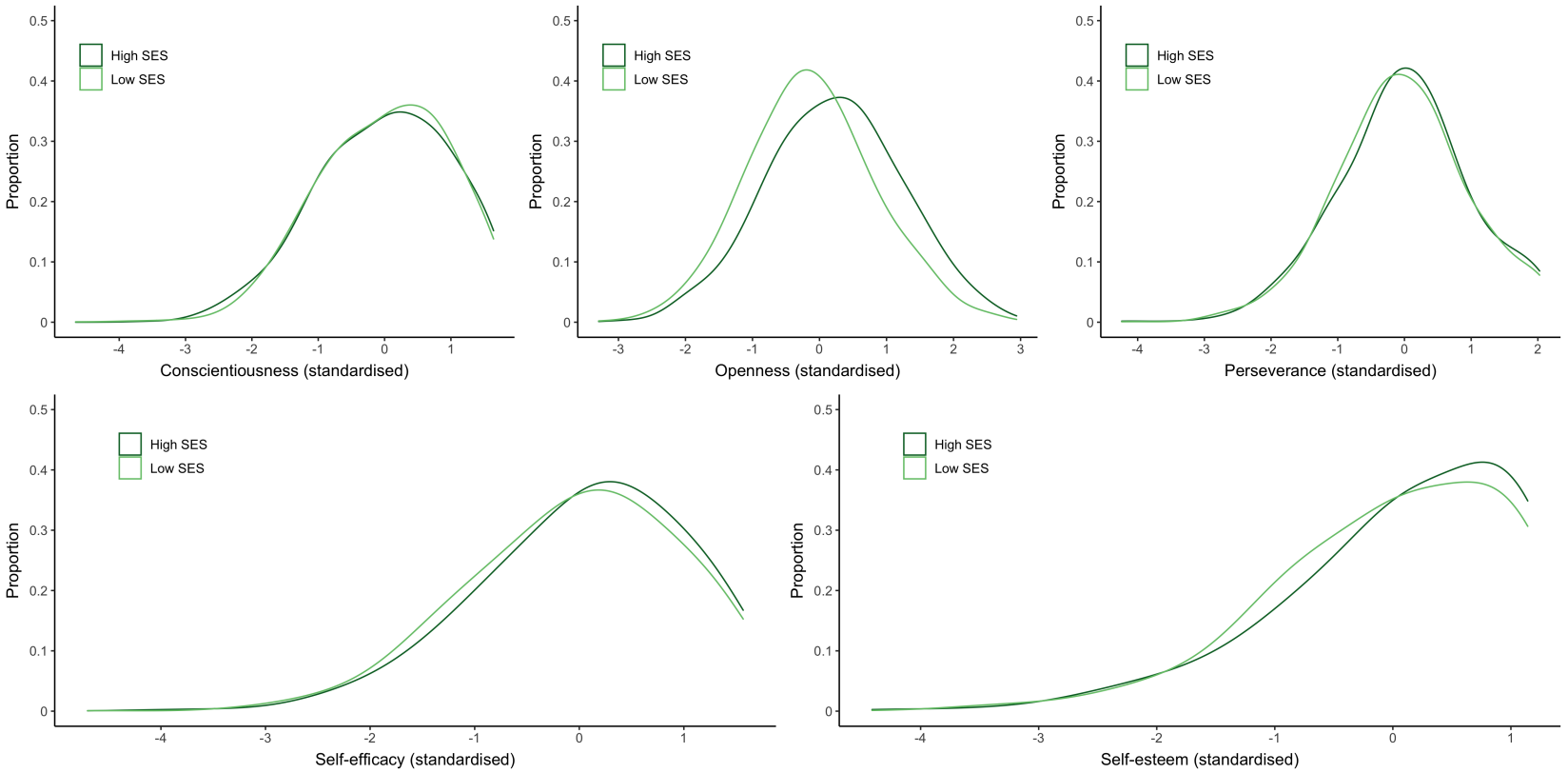


Figure 3.A.13: Comparing distributions of non-cognitive skill measures by SES in the WLS, where low SES means father's education below high school and high SES — high school and above.

Table 3.A.6: Estimated coefficients of direct and indirect effects of openness measures.

	b(SE)	p	% mediated
EA3 PGS →Openness (overall)	0.103***(0.014)	<0.001	
Openness (overall) →EA	0.671***(0.029)	<0.001	
EA3 PGS →Openness (overall) →EA	0.069***(0.003)	<0.001	12.0%
EA3 PGS →Openness (intellectualism)	0.070***(0.015)	<0.001	
Openness (intellectualism) →EA	0.469***(0.031)	<0.001	
EA3 PGS →Openness (intellectualism) →EA	0.032***(0.006)	<0.001	5.60%
EA3 PGS →Openness (adventurousness)	0.086***(0.015)	<0.001	
Openness (adventurousness) →EA	0.648***(0.030)	<0.001	
EA3 PGS →Openness (adventurousness) →EA	0.056***(0.009)	<0.001	9.70%

Notes: EA3 PGS = Educational attainment polygenic score. EA = Educational Attainment.
†p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Separate analyses were conducted for each mediator.
The EA3 PGS included in the analyses was regressed on the ten genetic ancestry principal components.
Data source: the Wisconsin Longitudinal Study.

Table 3.A.7: Estimated coefficients of direct and indirect effects: moderation by gender.

Direct effects/Total effects	Male		Female	
EA3 PGS →IQ	0.292***(0.020)		0.277***(0.017)	
EA3 PGS →Conscientiousness	-0.031(0.027)		0(0.019)	
EA3 PGS →Openness	0.080***(0.021)		0.138***(0.019)	
EA3 PGS →Perseverance	0.010(0.022)		0.058**(0.020)	
EA3 PGS →Self-efficacy	0.033(0.022)		0.017(0.022)	
EA3 PGS →Self-esteem	0.003(0.021)		0.061**(0.022)	
EA3 PGS →EA	0.727***(0.049)		0.441***(0.034)	
IQ →EA	1.249***(0.042)		0.753***(0.033)	
Conscientiousness →EA	-0.007(0.049)		0.003(0.040)	
Openness →EA	0.753***(0.052)		0.656***(0.032)	
Perseverance →EA	0.174**(0.058)		0.138***(0.038)	
Self-efficacy →EA	0.184**(0.058)		0.084***(0.036)	
Self-esteem →EA	0.157**(0.061)		0.182***(0.039)	
Indirect effects	Male	% mediated	Female	% mediated
EA3 PGS →IQ →EA	0.364***(0.002)	50.10%	0.209***(0.001)	47.30%
EA3 PGS →Conscientiousness →EA	-0.001(0.001)	0%	0(0)	0%
EA3 PGS →Openness →EA	0.060***(0.009)	8.25%	0.091***(0.002)	20.60%
EA3 PGS →Perseverance →EA	0.002(0.006)	0%	0.008**(0.002)	1.81%
EA3 PGS →Self-efficacy →EA	0.006(0.006)	0%	0.001(0.002)	0%
EA3 PGS →Self-esteem →EA	0.0005(0.005)	0%	0.011**(0.005)	2.49%

Notes: EA3 PGS = Educational attainment polygenic score. EA = Educational Attainment.
†p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Separate analyses were conducted for each mediator.
The EA3 PGS included in the analyses was regressed on the ten genetic ancestry principal components.
Data source: the Wisconsin Longitudinal Study.

Differences in mediated proportions between males and females in these measures are statistically significant (z-scores and associated p-values are reported in parentheses): IQ (z = 1.328, p-value = 0.092), openness (z = -5.92, p-value = <0.001), self-esteem (z = -3.451, p-value = 0.0003).

Table 3.A.8: Standardised path coefficients, direct and indirect effects: moderation by SES.

Direct effects/Total effects	High SES		Low SES	
EA3 PGS →IQ	0.294***(0.020)		0.247***(0.018)	
EA3 PGS →Conscientiousness	-0.037(0.022)		0.005(0.022)	
EA3 PGS →Openness	0.134***(0.023)		0.061***(0.019)	
EA3 PGS →Perseverance	0.017(0.022)		0.037†(0.022)	
EA3 PGS →Self-efficacy	0.015(0.022)		0.030(0.022)	
EA3 PGS →Self-esteem	0.052*(0.023)		0.016(0.022)	
EA3 PGS →EA	0.630***(0.047)		0.407***(0.036)	
IQ →EA	1.138***(0.045)		0.733***(0.034)	
Conscientiousness →EA	-0.064(0.056)		0.042(0.037)	
Openness →EA	0.659***(0.051)		0.578***(0.036)	
Perseverance →EA	0.178**(0.056)		0.214***(0.042)	
Self-efficacy →EA	0.064***(0.058)		0.157***(0.042)	
Self-esteem →EA	0.090(0.057)		0.210***(0.042)	
Indirect effects	High SES	% mediated	Low SES	% mediated
EA3 PGS →IQ →EA	0.335***(0.020)	53.20%	0.181***(0.013)	44.50%
EA3 PGS →Conscientiousness →EA	0(0)	0%	0(0)	0%
EA3 PGS →Openness →EA	0.088***(0.013)	14.0%	0.035***(0.004)	8.60%
EA3 PGS →Perseverance →EA	0.003(0.004)	0.48%	0.008(0.007)	1.96%
EA3 PGS →Self-efficacy →EA	0.001(0.004)	0%	0.005(0.005)	1.23%
EA3 PGS →Self-esteem →EA	0.005(0.003)	0.79%	0.003(0.004)	0.73%

EA3 PGS = Educational attainment polygenic score. EA = Educational Attainment.
†p<0.1, *p<0.05, **p<0.01, ***p<0.001.
Separate analyses were conducted for each mediator.
The EA3 PGS included in the analyses was regressed on the ten genetic ancestry principal components.
Data source: the Wisconsin Longitudinal Study.

Differences in mediated proportions between high-SES and low-SES respondents in these measures are statistically significant (z-scores and associated p-values are reported in parentheses): IQ (z = 1.790, p-value = 0.037), openness (z = 2.420, p-value = 0.008), perseverance (z = -1.841, p-value = 0.033), self-esteem (z = 2.330, p-value = 0.010).

	1st tercile			2nd tercile			3rd tercile		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
IQ	0.371*	0.349*	0.331*	0.839***	0.755**	0.592**	0.880***	0.835***	0.722***
	(0.007)	(0.100)	(0.101)	(0.223)	(0.221)	(0.197)	(0.054)	(0.055)	(0.052)
Conscientiousness	-0.019	-0.023	-0.024	0.003	0.021	0.058	-0.012	-0.001	0.034
	(0.043)	(0.040)	(0.038)	(0.059)	(0.060)	(0.052)	(0.067)	(0.067)	(0.062)
Openness	0.209***	0.206***	0.178***	0.554***	0.536***	0.460***	0.621***	0.610***	0.563***
	(0.040)	(0.039)	(0.038)	(0.060)	(0.059)	(0.063)	(0.059)	(0.058)	(0.055)
Perseverance	-0.054	-0.050	-0.049	-0.057	-0.087	-0.118	-0.055	-0.059	-0.080
	(0.043)	(0.043)	(0.038)	(0.052)	(0.052)	(0.050)	(0.063)	(0.063)	(0.056)
Self-efficacy	-0.017	-0.017	-0.012	-0.050	-0.056	-0.050	0.030	0.024	0.057
	(0.041)	(0.041)	(0.043)	(0.075)	(0.074)	(0.072)	(0.083)	(0.080)	(0.076)
Self-esteem	0.080*	0.076	0.057	0.107	0.114	0.077	0.056	0.059	-0.030
	(0.039)	(0.039)	(0.039)	(0.062)	(0.065)	(0.059)	(0.068)	(0.066)	(0.067)
EA3 PGS		0.197***	0.189***		0.362***	0.333***		0.448***	0.422***
		(0.038)	(0.037)		(0.047)	(0.045)		(0.091)	(0.085)
Controls			✓		✓		✓		✓
N	1,805	1,805	1,805	1,805	1,805	1,805	1,804	1,804	1,804
R^2	0.057	0.075	0.111	0.092	0.121	0.211	0.239	0.249	0.358

†p<0.1, *p<0.05, **p<0.01, ***p<0.001.

Table 3.A.9: Moderation of the association between cognitive/non-cognitive skills and years of education by educational attainment in the WLS.

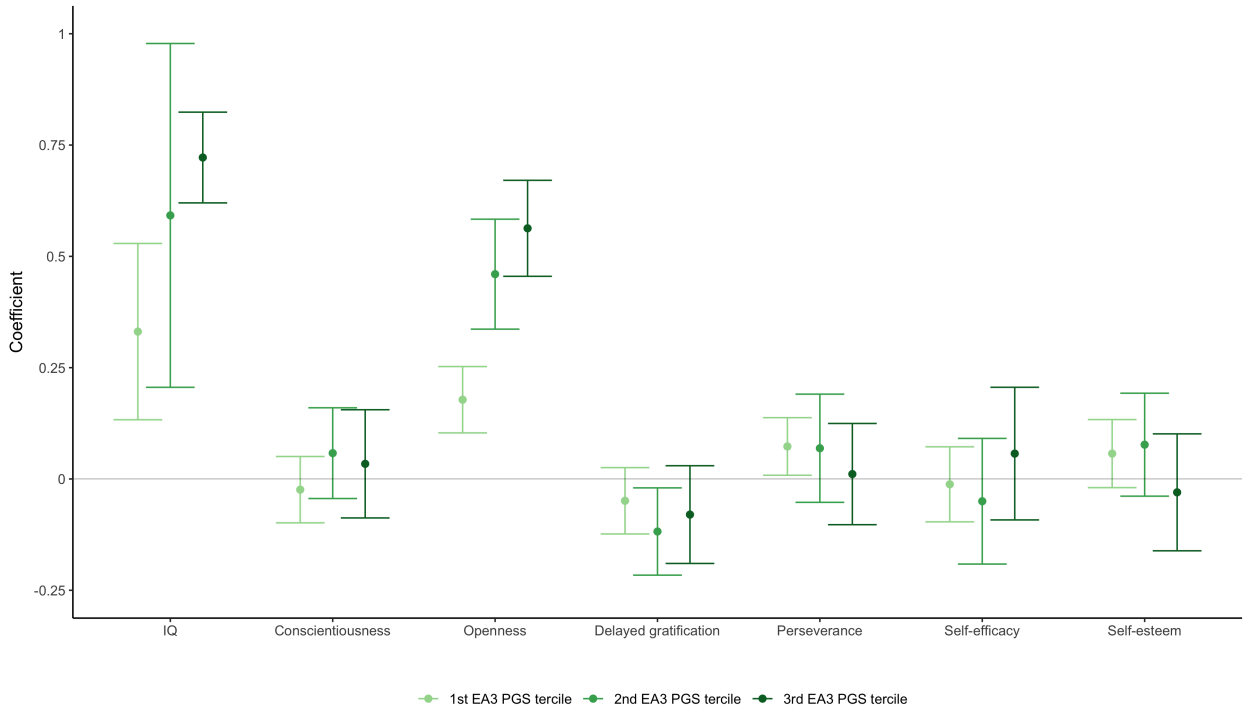


Figure 3.A.14: Coefficient sizes of the association between cognitive/non-cognitive skills and years of education by the educational attainment polygenic score tercile in the WLS.

The explanatory power of IQ increases the higher the polygenic score tercile is. For example, for the individuals in the lowest tercile, a 1 standard deviation increase in IQ corresponds to a 0.371 year increase in educational attainment while for the individuals at the highest tercile it corresponds to a 0.722 year increase in years of education. The same pattern emerges for openness. Here, a 1 standard deviation increase in openness is associated with a 0.208 year in education increase for the individuals with the lowest polygenic scores whereas for the individuals with the highest score it is associated with a 0.563 year increase. The effect of the educational polygenic score on years of education also increases at the higher polygenic score levels. In other words, even at the right end of the distribution in the polygenic score, the differentials in the polygenic score are very predictive of educational attainment.)

Chapter 4. The Role of Genetic Propensity to Education for Status Attainment

The Role of Genetic Propensity to Education for Status Attainment

Abstract

Status attainment literature has largely ignored an important stratification mechanism — the genetic pathway. Recent innovations in the field of social science genetics allow us to fill this gap by incorporating the educational attainment polygenic score into traditional status attainment models. This paper aims at understanding the roles of genetic predisposition to education, educational attainment and parental socio-economic background for occupational outcomes over the life course and exploring differences by gender. Linear regression models run on the data from the Wisconsin Longitudinal Study (N$\approx$ 5000) show that the educational attainment polygenic score is highly predictive of occupational outcomes, captured by three different measures: the occupational education, the occupational income, and Siegel prestige scores ($R^2 \approx 5-7\%$). However, the polygenic score does not seem to have an independent association with the outcomes net of educational attainment. Furthermore, I find that the importance of education-linked genetics wanes over time for men but not for women. The role of educational attainment and social origins is universally diminished over the life course in favour of existing work experience. Finally, interesting gendered patterns are uncovered by using different occupational attainment measures: gender differences are the smallest when the occupational educational scores are considered and the largest for the occupational income scores. These findings highlight the importance of approaching status attainment as a multidimensional gendered phenomenon which can be explained only by using a rich set of outcome measurements and explanatory factors.

Keywords: educational attainment polygenic score, educational attainment, occupational attainment, gender differences, Wisconsin status attainment model, Wisconsin Longitudinal Study

4.1 Introduction

Social stratification researchers have for a long time been interested in the so-called O-E-D triangle - the relationship between one's origins, education, and destinations (Erikson and Jonsson, 1996; Hout, 1988). One of the most successful attempts to disentangle this relationship has been the Wisconsin status attainment model with its emphasis on social psychological aspects of intergenerational transmission of advantages and disadvantages (Sewell et al., 1969; Featherman, 1971; Featherman and Hauser, 1978). The model creators' intention from the beginning was to have an inconclusive model: as sociological theory becomes more nuanced and methods more powerful, new explanatory factors would be continuously added to the model (Haller, 1982). Unfortunately, these wishes were not fully realised: the Wisconsin model is still best known in its original form which includes only men with rural origins and which does not extend beyond the first job.

Another glaring gap in the status attainment literature is the lack of understanding of the role that genetics play in this context. Recent methodological developments allow us to directly measure one's genetic propensity to education in the form of polygenic scores and include them in our models. With several notable exceptions (Belsky et al., 2016, 2018), such polygenic scores have not been widely used to explain life success beyond strictly educational outcomes. However, they should be incorporated in the occupational attainment models as they are theoretically linked to all important factors in the origins-education-destinations puzzle. Due to assortative mating, the educational attainment polygenic score has been shown to be correlated with social origins and, thus, to explain a proportion of intergenerational transmission of resources (Conley et al., 2015; Belsky et al., 2016). Moreover, an enriched home environment is crucial for genetic potential to turn into tangible advantages (Turkheimer et al., 2003; Guo and Stearns, 2002). The predictive power of the polygenic score for educational attainment is undeniable: it can explain up to 11% of the variance in years of education (which is comparable to many widely used traditional sociological explanatory measures). Finally, due to the black-box nature of the educational attainment polygenic score, we do not fully understand yet what it truly captures. It is likely that it contains information on the predispositions which are not exclusively important in an academic realm but also in a professional context (Belsky

et al., 2018). In other words, it would be interesting to know whether the educational attainment polygenic score has any effect on one's status attainment beyond its association with years of education.

The role of genetics in status attainment can be expected to be multi-faceted, just as individual career trajectories are. As a consequence, it is important to not only look at first jobs but also to explore how one's status evolves over time. In other words, the intergenerational perspective, which links origins and the first job, should be complemented with the intragenerational approach by adding information on career moves that occur later in life (Erikson et al., 1979; Carroll and Mayer, 1987). This allows us to explore how the role of the genetic propensity, educational attainment and social origins changes in relation to actual job experience and performance (Herd et al., 2019). The measurement of occupational outcomes has been shown to be sensitive to gender differences (Warren et al., 1998; Warren, 2001). Whereas measurements which rely more heavily on the educational credentials of a given occupation's incumbents (for example, educational income scores) might display minimal gender differences (as women are as likely or even more likely than men to enter professions that require higher levels of education), those focusing on income or prestige often demonstrate stark gender disparities (Warren et al., 2002; Hauser et al., 1999). Rather than being an impediment to drawing definitive conclusions, such measurement heterogeneity should be embraced and exploited to illuminate interesting gendered patterns in status attainment over the life course.

This paper makes several contributions both to the field of social science genetics and to the social stratification research. First, I explore the role of genetic propensity to education on occupational outcomes and make an inquiry into whether some part of this role is independent of educational attainment. Second, I do not limit this exploration to the first job but rather investigate the intra-generational dynamics of potentially waning or growing influences of the educational attainment polygenic score, educational attainment and parental socio-economic background on one's career success. Third, I treat status attainment as a gendered process, where gender affects both early-life processes of turning genetic propensities to skills and later-life processes of turning skills into status rewards. Finally, I conceptualise occupational attainment as a multi-faceted phenomena and, in

turn, instead of relying on one specific indicator, I use three different measures — the occupational educational score, the occupational income score, and the Siegel prestige score — in order to discover important gender-specific patterns in status attainment.

4.2 Theoretical framework

4.2.1 Traditional status attainment models

The concept of status and the processes that generate inequalities in status among social units is at the very core of the sociological discipline (Haller and Portes, 1973). Status in this context denotes power, prestige, and general social standing (Haller, 1982). One of the most important dimensions of status is occupation since in modern societies it greatly influences one's life chances (Haller and Portes, 1973). In particular, sociologists have long been interested in how occupational attainment is transmitted intergenerationally — from parents to children. At the beginning of this research tradition researchers used mobility tables to describe intergenerational status attainment patterns. From the late 1960s the focus shifted from describing to explaining. With this goal in mind the Wisconsin status attainment model was developed (Sewell et al., 1969; Sewell and Hauser, 1972; Woelfel and Haller, 1971). It aims at understanding the relationship between one's origins, educational attainment, and occupational attainment by specifying causal chains. While a lot is already known about the transmission of educational attainment, less is known about occupations. Undoubtedly, educational attainment mediates a big portion of the association between origins and destination. The real challenge, however, is specifying other factors which go beyond educational attainment to influence occupational attainment (Sewell and Hauser, 1972; Kerckhoff, 1976).

With several exceptions (Belsky et al., 2018, 2016), one mechanism of occupational transmission has been virtually unexplored in status attainment literature — genetic transmission. From twin studies and recent molecular genetic advances we know that educational attainment is to a large extent heritable (Lee et al., 2018; Okbay et al., 2016). Since educational attainment is highly correlated with career success, occupational attainment should also be expected to have a heritable component

(Sewell and Hauser, 1972; Erikson and Jonsson, 1996). However, genetic effects can go over and above educational attainment while exerting their power over one's career success. First, certain genetic predispositions might affect one's non-cognitive skills (Krapohl and Plomin, 2016; Okbay et al., 2016; Cesarini and Visscher, 2017; Plomin et al., 1994; Benjamin et al., 2012). Those traits, such as self-discipline, delayed gratitification, openness to experience, might then in turn be highly rewarded by employers, thus, enhancing one's occupational trajectory (Jencks, 1979; Bowles et al., 2001; Heckman and Kautz, 2012b; Heckman, 2006). Second, parents might also genetically transfer intrinsic interests, motivations and preferences, which can in turn shape children's career choices. For example, genetic propensities have been identified for creative thinking or systematising, a desire for autonomy, and entrepreneurship (Lichtenstein et al., 1995; Rimfeld et al., 2016).

4.2.2 Extending traditional models: the educational attainment polygenic score and the job market entry

Recent developments in social science genetics can enrich existing status attainment models by allowing us to make causal mechanisms operating between origins and destinations even more detailed. Specifically, the educational attainment polygenic score, which measures one's genetic propensity to educational attainment, can be incorporated as another mediator of parental influence on attainment (Lee et al., 2018; Belsky et al., 2018). The educational attainment polygenic score represents both the power of ascription and socialisation. One the one hand, because it is fixed at birth, the polygenic score can be seen as a classic ascriptive attribute. On the other hand, the realisation of one's genetic potential highly depends on the social environments in which an individual grows up (Guo and Stearns, 2002; Turkheimer et al., 2003). In other words, the educational attainment polygenic score represents a transmission pathway of social advantages and disadvantages which is not determinstic but rather contingent on one's social circumstances. Consequently, as Belsky et al. (2016) once put it, the educational attainment genetic discoveries are not simply about education but rather about socioeconomic success more broadly. Belsky et al. (2018) find that the educational attainment polygenic score is associated with occupational attainment. However, we do not truly know how this association is produced. Does the influence of

the polygenic score on occupational attainment is mediated entirely by its strong association with educational attainment? Or does this transmission have independent power net of other variables (Figure 4.1)?

Studying the role of the educational attainment polygenic score can help achieve two goals stated in the status attainment literature (Kerckhoff, 1976). First, it could make predictions of status attainment more accurate. One's origins, or socio-economic background, have been shown to be associated with the educational attainment polygenic score (Krapohl and Plomin, 2016; Selzam et al., 2017). Incorporating information on genetic transmission, if it is not entirely mediated by the educational attainment, could improve the power of status attainment models which aim at predicting one's destination from one's origins. Second, the status attainment research focuses on theoretically specifying "finer mechanisms" in the causal chain from origins to destination (Kerckhoff, 1976). Haller and Portes (1973) believed that the Wisconsin model "constitutes a heuristic approximation in need of further development" (p. 66). Consequently, even if the polygenic score proves to have no independent effect on occupational attainment from its association with educational attainment, it should not be simply delegated to the error term but rather theoretically specified as a contributor to disparities in educational attainment.

Occupational attainment processes are unique to men and women (Haller and Portes, 1973; Bukodi and Goldthorpe, 2011). Life-course transitions shape men's and women's mobility differently (Fuller, 2008). The reasons for such disparities are numerous: family formation (Warren, 2001; Budig and Hodges, 2010; Evertsson and Nermo, 2004), gender discrimination and segregation (Warren et al., 2002; Alon and DiPrete, 2014), or gender role socialisation (Evertsson and Nermo, 2004; Legewie and DiPrete, 2014). The educational attainment polygenic score can also operate differently for men and women. Especially in our sample (the Wisconsin Longitudinal Study), collected during the time where women's occupational opportunities were still very limited, men might have had more opportunities to realise their genetic potential. Due to the possibly differential treatment by parents and teachers and socialisation into how much education should be aspired and expected, men might display a stronger association between their polygenic score and eventual educational attainment (Herd et al., 2019). Because men in that cohort in general have stronger association

between education and occupation than women (Breen and Jonsson, 2005), by definition, their association betweeen the polygenic score and occupation should also be stronger than for women.

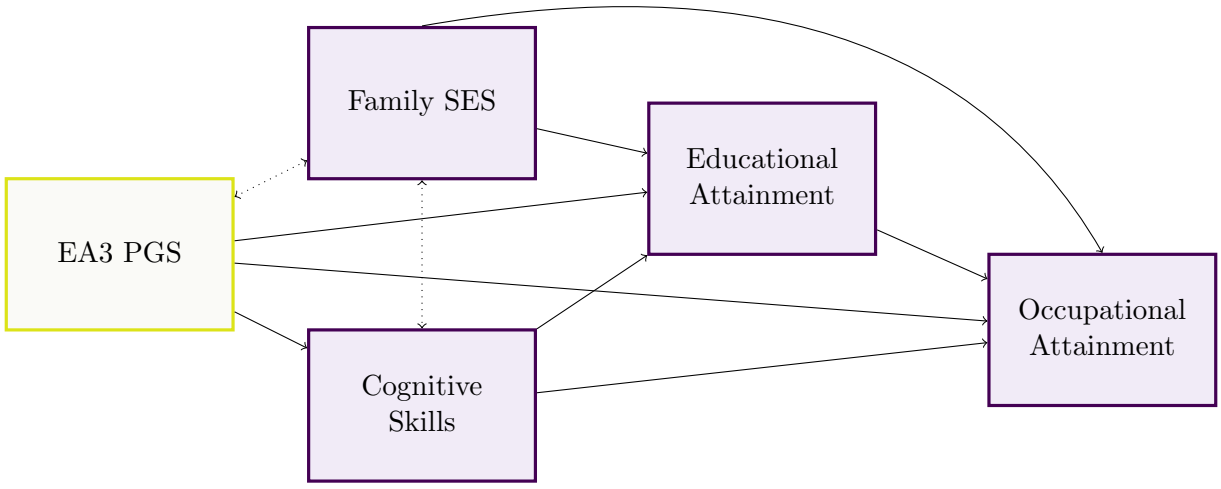
It is also important to mention that such gender patterns might be sensitive to the choice of occupational attainment measurement. There exist three main approaches in sociological literature for operationalising occupational attainment: social classes, prestige indexes, and socio-economic indexes. Theoretically derived social classes focus on employment relations in the labour market (Erikson et al., 1979; Erikson and Goldthorpe, 1992; Grusky and Sorensen, 1998; Grusky and Weeden, 2006). When constructing prestige indexes, survey respondents are asked to rate jobs. A prestige score then represents a percentage of respondents who say that a given occupation is above a certain prestige threshold. The most widely used prestige indicators are the Standard Occupational Prestige Scale (Treiman, 1977; Ganzeboom and Treiman, 2003) and scales developed by Reiss (1961); Siegel (1971); Nakao and Treas (1991). The WLS provides the Siegel prestige score for occupations. The third category is composite indexes for socio-economic standing. Some examples for such indexes include the International Socio-Economic Index (Ganzeboom et al., 1992) and Duncan Socio-Economic Index (Duncan, 1961). Such SEIs combine three metrics: occupational education, occupational income, and occupational prestige. Occupational education corresponds to a proportion of individuals within a given occupation whose education is higher than a specified threshold. For instance, the WLS occupational education score of 50 would mean that 50% of occupation's incumbents have one or more years of college education (the educational threshold here is one year of college). Similarly, occupational income represents a proportion of incumbents of a given occupational that earn above a certain threshold. For example, the WLS occupational income score of 50 would mean that 50% of all incumbents of that occupation earn more than \$10,000 in dollars of 1969 (the income threshold here is \$10,000). Constructing Duncan SEI involves regressing prestige scores (in the WLS it is Siegel prestige scores) on occupational education and occupational income scores. The estimates of the association of these two scores and occupational prestige are then used as weights to combine occupational education and occupational income into a composite scalar variable of socio-economic standing.

Although used in the early status attainment research, Duncan SEI (and other socio-economic

indexes) have received a lot of criticism (Hauser and Warren, 1997; Warren et al., 1998). In particular, researchers are cautioned against using SEIs when gendered occupational patterns are of interest. According to Warren et al. (1998), in such global indexes, the weights given to occupational education and occupational income are arbitrary. Since men have usually had a much larger advantage in the educational income scores than in the educational education scores (Warren et al., 2002), Duncan SEI, by combining these two measures, in a way conceals important gender dynamics. Consequently, Warren (2001) recommends using both measures separately to truly understand their advantages and disadvantages in a multi-dimensional occupational structure when gender is concerned. In this paper I will follow this advice and instead of using a composite measure of Duncan SEI which, as shown before, combines occupational education, occupational income and occupational prestige, I will explore each of these three measurements individually.

To properly consider all important aspects of occupational attainment at the labour force entry, I will test four hypotheses in the first part of my analysis. First, I will investigate the association between the educational attainment polygenic score and occupational attainment. I expect the estimated association to be highly significant for all three measures and more so for men than for women (Hypothesis 1a). Second, after including educational attainment, cognitive ability, and parental SES, I expect the educational attainment polygenic score to still have some independent effect on occupational outcomes for men but not for women (Hypothesis 1b). Third, the polygenic score should be the most highly associated with the occupational education score, as, compared to other two measures, it relies most heavily on education of a given occupation's incumbents. At the same time, I expect to observe the most pronounced gender differences when the association between the polygenic score and the occupational income is considered (Hypothesis 1c). Fourth, I expect social origins to be correlated both with our outcome measures and the educational attainment polygenic score, thus confounding the association both for men and women (Hypothesis 1d).

Figure 4.1: Theoretical causal graph.



Notes: Straight solid lines — theoretically expected causal links. Bidirectional dotted lines — correlations without assigned causal priority.

4.2.3 Occupational attainment over the life course

The status attainment literature has most often investigated the determinants of first job success while ignoring intragenerational mobility of individuals (Burkhauser et al., 2009). However, the concept of destination should be understood not just as a characteristic fixed upon entering the labour market but rather as a work history which develops throughout one's life course (Blossfeld, 1986; Carroll and Mayer, 1987). The limited existing research has investigated the predictors of the evolution of occupational attainment in terms of wages (Cheng, 2014; Cheng and Song, 2019), prestige or desirability (Manzoni et al., 2014; Barone et al., 2011; Härkönen et al., 2016; Sacchi et al., 2016), and social class (Bukodi and Goldthorpe, 2011). The aim of this paper is to better understand when the effects of explanatory factors on status attainment variables change (Haller and Portes, 1973). Specifically, the change in the role of the educational attainment polygenic score on occupational attainment will be investigated.

There are two main possibilities for the evolution of the effect of the polygenic score over the life course. The first possibility is for the polygenic score to diminish in importance. It has been previously demonstrated that, generally, the role of ascriptive attributes tends to lose its

significance as employers learn about the true producitivity of their employees (Blau and Duncan, 1967; Featherman, 1971; Featherman and Hauser, 1978; Kelley, 1973; Hauser et al., 1999; Sewell et al., 1980; Warren, 2001). For example, the association between educational attainment and occupational attainment, while very strong for the first job, is reduced for later occupations. The theory of certifying suggests that many employers require credentials to prove relevant competences. While those credentials serve as a minimum requirement in the beginning of one's career, once the labor market is entered they do not matter anymore (Spilerman, 1977; Jackson et al., 2005). The theory of signalling also sheds some light on why education is so important for the first job but not so much later in one's career. When important worker characteristics such as motivation or perseverance are not directly observable, credentials serve as an imperfect but preferred indicator of these desired characteristics. Once employees have opportunity to prove these skills on the job, diplomas become less important and predictive of further success (Berg, 1970; Billis, 1988; Bowles and Gintis, 1976). A similar logic can be applied to the educational attainment polygenic score. If its association with occupational attainment is entirely mediated by educational attainment, its effect should decrease over one's life course. However, the polygenic score might capture a broader set of characteristics whose importance, rather then diminish over time, actually becomes stronger. The first mechanism which could make this possible is sorting. If the polygenic score does not perfectly correlate with educational attainment but represents some desired attributes beyond education, the longer one is in the job market the more effectively these attributes can be identified by employers, put into good use and rewarded (Hauser et al., 1999; Warren, 2001; Carbonaro, 2007; Farkas et al., 1997). The second mechanism is on-the-job skill development. The third educational attainment GWAS identified genetic variants which encode synaptic plasticity — main cellular mechanisms for learning and memory (Lee et al., 2018). Consequently, the polygenic score might reflect ability to learn and acquire human capital on the go, certain agility and plasticity which is very important for ongoing self-improvement and which extends beyond existing educational attainment. Indeed, occupational mobility can be a product of learning on the job which can even compensate for a previous lack of educational opportunities (Bukodi and Goldthorpe, 2011).

Potential gender differences are also important to consider. For men, it is reasonable to expect that as they progress through their careers, their education, and relatedly the polygenic score, loses its

importance while their productivity, experience, or personal connections become more important for occupational attainment. For women, however, the mechanism might work differently. Since the WLS women in their early years might not have had as many opportunities to realise their genetic potential as men, their association between the polygenic score and educational attainment and, by extension, occupational attainment at the start of their career (which for most of them was in the early 1960s) might be lower. Moreover, in the 1960s and the 1970s these women would have been expected to stay home and care for their kids. Figure 4.2 demonstrates that more than half of the women in our sample were not in the labour force in 1970 and many of them did not come back by 1975. This trend can probably be explained by their commitments of raising children at that particular time in their lives. However, for those who stayed in the labour force or those who reentered after some time, the initial mismatch between their intrinsic skills and their first jobs can be adjusted through the sorting mechanism: their genetic propensity can become visible on the job and then be more fairly rewarded (Manzoni et al., 2014). Moreover, as structural constraints weakened in the 1970s and 1980s (DiPrete and Buchmann, 2006; Blau et al., 2013), some women were potentially able to compensate for their early disadvantages by acquiring human capital on the job and translating these new skills into occupational benefits. The process of learning on the job might have been facilitated by their genetic propensity to education, previously not fully utilised.

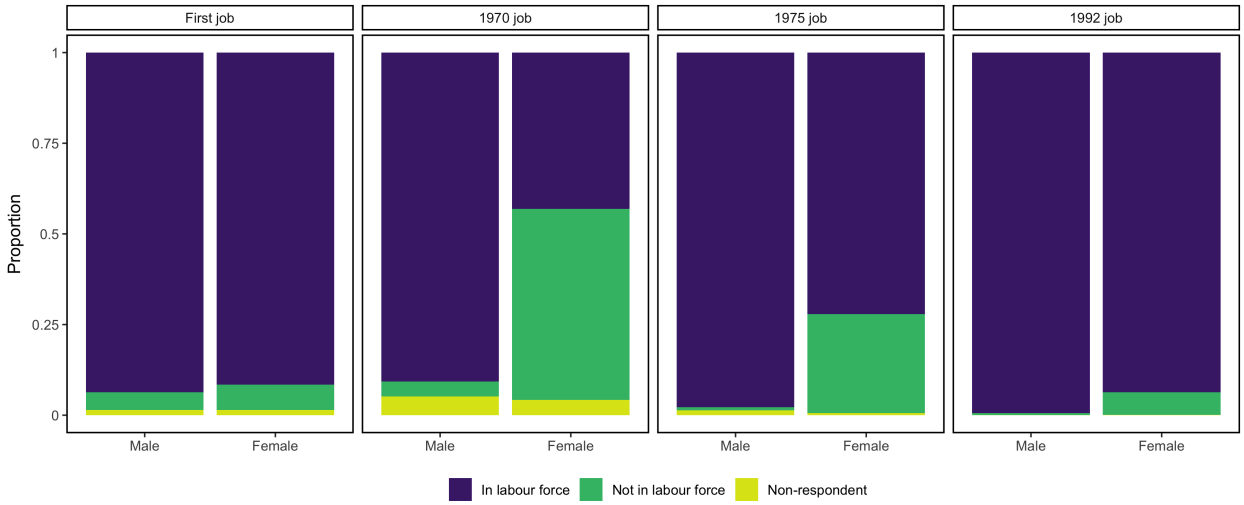


Figure 4.2: The evolution of proportions of respondents in labour force, not in labour force or with missing data by gender in the WLS.

All in all, in the second part of my analyses I will test three hypotheses. First, I expect the importance of the educational attainment polygenic score to wane over time for men but not for women (Hypothesis 2a). Second, the role of social origins and educational attainment should also decrease over the life course as previous jobs become more important (Hypothesis 2b). Finally, I expect to see some differences depending on which occupational attainment measurement we use. The association between the polygenic score and the outcome should weaken the most when the occupational income score is considered, as previous occupational success, job experience and other less systematic personal circumstances should be very important for this aspect of occupational attainment (Hypothesis 2c).

4.3 Data

The previously stated hypotheses will be tested using the Wisconsin Longitudinal Study (WLS). This dataset is uniquely suited for the needs of this paper because it contains genetic information as well as occupational attainment measures at several points in time. The original survey was first administered in 1957 and collected information on a random sample of one third of all graduating seniors (called graduates) at Wisconsin's high schools who were born in 1939-1940. The response rate from the sampling frame was 100% (Herd et al., 2019). The dataset follows the graduates from the first data collection in their late adolescence in 1957 throughout adulthood with additional data collections in 1964, 1975, 1992, 2004, and 2011. Starting in 1977, the WLS collects information on a selected sibling of each graduate. It has rich information on their family background, adolescent characteristics, educational history, employment experiences, income and wealth, family formation, physical and mental health, psychological well-being, cognition, end of life planning, and mortality (Herd et al., 2014). Saliva samples for DNA sequencing were collected from both graduates and siblings in 2006 to 2007 using mailed Oragene kits and in 2011 during individual interviews. ∼710,000 markers were genotyped using Omni-Express BeadChip at Johns Hopkins' Center for Inherited Disease Research (Herd et al., 2014). Despite its numerous strengths, the dataset poses challenges for the extrapolation of the results to wider populations because almost all respondents are non-Hispanic whites and because of the truncated measure of educational attainment (Hauser

and Palloni, 2011). The sample used in this paper will only include graduates and siblings who have had their genetic information collected. As can be seen in Figure 4.A.1 in the Appendix, this requirement leaves us with 7,478 individuals — 5,414 graduates and 2,064 siblings. Most of the analyses will use the data of only graduates because they have more thorough records of occupational history and a wider range of needed explanatory variables. However, the first part of the analysis, which will investigate associations with occupational attainment at the job market entry, I will also use the sibling sample to run within-family models.

4.4 Measurement

4.4.1 Occupational attainment

According to the recommendations provided by Warren et al. (2002), occupational attainment will be measured with three separate variables: occupational education, occupational income, and the Siegel prestige score. Occupational attainment is measured for jobs at four different time points: the first job, the job held in 1970, the job held in 1975, and the job held in 1992 or 1993. Their attainment scores are based on the 1970 Census occupational classification.

Occupational education

Occupational education is computed as the proportion of incumbents in a given occupation who completed one or more years of college and, in theory, can range from 0 to 100, where 100 means that all incumbents of that occupation have one or more years of college education. Descriptive statistics for this measure can be seen in Table 4.1.

Occupational income

Occupational income is computed as the proportion of incumbents in a given occupation who earned more than $10,000 in 1969 and, in theory, can range from 0 to 100, where 100 means that all incumbents of that ocupation earn $10,000 or more. Descriptive statistics for this measure can be seen in Table 4.1.

Siegel prestige score

The Siegel prestige score captures the general perceived prestige of a specific job. The prestige mapping I will use in the analysis is from 1970. The score can range from 0 to 100. Descriptive statistics for this measure can be seen in Table 4.1.

4.4.2 Educational attainment

Educational attainment in this paper is operationalised as a continuous variable of years of education. Since by design of the dataset the respondents have completed high school, the values of the educational attainment variable range from 12 to 21 years, where 16 years correspond to holding a college degree. Educational attainment in the WLS is measured in three interviews: 1975, 1992, and 2003. The distributions for each measure can be seen in Figure 4.A.2 in the Appendix. The exact details of how an originally categrocal variable of educational attainment was recoded into a numerical one and how some of the categories differ from one interview to another can be found in the Appendix.

4.4.3 Educational attainment polygenic score

The educational attainment polygenic score is a single aggregate measure of the multiple genetic effects correlated with educational attainment (Lee et al., 2018; Okbay et al., 2016). The polygenic score for an individual i is equal to the sum of the allele counts a_{ij}(0,1 or 2) from each SNP j

= 1,..M, multiplied by a weight w_j: $PGS_i = \sum_{j=1}^{m} w_j a_{ij}$ (Mills et al., 2020a), where the weights are association coefficients produced in the educational attainment genome-wide association study (GWAS). In this paper the polygenic score from the latest 1.1 million-person educational attainment GWAS will be used (Lee et al., 2018). It's the third educational attainment GWAS and, as is convention, the educational attainment polygenic score in this paper will be abbreviated as the EA3 PGS. Individual's educational attainment polygenic score can then be interpreted as a single quantitative measure of one's genetic propensity to educational attainment (Mills et al., 2020a). More detailed information on how the EA3 PGS was constructed in the WLS can be found online (https://www.ssc.wisc.edu/wlsresearch/ documentation/GWAS). To facilitate the analysis, the score was standardised to have a mean of 0 and a standard deviation of 1 (Table 4.1). Our analyses also include 10 principal components to control for potential population stratification.

4.4.4 Cognitive skills

Cognitive skills, or mental ability, is measured as a score from the Henmnon-Nelson IQ test completed in 1957. The measure is standardised for the analyses to have a mean of 0 and a standard deviation of 1.

4.4.5 Parental SES

The WLS provides a factor-weighted parental SES score which combines information on father's years of education, mother's years of education, Duncan's SEI for father's 1957 occupation, and average parental income. This measure broadly captures the origins of the respondents. For our analyses the score is standardised to have a mean of 0 and a standard deviation of 1.

Descriptive statistics for each variable is further broken down by respondent type (graduate or sibling) and gender (male and female) and can be found in Table 4.A.1 in the Appendix.

Variables	Mean/Proportion	SD	Range	N
EA3 PGS	0	1.00	-3.32 - 4.05	5,414
Occupational education of first job	33.48	30.37	1.90 - 99.50	5,012
Occupational education of 1970 job	39.49	30.85	2.10 - 99.50	3,563
Occupational education of 1975 job	37.61	29.37	2.10 - 99.50	4,567
Occupational education of 1992/1993 job	40.01	28.60	2.10 - 99.30	5,041
Occupational income of first job	17.73	20.46	0.20 - 87.70	5,012
Occupational income of 1970 job	30.41	23.84	0.20 - 87.70	3,563
Occupational income of 1975 job	28.35	24.17	0.20 - 87.70	4,567
Occupational income of 1992/1993 job	30.36	24.40	0.20 - 87.70	5,041
Siegel prestige score of first job	41.12	14.55	14.10 - 81.20	5,012
Siegel prestige score of 1970 job	45.49	13.82	12.20 - 81.20	3,563
Siegel prestige score of 1975 job	44.29	13.89	14.10 - 81.20	4,567
Siegel prestige score of 1992/1993 jobs	45.38	13.65	14.10 - 81.20	5,041
Years of education by 1975	13.55	2.23	12.00 - 20.00	5,384
Years of education by 1992/1993	13.75	2.32	12.00 - 21.00	5,233
IQ	102.22	14.70	61.00 - 145.00	5,414
Parent SES	0	1.00	1.39 - 5.59	5,414
Gender				5,414
Female	0.52			2,830
Male	0.48			2,584

Table 4.1: Descriptive statistics for graduates who have the polygenic score available in the WLS.

4.5 Analytical strategy

The first part of the analysis will look at whether the educational attainment polygenic score has an independent effect on occupational attainment of the first job or, in contrast, it is entirely mediated by educational attainment and cognitive ability. First, I will explain differences in the occupational education, the occupational income and Siegel prestige scores for men and women separately using only the educational attainment polygenic score and controlling for 10 principal components in the following OLS regression (Hypothesis 1a):

$$FirstJob_i = \alpha + \beta \times PGS_i + \epsilon_i \tag{4.1}$$

where $FirstJob_i$ indicates either occupational education or occupational income of the first job;

PGS_i represents the educational attainment polygenic score for the respondent i.

After that, I will also include the educational attainment, cognitive ability, and parental SES measures to ascertain whether the polygenic score has any independent association with the occupational attainment net of these variables (Hypothesis 1b) and the role of SES at the job market entry (Hypothesis 1d) by also exploring heterogeneity by different outcome measures (Hypothesis 1c):

$$FirstJob_i = \alpha + \beta \times PGS_i + \gamma \times Edu_i + \mu \times IQ_i + \omega \times SES_i + \epsilon_i \tag{4.2}$$

where $FirstJob_i$ indicates either occupational education or occupational income of the first job; PGS_i — the educational attainment polygenic score; Edu_i - years of education; SES_i — parental SES for the respondent i.

To account for possible measurement error in the parental background measure, I will also run sibling fixed effects to control for all family characteristics shared by both siblings. Sibling pairs will be divided into three categories following Warren et al. (2002): female-female, male-male, mixed-gender pairs. As Warren et al. (2002) writes, sibling disparities can be heterogeneous by the specific gender pattern in a family. In the context of the educational attainment polygenic score, within families the association between the score and occupational outcomes might be more pronounced when one of the siblings is male and another — female, as higher polygenic scores for men can further increase their advantages in the labour market, whereas differences in the educational attainment polygenic score can have little or no impact in female sibling pairs as they might not have opportunities to realise their genetic potential anyway. For these three sibling groups the following models will be run:

$$FirstJob_i = \alpha + \beta \times PGS_i + \gamma \times Edu_i + \mu \times IQ_i + + \sum_{k=2}^{N} \omega_k \times S_k(i) + \epsilon_i \quad (4.3)$$

where $S_k(i)$ is 1 if sibling i is in family k and 0 otherwise, with $k = 1$ family excluded as the reference.

In the second part of my analysis, to investigate the association between the educational attainment polygenic score and later occupations and to test hypotheses 2a, 2b, and 2c the same modeling strategy will be used except that I will also account for the occupational attainment measured in the previous time point:

$$1970Job_i = \alpha + \beta \times PGS_i + \gamma \times Edu_i + \mu \times IQ_i + \omega \times SES_i + \theta \times FirstJob_i + \epsilon_i \quad (4.4)$$

$$1975Job_i = \alpha + \beta \times PGS_i + \gamma \times Edu_i + \mu \times IQ_i + \omega \times SES_i + \theta \times 1970Job_i + \epsilon_i \quad (4.5)$$

$$1992Job_i = \alpha + \beta \times PGS_i + \gamma \times Edu_i + \mu \times IQ_i + \omega \times SES_i + \theta \times 1975Job_i + \epsilon_i \quad (4.6)$$

4.6 Results

4.6.1 Descriptive results: the polygenic score, occupational attainment and gender over the life course

The previously described analytical strategy is premised on the assumption that occupational attainment changed from the first job to the jobs held at later time points. Before the actual analysis,

we should first check whether this assumption is supported by evidence. Figure 4.1 demonstrates that there has been significant intragenerational occupational mobility over the years, especially when occupational income is considered. The upward trend is present for both men and women. We should also notice a vast disparity between male and female occupational income: while both men and women on average experienced substantial mobility over their lifetimes, the gender gap never closed. One puzzling aspect seen in Figure 4.1 is the fact that for the first job women on average have higher scores of occupational education than men; however, this pattern is reversed for the jobs held in 1970 and 1975.

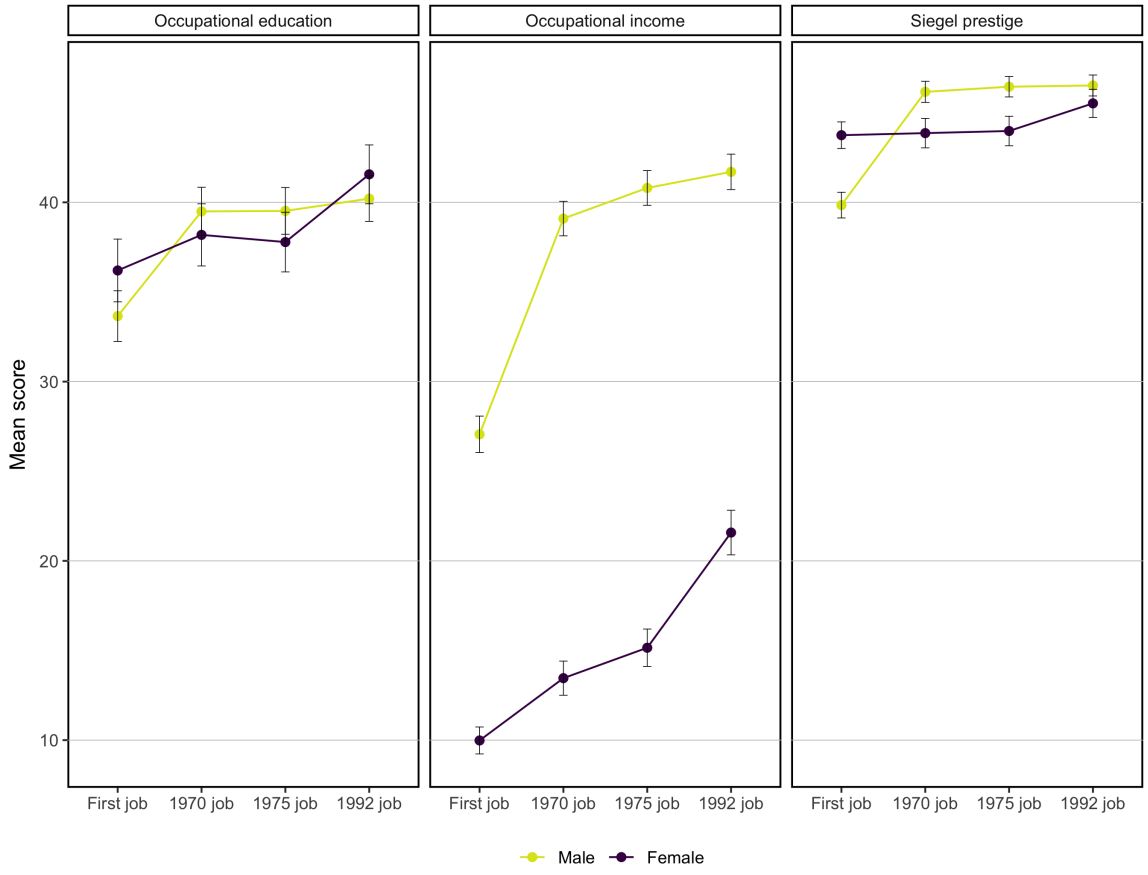


Figure 4.1: The evolution of average occupational education, occupational income and Siegel prestige scores by gender over time in the WLS.

Several trends can also be identified using Figure 4.2. First, across all three polygenic score terciles men have higher occupational outcomes than women. Second, the occupational attainment differences between the educational attainment polygenic score terciles is more pronounced for men than

for women. Third, the strength of the association between the polygenic score and occupational outcomes depends on the way those outcomes are measured: the relationship is strongest when the occupational education measure is used.

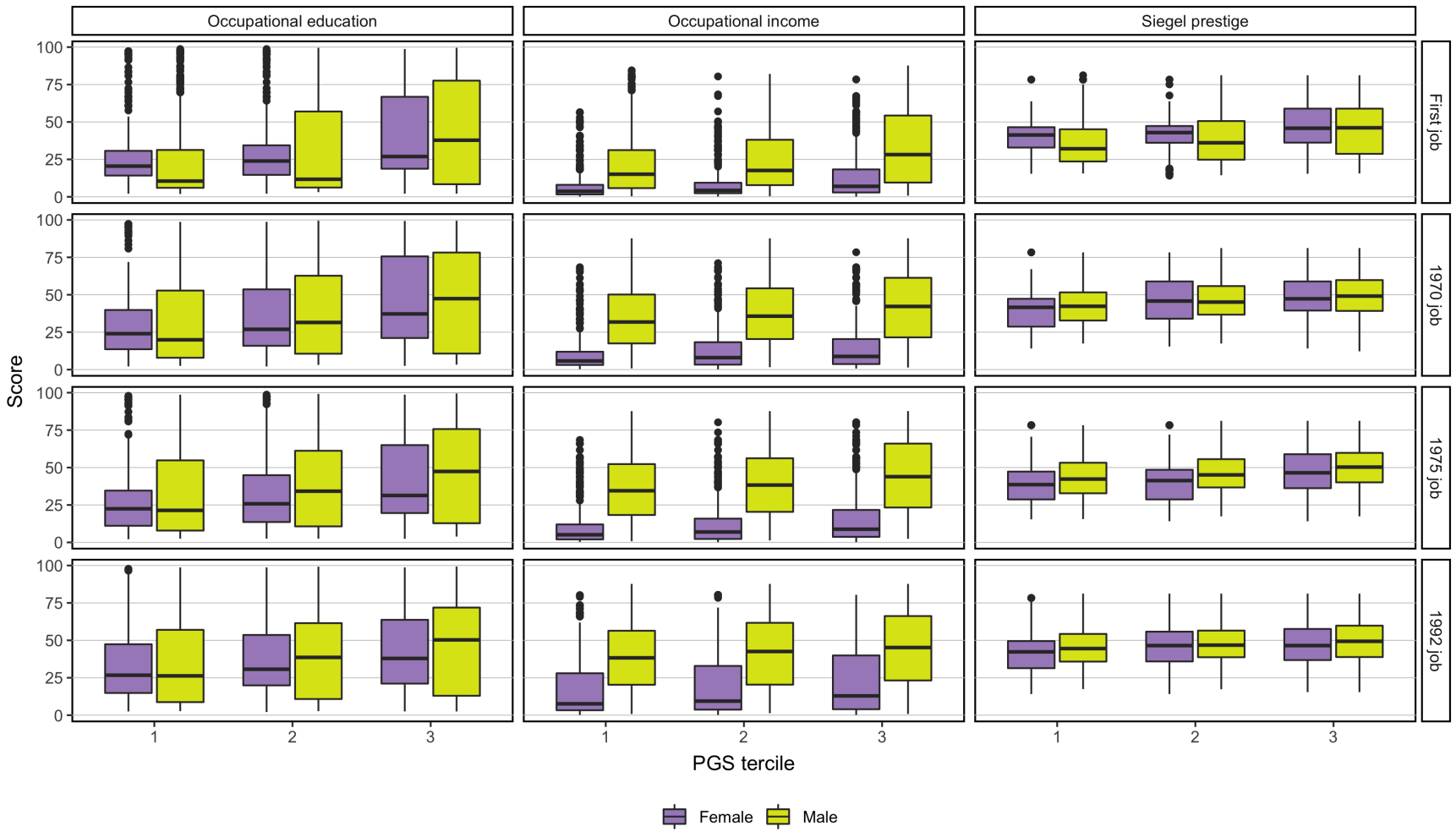


Figure 4.2: Boxplots of occupational attainment scores by gender and the educational attainment polygenic score tercile over the life course in the WLS.

4.6.2 The role of the polygenic score at the job market entry

Table 4.1 presents results from the models analysing the relationship between the educational attainment polygenic score and occupational attainment at first job. The main results are also summarised in Figure 4.3. Model 1 shows that the educational attainment polygenic score is highly associated with our outcomes of interest. As predicted by Hypothesis 1a, the association is much stronger for men than for women. Model 2, which includes an indicator for years of education, demonstrates that most of the association between the educational attainment polygenic score and first-job occupational measures is explained by educational attainment. In other words, the polygenic score does not seem to have an independent effect on career success net of education. However, here gender differences are also visible: the association between the polygenic score and all three occupational outcomes, although small, remains significant for women but not for men, lending gender-specific evidence for Hypothesis 1b. Furthermore, our results confirm the hypothesised heterogeneity by the measurement of outcomes (Hypothesis 1c). The polygenic score has the highest association with the occupational education score (a 8.751 point increase on the scale of 0-100 for men and a 6.892 point increase for women for a 1 standard deviation increase in the educational attainment polygenic score) compared to the two other occupational indicators for first job. After controlling for the educational attainment, cognitive abilities and parental socio-economic status, the polygenic score remains significant and most pronounced for women when the occupational education score is concerned. To put it differently, the polygenic score, after taking into account other important attainment variables, is independently predictive only of the occupational education score of the first job and only for women (Hypothesis 1c). Finally, we can also observe the heterogeneous role of parental background on our outcomes of interest. Net of educational attainment and cognitive ability, parental socio-economic status is only somewhat important for the income aspect of the first-job attainment: a 1 standard deviation increase in parental SES corresponds to an increase in the occupational income score by 1.551. For other outcomes it does not seem to matter that much (independently of other variables). Parental SES, net of educational attainment, is also not correlated with the polygenic score, allowing as to reject Hypothesis 1d.

Table 4.1: Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the first job held by the graduates with explanatory variables.

	Occupational education							
	(1)		(2)		(3)		(4)	
	Male	Female	Male	Female	Male	Female	Male	Female
Constant	33.475***	33.861***	-125.713***	-123.986***	-124.096***	-119.839***	-124.285***	-117.569***
	(0.648)	(0.527)	(2.070)	(2.581)	(2.349)	(2.720)	(2.451)	(2.898)
EA3 PGS	8.751***	6.892***	-0.183	1.305***	-0.262	0.955**	-0.260	0.936**
	(0.663)	(0.522)	(0.372)	(0.344)	(0.376)	(0.351)	(0.376)	(0.351)
Years of education			11.447***	12.131***	11.332***	11.816***	11.345***	11.643***
			(0.147)	(0.197)	(0.167)	(0.207)	(0.174)	(0.221)
Cognitive ability					0.578	1.780***	0.592	1.651***
					(0.397)	(0.381)	(0.401)	(0.385)
Parental SES							-0.103	0.857*
							(0.380)	(0.380)
$Adj.R^2$	0.067	0.063	0.735	0.621	0.735	0.624	0.735	0.624
N	2,420	2,592	2,420	2,592	2,420	2,592	2,420	2,592
	Occupational income							
	(1)		(2)		(3)		(4)	
	Male	Female	Male	Female	Male	Female	Male	Female
Constant	27.067***	9.213***	-60.839***	-42.765***	-52.846***	-41.622***	-50.065***	-40.980***
	(0.468)	(0.219)	(2.143)	(1.334)	(2.408)	(1.410)	(2.505)	(1.504)
EA3 PGS	5.975***	2.353***	1.042**	0.513**	0.649†	0.417*	0.620	0.411*
	(0.480)	(0.217)	(0.385)	(0.178)	(0.385)	(0.182)	(0.384)	(0.182)
Years of education			6.321***	3.995***	5.750***	3.908***	5.551***	3.859***
			(0.152)	(0.102)	(0.171)	(0.107)	(0.178)	(0.115)
Cognitive ability					2.858***	0.490*	2.658***	0.454*
					(0.407)	(0.198)	(0.409)	(0.200)
Parental SES							1.515***	0.242
							(0.389)	(0.197)
$Adj.R^2$	0.060	0.043	0.453	0.401	0.463	0.402	0.467	0.402
N	2,420	2,592	2,420	2,592	2,420	2,592	2,420	2,592
	Siegel prestige							
	(1)		(2)		(3)		(4)	
	Male	Female	Male	Female	Male	Female	Male	Female
Constant	39.869***	42.447***	-34.044***	-5.306***	-31.229***	-1.293	-30.819***	0.459
	(0.332)	(0.230)	(1.276)	(1.495)	(1.444)	(1.564)	(1.506)	(1.665)
EA3 PGS	4.121***	2.321***	-0.027	0.631***	-0.165	0.292	-0.170	0.278
	(0.340)	(0.228)	(0.229)	(0.199)	(0.231)	(0.202)	(0.231)	(0.202)
Years of education			5.315***	3.670***	5.114***	3.365***	5.085***	3.232***
			(0.091)	(0.114)	(0.103)	(0.119)	(0.107)	(0.127)
Cognitive ability					1.007***	1.722***	0.977***	1.623***
					(0.244)	(0.219)	(0.246)	(0.221)
Parental SES							0.223	0.662**
							(0.234)	(0.218)
$Adj.R^2$	0.057	0.038	0.611	0.314	0.614	0.329	0.614	0.331
N	2,420	2,592	2,420	2,592	2,420	2,592	2,420	2,592

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

The approach of sibling fixed effects produces similar results to the ones obtained using only the sample of graduates (Table 4.2). Between siblings the educational attainment polygenic score is highly correlated with occupational outcomes at first job (Hypothesis 1a). Contrary to the Hypothesis 1b, once educational attainment is taken into account, the polygenic score loses all of its association with our outcomes of interest for all sibling-pair types and outcome variables. The sizes of estimates vary by the measurement of the dependent variable (Hypothesis 1c). As was the

case in the sample of only graduates, the correlation is highest with the occupational education score, especially in the pairs where both siblings are male, and the lowest with the occupational income score again for male-male sibling pairs. It is also noteworthy that, in general, the role of the educational attainment polygenic score, years of education, and cognitive ability is much lower in exclusively female sibling pairs than in those with two brothers. In mixed pairs, the polygenic score also matter less for a female sibling than for a male sibling.

Table 4.2: Estimated coefficients from the fixed effects regressions associating occupational education, occupational income and the Siegel prestige scores of the first job with explanatory variables between siblings.

	Occupational education								
	F-F pairs			M-M pairs			M-F/F-M pairs		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Constant	3.881	-109.150***	-108.683***	7.740	-137.504***	-130.773***	60.851***	-117.669***	-115.579***
	(16.632)	(14.996)	(15.001)	(19.858)	(15.073)	(15.411)	(19.831)	(14.649)	(14.743)
EA3 PGS	5.472**	0.566	0.316	8.176***	0.537	-0.065	7.808***	0.395	0.226
	(1.704)	(1.377)	(1.398)	(2.416)	(1.647)	(1.669)	(1.886)	(1.286)	(1.293)
Years of education		9.370***	9.244***		11.120***	10.800***		10.911***	10.766***
		(0.608)	(0.620)		(0.563)	(0.585)		(0.383)	(0.401)
Cognitive ability			1.392			2.783†			1.183
			(1.338)			(1.449)			(0.969)
Gender									
Female (ref.)							2.849†	9.890***	9.721***
							(1.548)	(1.063)	(1.072)
EA3 PGS x Female							-0.613	0.576	0.585
							(1.862)	(1.244)	(1.244)
R^2	0.694	0.811	0.812	0.653	0.848	0.850	0.610	0.826	0.827
N (pairs)	384	384	384	305	305	305	656	656	656
	Occupational income								
	F-F pairs			M-M pairs			M-F/F-M pairs		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Constant	-1.209	-40.282***	-40.457***	17.700	-50.050***	-43.375**	46.602***	-40.995***	-39.554***
	(9.367)	(9.948)	(9.961)	(14.194)	(14.205)	(14.514)	(12.917)	(11.837)	(11.916)
EA3 PGS	3.364***	1.668†	1.762†	2.211	-1.352	-1.949	4.883***	1.246	1.129
	(0.959)	(0.913)	(0.928)	(1.727)	(1.552)	(1.572)	(1.228)	(1.039)	(1.045)
Years of education		3.239***	3.286***		5.187***	4.870***		5.354***	5.254***
		(0.404)	(0.412)		(0.531)	(0.551)		(0.309)	(0.324)
Cognitive ability			-0.522			2.760*			0.816
			(0.889)			(1.365)			(0.783)
Gender									
Female (ref.)							-17.519***	-14.065***	-14.182***
							(1.009)	(0.859)	(0.866)
EA3 PGS x Female							-2.388*	-1.805†	-1.799†
							(1.213)	(1.005)	(1.005)
R^2	0.585	0.645	0.645	0.626	0.715	0.719	0.637	0.751	0.752
N (pairs)	384	384	384	305	305	305	656	656	656
	Siegel prestige								
	F-F pairs			M-M pairs			M-F/F-M pairs		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Constant	25.748***	-9.016	-8.816	37.025***	-24.124**	-20.290*	61.476***	-9.214	-7.393
	(7.520)	(7.831)	(7.837)	(10.002)	(8.943)	(9.148)	(9.279)	(7.974)	(8.011)
EA3 PGS	2.518**	1.009	0.902	3.589**	0.373	0.031	4.069***	1.134	0.987
	(0.770)	(0.719)	(0.730)	(1.217)	(0.977)	(0.990)	(0.882)	(0.700)	(0.702)
Years of education		2.882***	2.828***		4.681***	4.499***		6.109***	4.194***
		(0.318)	(0.324)		(0.334)	(0.347)		(0.579)	(0.218)
Cognitive ability			0.594			1.585†			1.031†
			(0.699)			(0.860)			(0.527)
Gender									
Female (ref.)							3.322***	4.320***	5.962***
							(0.724)	(0.208)	(0.582)
EA3 PGS x Female							-1.181	-0.711	-0.703
							(0.871)	(0.677)	(0.675)
R^2	0.652	0.714	0.715	0.631	0.776	0.779	0.609	0.764	0.766
N (pairs)	384	384	384	305	305	305	656	656	656

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

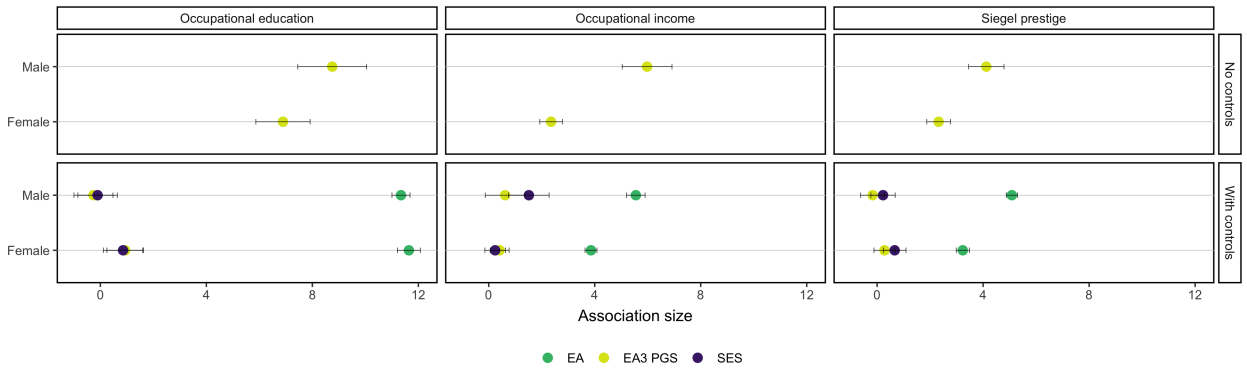


Figure 4.3: Summary of the estimated coefficients for the association between the polygenic score (EA3 PGS), educational attainment (EA) and family socio-economic status (SES) using OLS regressions in the WLS.

4.6.3 Predictors of occupational success over the life course

The next step in my analysis is to consider how occupational attainment evolved throughout an individual's life course. Regression results for the jobs held in 1970, 1975 and 1992 can be seen in Table 4.3, Table 4.4, and Table 4.5 respectively. These results confirm Hypothesis 2a, stating that the importance of the educational attainment polygenic score for career success should wane over time for men but not for women. The trend is visible in Figure 4.4: for all three occupational attainment outcomes their association with the polygenic score increases for the 1970 and the 1975 jobs compared to the first job and then comes back to the first job level in 1992. As already observed in Figure 4.2, in 1970 and 1975 a lot of women were out of the labour market. Consequently, an increased explanatory power of one's genetic propensity could result from a self-selection in our sample for these years: women who have opportunities to turn their genetic potential into concrete career benefits might be disproportionately staying in the labour force or reentering it much quicker. I explore potential differences between the women who are and who are not in the labour force at different time points (Table 4.A.4 in the Appendix). While women's years of education is somewhat predictive of the likelihood of being in the labour force, the effect is not substantial. In other words, the two groups of women are not that different based on their observable characteristics. Hypothesis 2b is also confirmed: the role of educational attainment and parental SES decreases over time whereas the importance of the previous job (net of educational attainment and socio-

economic origins) increases and peaks in 1975 both for men and women. Furthermore, as predicted by Hypothesis 2c, the life-course career dynamics and related gender differences are contingent on the specific way in which occupational attainment is operationalised and measured. Occupational education has the strongest relationship with education-related genetics (both independently and mediated by educational attainment) and it remains the strongest throughout the lives of the WLS respondents. Occupational income and Siegel prestige scores display an association with the polygenic score of similar magnitude. The starkest gender difference in the association is for occupational income: at the start of one's career the educational attainment polygenic score is very strongly correlated with the occupational income score for men but not for women. However, the difference disappears over time as the association magnitude decreases dramatically for men and increases for women. Finally, the heterogeneity by occupational outcome is further captured by comparing the overall predictive power of different models. The variance explained of all such models, represented by the adjusted R-squared, are visualised in Figure 4.5. The purple color corresponds to the adjusted R-squared in the models where only the polygenic score (and 10 principal components) are used as explanatory variables. The green color represents an increase in the adjusted R-squared brought by an inclusion of years of education, cognitive ability (IQ) and parental SES. The yellow color represents an increment in the adjusted R-squared produced by taking into account the occupational attainment measured in the previous time point. Together our variables are best at explaining the occupational education scores, especially when the first job is considered, and worst — the occupational income scores, especially in later stages of one's career (in 1992). The overall trend is that ascriptive characteristics and educational attainment explain less and less variance in occupational outcomes over time as other employee characteristics become more important. This pattern is especially true for the occupational income score. Gender differences are also the most pronounced when occupational income is concerned and the disparity grows over time: women's attainment is harder to explain with our measures than men's.

Table 4.3: Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1970 job held by the graduates with explanatory variables.

	1970 Job											
	Occupational education						Occupational income					
	(1)		(2)		(3)		(1)		(2)		(3)	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Constant	39.690***	39.189***	-91.122***	-96.488***	-30.609***	-42.717***	39.151***	13.664***	1.809	-32.056***	22.033***	-18.376***
	(0.636)	(0.810)	(2.890)	(4.451)	(3.923)	(5.240)	(0.462)	(0.446)	(2.925)	(3.175)	(3.020)	(3.853)
EA3 PGS	7.477***	7.786***	-0.293	1.488*	-0.210	1.278*	3.591***	2.879***	0.162	0.725†	-0.167	0.611
	(0.653)	(0.812)	(0.445)	(0.597)	(0.413)	(0.552)	(0.475)	(0.447)	(0.450)	(0.426)	(0.425)	(0.435)
Years of education			9.386***	10.221***	3.886***	4.795***			2.685***	3.441***	0.415†	2.173***
			(0.205)	(0.333)	(0.322)	(0.449)			(0.207)	(0.237)	(0.236)	(0.310)
Cognitive ability			2.160***	1.371*	2.122***	0.741			3.694***	0.888*	2.633***	0.672
			(0.474)	(0.621)	(0.441)	(0.566)			(0.480)	(0.443)	(0.457)	(0.446)
Parental SES			0.912*	0.520	0.952*	0.088			2.610***	-0.296	1.916***	-0.285
			(0.452)	(0.668)	(0.419)	(0.627)			(0.458)	(0.476)	(0.433)	(0.494)
Occupational score of first job					0.476***	0.488***					0.419***	0.330***
					(0.023)	(0.029)					(0.023)	(0.045)
R^2	0.053	0.070	0.608	0.553	0.671	0.637	0.024	0.033	0.217	0.220	0.327	0.251
N	2,344	1,219	2,344	1,219	2,255	1,146	2,344	1,219	2,344	1,219	2,255	1,146
	Siegel prestige											
	(1)		(2)		(3)							
	Male	Female	Male	Female	Male	Female						
Constant	46.243***	44.066***	-0.186	4.940†	8.666***	4.360†						
	(0.279)	(0.381)	(1.511)	(2.591)	(1.618)	(2.504)						
EA3 PGS	3.180***	2.996***	0.282	0.674†	0.298	0.723*						
	(0.287)	(0.382)	(0.232)	(0.348)	(0.227)	(0.333)						
Years of education			3.332***	2.942***	1.856***	1.395***						
			(0.107)	(0.194)	(0.148)	(0.213)						
Cognitive ability			1.062***	2.040***	0.810***	1.312***						
			(0.248)	(0.362)	(0.243)	(0.345)						
Parental SES			0.709**	0.550	0.596**	0.256						
			(0.236)	(0.389)	(0.230)	(0.379)						
Occupational score of first job job					0.292***	0.481***						
					(0.020)	(0.031)						
R^2	0.050	0.048	0.443	0.300	0.490	0.409						
N	2,344	1,219	2,344	1,219	2,255	1,146						

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

Table 4.4: Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1975 job held by the graduates with explanatory variables.

	1975 Job											
	Occupational education						Occupational income					
	(1)		(2)		(3)		(1)		(2)		(3)	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Constant	39.749***	35.157***	-79.658***	-82.854***	-16.915***	-20.771***	40.454***	13.449***	8.427**	-22.867***	8.414***	-4.029
	(0.595)	(0.589)	(2.881)	(3.569)	(2.686)	(3.988)	(0.449)	(0.352)	(2.875)	(2.665)	(2.267)	(3.282)
EA3 PGS	6.995***	6.792***	-0.390	1.711***	-0.171	0.689	3.550***	2.986***	0.277	1.134***	0.296	0.470
	(0.608)	(0.587)	(0.447)	(0.458)	(0.346)	(0.459)	(0.458)	(0.351)	(0.446)	(0.342)	(0.348)	(0.425)
Years of education			8.533***	8.961***	2.080***	2.590***			2.290***	2.756***	0.412*	0.900***
			(0.204)	(0.269)	(0.220)	(0.338)			(0.203)	(0.201)	(0.166)	(0.256)
Cognitive ability			2.137***	3.638***	0.518	2.262***			3.550***	2.426***	1.032**	2.352***
			(0.475)	(0.496)	(0.371)	(0.475)			(0.474)	(0.371)	(0.376)	(0.441)
Parental SES			0.653	-0.101	0.353	0.045			2.712***	-0.383	1.241***	0.065
			(0.454)	(0.506)	(0.353)	(0.517)			(0.453)	(0.378)	(0.357)	(0.479)
Occupational score of 1970 job					0.693***	0.640***					0.678***	0.535***
					(0.016)	(0.022)					(0.016)	(0.029)
R^2	0.050	0.062	0.547	0.487	0.747	0.725	0.023	0.034	0.183	0.175	0.537	0.371
N	2,527	2,040	2,527	2,040	2,343	1,192	2,527	2,040	2,527	2,040	2,343	1,192
	Siegel prestige											
	(1)		(2)		(3)							
	Male	Female	Male	Female	Male	Female						
Constant	46.429***	41.734***	6.229***	4.277*	6.393***	6.184**						
	(0.261)	(0.303)	(1.471)	(2.171)	(1.214)	(1.962)						
EA3 PGS	3.088***	2.817***	0.364	0.729**	0.232	0.446†						
	(0.267)	(0.302)	(0.228)	(0.278)	(0.187)	(0.265)						
Years of education			2.873***	2.843***	0.840***	0.583***						
			(0.104)	(0.164)	(0.102)	(0.160)						
Cognitive ability			1.297***	2.830***	0.606**	1.249***						
			(0.243)	(0.302)	(0.200)	(0.278)						
Parental SES			0.639**	0.157	0.346†	0.229						
			(0.232)	(0.308)	(0.190)	(0.299)						
Occupational score of 1970 job					0.613***	0.683***						
					(0.017)	(0.022)						
R^2	0.051	0.041	0.385	0.269	0.613	0.612						
N	2,527	2,040	2,527	2,040	2,343	1,192						

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

Table 4.5: Estimated coefficients from the OLS regressions associating occupational education, occupational income and the Siegel prestige scores of the 1992 job held by the graduates with explanatory variables.

	1992 Job											
	Occupational education						Occupational income					
	(1)		(2)		(3)		(1)		(2)		(3)	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Constant	41.011***	39.073***	-66.778***	-59.546***	-23.583***	-32.712***	42.038***	19.181***	4.482	-14.136***	0.264	-8.796**
	(6.331)	(0.524)	(2.999)	(3.223)	(2.889)	(3.644)	(0.467)	(0.387)	(2.952)	(2.834)	(2.590)	(3.228)
EA3 PGS	6.331***	5.519***	-0.432	0.240	-0.220	-0.056	2.511***	2.151***	-0.885†	-0.218	-1.059**	-0.656
	(0.604)	(0.515)	(0.469)	(0.434)	(0.396)	(0.461)	(0.476)	(0.381)	(0.462)	(0.382)	(0.404)	(0.440)
Years of education			7.628***	7.332***	2.996***	4.359***			2.658***	2.474***	1.497***	1.878***
			(0.210)	(0.238)	(0.231)	(0.296)			(0.207)	(0.209)	(0.186)	(0.244)
Cognitive ability			2.568***	3.618***	1.394**	2.684***			3.562***	3.020***	1.816***	2.848***
			(0.503)	(0.485)	(0.425)	(0.509)			(0.495)	(0.426)	(0.437)	(0.486)
Parental SES			0.635	0.610	0.090	0.776			1.893***	0.619	0.470	1.048*
			(0.479)	(0.463)	(0.404)	(0.502)			(0.471)	(0.407)	(0.415)	(0.480)
Occupational score of 1975 job					0.558***	0.393***					0.507***	0.257***
					(0.018)	(0.021)					(0.018)	(0.028)
R^2	0.043	0.043	0.488	0.402	0.641	0.511	0.011	0.012	0.178	0.128	0.377	0.174
N	2,465	2,575	2,465	2,575	2,431	1,947	2,465	2,575	2,465	2,575	2,431	1,947
	Siegel prestige											
	(1)		(2)		(3)							
	Male	Female	Male	Female	Male	Female						
Constant	46.884***	43.945***	6.352***	11.953***	2.822*	9.525***						
	(0.273)	(0.260)	(1.543)	(1.798)	(1.378)	(1.878)						
EA3 PGS	2.429***	2.071***	-0.283	-0.054	-0.438*	-0.133						
	(0.279)	(0.256)	(0.241)	(0.242)	(0.215)	(0.258)						
Years of education			2.868***	2.376***	1.543***	1.420***						
			(0.108)	(0.133)	(0.110)	(0.148)						
Cognitive ability			1.470***	2.553***	0.876***	1.804***						
			(0.259)	(0.270)	(0.231)	(0.287)						
Parental SES			0.365	0.310	-0.042	0.295						
			(0.246)	(0.258)	(0.219)	(0.281)						
Occupational score of 1975 job					0.479***	0.376***						
					(0.019)	(0.021)						
R^2	0.030	0.025	0.357	0.233	0.497	0.362						
N	2,465	2,575	2,465	2,575	2,431	1,947						

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

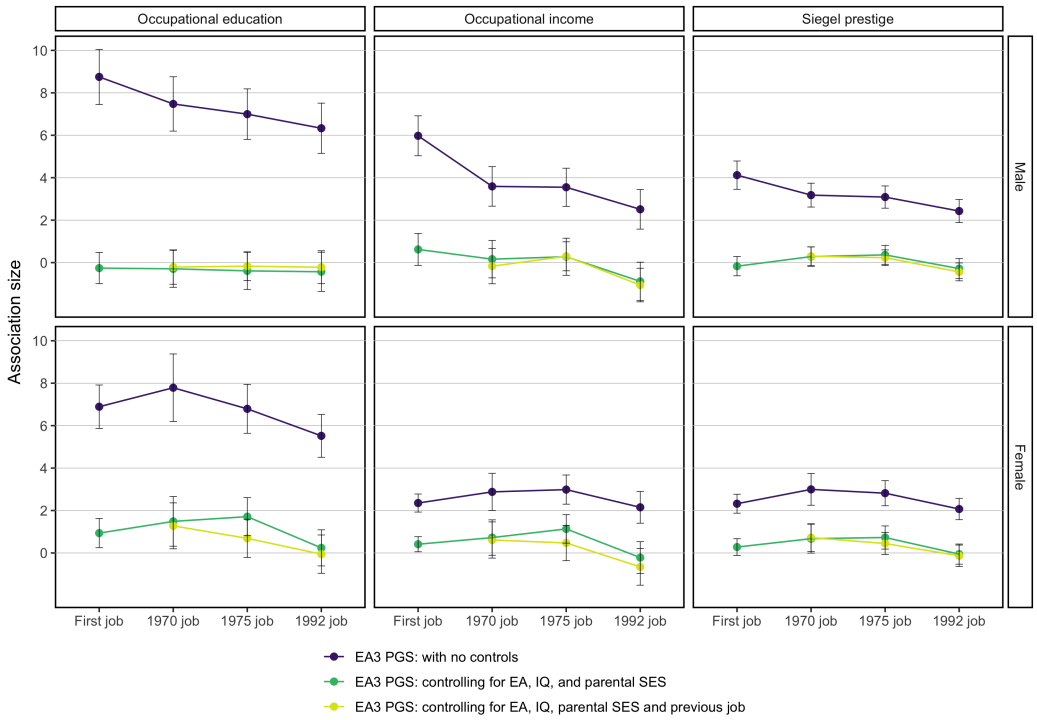


Figure 4.4: Associations between 1 standard deviation increase in the educational attainment polygenic score (EA3 PGS) and a corresponding change in occupational attainment in the WLS.

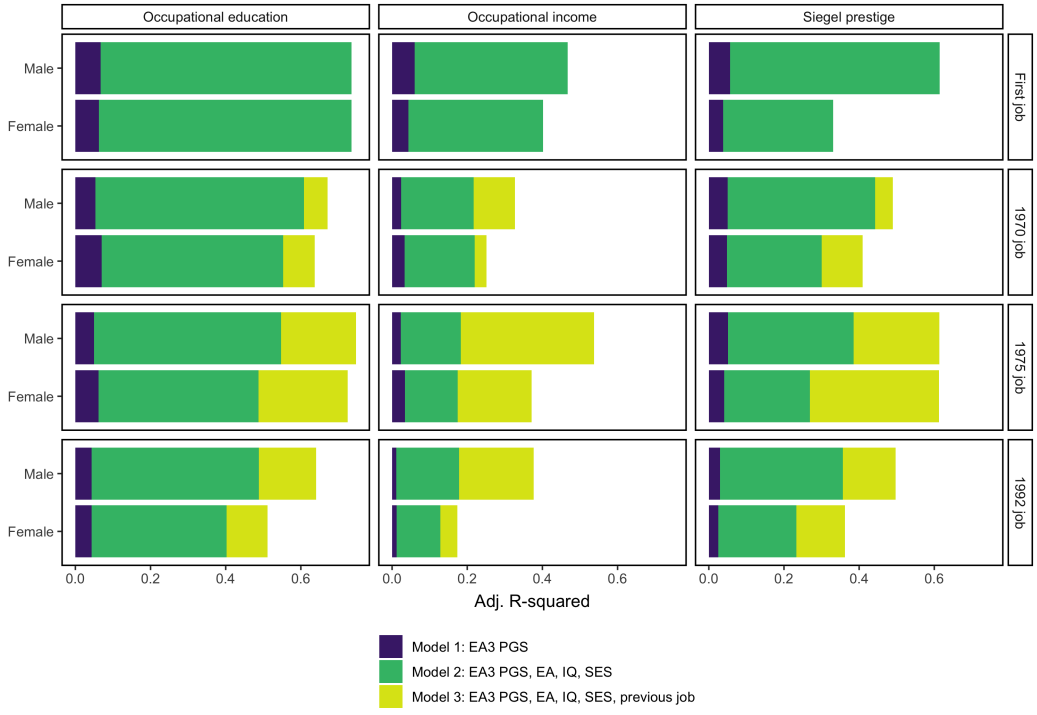


Figure 4.5: Adjusted R-squared of different OLS regressions explaining occupational attainment in the WLS.

4.7 Conclusion/Discussion

This paper aimed at bringing a genetic perspective into the research of status attainment. Its benefits are two-fold: it enriches our understanding of how occupational outcomes are produced as well as explores what is under the hood of the ever more popular educational attainment polygenic scores — whether they can solely predict educational outcomes or, in contrast, career success more broadly.

The goal of the first part of the analysis was to understand the explanatory factors of first-job attainment. First, I find that the educational attainment polygenic score is overall predictive of occupational attainment for both genders but more predictive for men than for women (Hypothesis 1a). Second, I do not find evidence to support the claim that for the success at first job the educational attainment polygenic score has an independent influence net of years of education. In other words, educational attainment mediates almost an entire association between the educational attainment polygenic score and occupational attainment (Hypothesis 1b). This finding is somewhat nuanced when gender is considered: for women the polygenic score remains statistically associated with the occupational education score and the Siegel prestige score (but not for the occupational income score). However, the remaining association size is very small. I also find support for the expected heterogeneity by attainment measurements (Hypothesis 1c). The polygenic score is best at explaining the occupational education scores probably because this measure relies on the proportion of incumbents in a specific occupation who have at least one year of college education. In other words, this occupational measure out of the three measures used is most connected to academic experiences and credentials. The occupational income and Siegel prestige scores, on the other hand, can be only moderately explained by one's genetic propensity to education as for these outcomes, theory would suggest, a more diverse set of non-academic characteristics is desired. Finally, the results show that one's origins, measured as a composite indicator of parental SES, does not correlate with occupational outcomes at first job or the educational attainment polygenic score once educational attainment is taken into account for both men and women, contrary to Hypothesis 1d. In other words, social origins and education-linked genetics do not influence first-

job attainment outcomes beyond their relationship to educational attainment. These findings are confirmed by also using sibling fixed effects.

The second part of the analysis takes an intragenerational approach to status attainment and explores its dynamics over one's life course. I find a gendered pattern in the changing importance of the educational attainment polygenic score over time: while for men it universally decreases across all three occupational measures, it increases in the case of women for the jobs held in 1970 and 1975 (Hypothesis 2a). Moreover, for both genders I find evidence for the waning influence of social origins and educational attainment as previous occupational attainment becomes more and more important (Hypothesis 2b). This finding can be explained by ascriptive characteristics and credentials losing their signaling power as respondents' skills can be more directly observed on the job. The differences stemming from different operationalisations of occupational attainment are also novel and informative (Hypothesis 2c). The role of the educational polygenic score, years of education and social origins in explaining the occupational education scores decreases but does so to a much lesser extent than for the occupational income or Siegel prestige scores. Gender differences are also worth mentioning: our variables are much more limited in their explanatory power of occupational prestige for women than for men. The explanatory disparity is reduced over time but even in 1992 remains substantial: men's occupational prestige is determined in a much more systematic way than that of women's.

These results and insights should be considered with several caveats in mind. There is a number of important limitations characteristic of most of the sociogenomic research. The genetic measurement used in this paper is imprecise and preliminary (Belsky et al., 2016). The educational attainment polygenic score is nowhere near at explaining as much variance in education as suggested by twin studies. In other words, the polygenic score potentially does not capture many important between-individual genetic differences, the phenomenon also known as missing heritability (Tropf et al., 2017). As a consequence, the estimates in this paper should probably be considered as conservatively downward-biased. At the same time it is hard to pin down what exactly these polygenic scores measure. Due to the atheoretical way in which such scores are constructed, their causal meaning is uncertain (Conley, 2016). Furthermore, because such polygenic scores are created using

GWASs that average across different populations and birth cohorts, they might not be as reflective of genetic predispositions for the WLS cohort born in 1930s. The WLS cohort is very specific: all respondents are born at the same time, more than 90% are of European descent and have mostly rural origins in Wisconsin, all have graduated from high school. Even if the polygenic score used here is appropriate for this cohort, the resulting associations between the score and the outcomes of interest might not be generalisable to other countries, cohorts, and contexts (Herd et al., 2019; Lee et al., 2018).

Genetic findings could also be spurious due population stratification, or the nonrandom patterning of allele frequency based on ancestry (Mills et al., 2020a). As the WLS sample is ethnically very homogeneous and principal components are used to account for population stratification, potential confounding by ancestry should be minimal but its possibility remains. Moreover, this paper would have benefited a lot from having not only graduates' but also their parents' genetic information. Consequently, I was not able to distinguish between environmental and genetic channels of intergenerational transmission. While sibling fixed effects analyses demonstrate an existence of direct education-related genetic effect on status attainment, not having parental genotypic information prevents me from making conclusion about the nature of nurture or the social-genetic effects that non-transmitted alleles have on childhood environments and, in turn, later life outcomes (Kong et al., 2018a). The limitations of this paper concern not only the genetic aspect of it. Sociological conclusions should also be interpreted with caution. The uncovered gender-patterns could be further complicated by the potential self-selection of women into the labour force: in 1970 and 1975 around half of the women's sample was not employed. Thus, the conclusions drawn about the role of genetics and education at those specific time points could apply only specifically to those who were in the labour force at that time. Moreover, the spacing of the occupational attainment measurements is far from ideal: occupational outcomes are observed only at four points in time and not in equal intervals. It prevented me from studying the evolution of careers more explicitly by, for example, using growth curve modeling, and having an additional analysis of how genetic predispositions potentially impact the rate of change in status attainment. Finally, the measurement of labour market entry is rather ambiguous: different individuals started their first job at different times and they could have interpreted the question about the first job differently. While

in the sample most of the respondents started their first job after high school, the possibility of differential interpretation of what is meant by "first job" still remains.

Despite of these limitations, this paper still makes important contributions and offers three take-home messages. The first message I would like to convey is that genetics matter for status attainment. Sceptics might point out that this paper finds barely any independent effect of the educational attainment polygenic score on occupational outcomes net of educational attainment. Indeed, including the polygenic score into a status attainment model would not add additional predictive power. However, sociologists have traditionally cared more about explanations rather than predictions (Goldthorpe, 2015; Hedstrom, 2005). To put it differently, an incorporation of information on genetic predispositions into traditional status attainment models enables us to better understand the mechanisms and, in this particular case, illuminates genetic antecedents of a central explanatory variable — years of education. The conclusion reached in this paper does not mean that genetics do not have a direct effect on occupational outcomes. Future research should test the role of different genetic measures in the status attainment context, for example, the income polygenic score (Hill et al., 2019).

The second take-home message concerns gender dynamics in occupational attainment over the life course. Because of women's complex relationship with the labour force, historically social stratification researchers have found it difficult to incorporate women into their studies. Potential gender differences were perceived as a nuisance that makes elegant models much messier (Acker, 1973). However, as this paper demonstrates, studying gendered patterns in occupational attainment is worthwhile as it truly demonstrates the richness and diversity of individual experiences. The WLS serves as an important case study. The late 1950s and early 1960s in the US, especially rural Wisconsin, was a conservative period when women married early and were expected to be full-time home makers (Gamoran, 2001). As a consequence, the gender differences upon the job market entry, which I discover in this paper, are probably as stark as they can be in all 20th century (Herd et al., 2019). However, what followed was a politically and culturally tumultuous time characterised by rapid shifts in social norms and a large-scale absorption of women into the labour force (DiPrete and Buchmann, 2006, 2013). The patterns we see in the WLS data, thus,

are not simply a function of life course but also of changing structural factors. In a way, then, the WLS provides us with an opportunity to observe gene-environment interaction, environment being social contexts representing ever-changing conceptions, expectations, and treatments of gender. A fruitful direction for future research would be to use later cohorts and explore similar questions to the ones asked in this paper. That way competing hypotheses of life-course forces versus structural changes might be ascertained.

The third and final message this paper sends is that, instead of avoiding it, sociologists should embrace the fact that there can be many ways of measuring the same social concept. As explained by Haug (1977), social stratification theories are often grand, contradictory and difficult to operationalise. Although the trend in sociology has been to use composite indicators when possible, disaggregated measures, each capturing a different facet of social phenomena, can contribute to our understanding of the complexities inherent in the world (Ganzeboom et al., 1991). This insight is particularly salient when studying gender differences in occupational attainment. Since occupational distributions are different for men and women, the usage of existing class or status categories can be seen as problematic (Bielby and Baron, 1986). However, instead of looking for one perfect measure which would be relevant for both genders and which would not conceal gender differences, the solution could be to instead use a rich set of measures (Belsky et al., 2016). An incorporation of three different measures in this paper — the occupational education, occupational income and Siegel prestige scores — provides evidence on the multidimensionality of occupational attainment. Future research in this area could carry this multidimensionality a step further by studying not just directly observable characteristics of occupations but individuals' subjective feelings. Inglehart (1977) has demonstrated that in developed societies, as subsistence becomes more or less guaranteed, non-material goals gain more importance. Occupational attainment and real career success in present days, thus, extend beyond income, prestige and employment relations to include job satisfaction and general well-being, occupation-related health, flexibility and autonomy, or intrinsic motivation (Clark et al., 2018). A worthwhile future endeavour would be to study an interplay of genetics and environments in producing disparities in these post-material occupational outcomes.

4.A Appendix

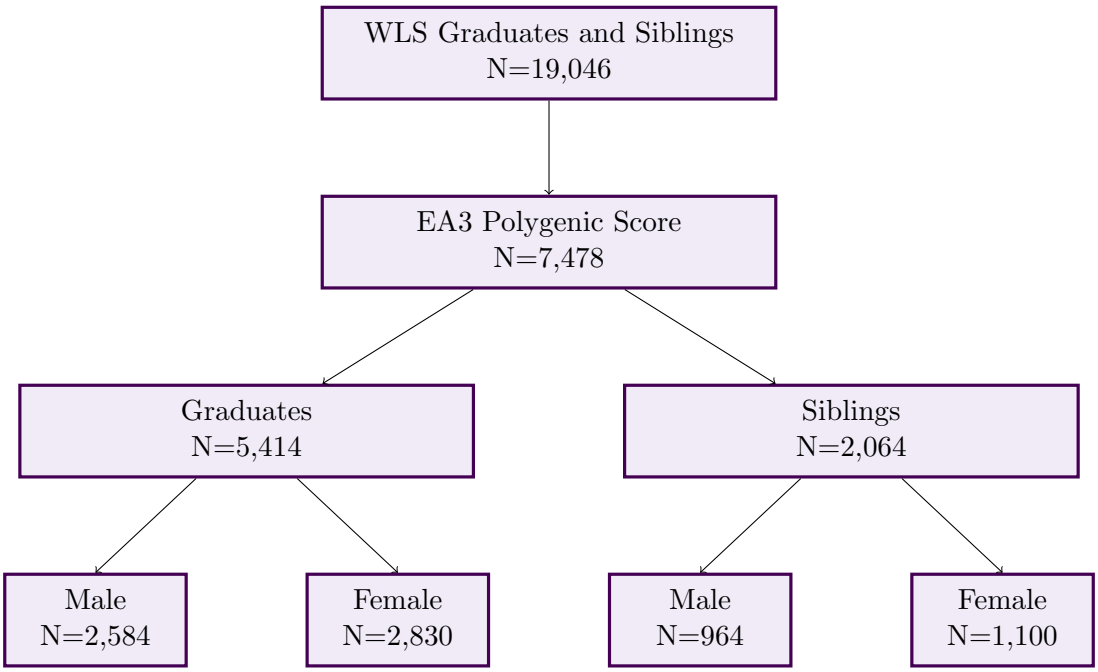


Figure 4.A.1: Sample reduction by gender and respondent type (graduate or sibling) in the WLS.

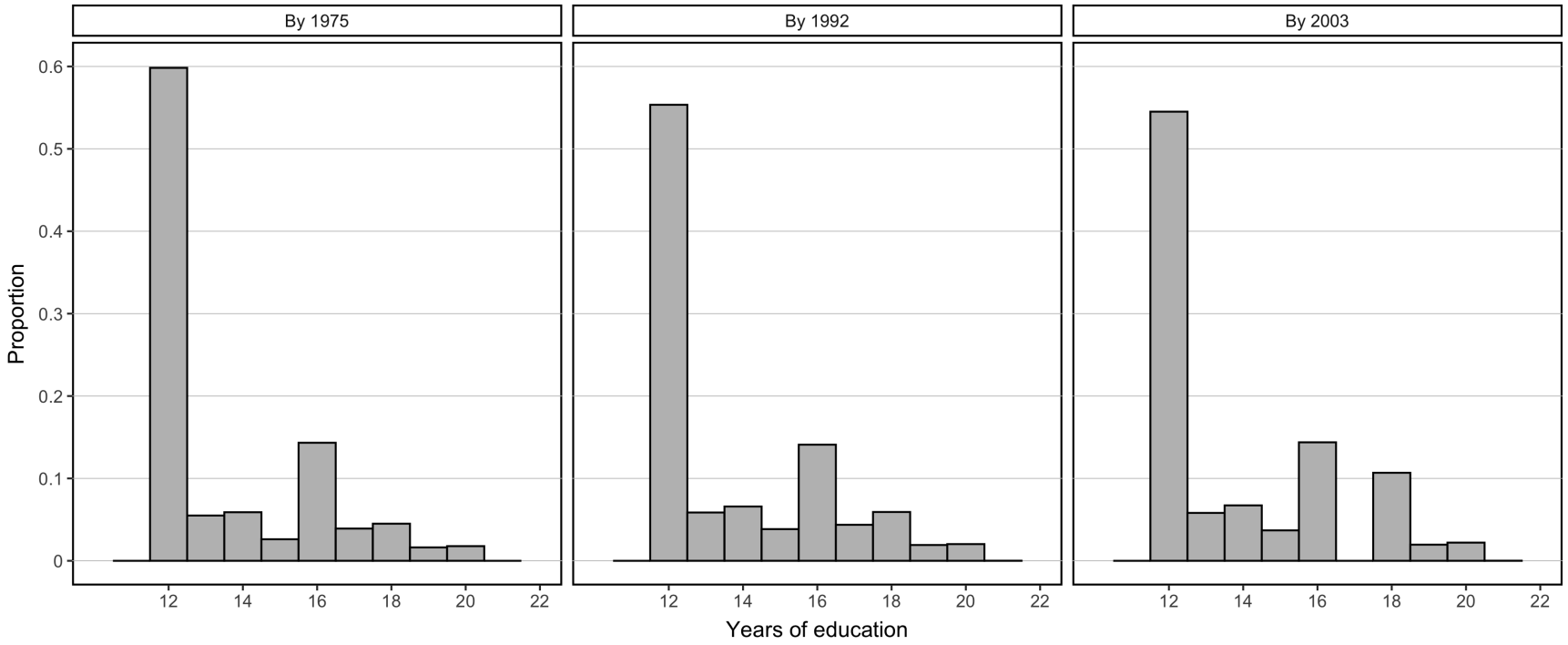


Figure 4.A.2: Distributions of years of education attained by three time points — 1975, 1992, and 2003 — in the WLS.

4.A.1 Operationalisation of educational attainment

1975 interview.

- High school graduate or less, or less than one year of college — 12 years
- One-year college certificate, one year of college (no degree or certificate) — 13 years
- Two-year teaching certificate, associate degree, two-year diploma or certificate, two years of college (no degree or certificate) — 14 years
- College trained registered nurses, one or more years post two-year degree (but no higher degree), three or more years of college (no degree or certificate) — 15 years
- Bachelor's degree — 16 years
- Master's degree with major field: continuing/vocational education, curriculum, education, elementary education, library science, or one or more years post-bachelor's degree (no higher degree or certificate) — 17 years
- Two-year master's (with major other than those above), AE, OD, EdS, DVM, DC, MDiv, ThB or one or more years post one-year master's (no higher degree or certificate) — 18 years
- DDS, LLD, LLB, JD, or one or more years post two-year-master's level degree (no higher degree or certificate) — 19 years
- PhD, MD, other doctorates not previously included, or one or more years post-professional degree — 20 years

1992 interview.

- High school graduate or less, or less than one year of college — 12 years
- One-year college certificate, one year of college (no degree or certificate) — 13 years

- Two-year teaching certificate, associate degree, two-year diploma or certificate, two years of college (no degree or certificate) — 14 years
- College trained registered nurses, one or more years post two-year degree (but no higher degree), three or more years of college (no degree or certificate) — 15 years
- Bachelor's degree — 16 years
- Master's degree with major field: continuing/vocational educ, curriculum, education, elementary education, library science, or one or more years post-bachelor's degree (no higher degree or certificate) — 17 years
- Two-year master's with major other than those above), AE, OD, EdS, DVM, DC, MDiv, ThB or one or more years post one-year master's (no higher degree or certificate) — 18 years
- DDS, LLD, LLB, JD, or one or more years post two-year-master's level degree (no higher degree or certificate) — 19 years
- PhD, MD, other doctorates not previously included, or one or more years post-professional degree — 20 years
- Post-doctorate education — 21 years

2003 interview.

- Completed High School / less than one year of college — 12 years
- One year of college — 13 years
- Associate's degree / two years of college — 14 years
- Three or more years of college — 15 years
- Bachelor's degree — 16 years
- Master's degree — 17 years
- Two-year master's — 18 years

- Professional degrees, one or more years post-two-year-master's — 19 years
- PhD, MD — 20 years
- Post-doctorate education — 21 years

	Graduates						Siblings					
	Male			Female			Male			Female		
Variables	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
EA3 PGS	-0.01	0.98	2,584	-0.001	1.02	2,830	0.03	1.00	964	0.04	1.02	1,100
Occupational education of first job	33.25	32.98	2,420	33.69	27.71	2,592	35.85	33.47	785	37.95	30.50	884
Occupational education of 1970 job	39.59	31.62	2,344	39.32	29.33	1,219	-	-	-	-	-	-
Occupational education of 1975 job	39.63	30.69	2,527	35.09	27.44	2,040	39.25	30.60	767	37.82	29.25	614
Occupational education of 1992/1993 job	40.94	30.01	2,466	39.12	27.16	2,575	40.51	30.17	897	41.01	28.20	1,040
Occupational income of first job	26.91	23.76	2,420	9.16	11.40	2,592	30.05	23.77	785	13.12	16.25	884
Occupational income of 1970 job	39.10	22.65	2,344	13.71	15.83	1,219	-	-	-	-	-	-
Occupational income of 1975 job	40.40	22.81	2,527	13.42	16.16	2,040	39.30	23.22	767	15.12	17.64	614
Occupational income of 1992/1993 job	42.00	23.31	2,466	19.20	19.77	2,575	40.19	23.64	897	20.02	20.59	1,040
Years of education (measured in 1993/1994)	14.13	2.54	2,480	13.41	2.04	2,753	14.31	2.71	898	13.62	2.34	1,050
IQ	102.14	15.10	2,584	102.29	14.33	2,830	103.82	16.34	813	104.80	14.91	932

Table 4.A.1: Descriptive statistics by respondent type (graduate or sibling) and gender (male or female) of the WLS.

Sibling pair	Count
Female-Female (F-F)	571
Female-Male (F-M)	519
Male-Female (M-F)	529
Male-Male(M-M)	445
Total	2,064

Table 4.A.3: Counts of sibling pairs by gender composition in the WLS.

Measures	Data source
Occupational attainment	
1970 Occupational education of first job	1975 Survey
1970 Occupational education of 1970 job	1975 Survey
1970 Occupational education of 1975 job	1975 Survey
1970 Occupational education of 1992/1993 job	1992/1993 Survey
1970 Occupational earnings of first job	1975 Survey
1970 Occupational earnings of 1970 job	1975 Survey
1970 Occupational earnings of 1975 job	1975 Survey
1970 Occupational earnings of 1992/1993 job	1992/1993 Survey
Educational attainment	
Highest degree by 1975	1975 Survey
Highest degree by 1992/1993	1992/1993 Survey
Occupational aspirations	
1970 Duncan SEI for 1957 aspiration	1957 Survey
1970 Duncan SEI for job to be held after graduation	1957 Survey
1970 Occupational education for job to be held after graduation	1957 Survey
1970 Occupational earnings for job to be held after graduation	1957 Survey
Educational aspirations	
Plans after high school graduation	1957 Survey
Significant others' influence	
Parents' thoughts about college	1957 Survey
Teachers' encouragement to attend college	1957 Survey
Friends' plans after high school	1957 Survey
Academic performance	
Rank-normalised percentile of high school grades	1957 Survey
Social background	
Factor-weighted SES score for parents in 1957	1966 Tax data
Father's education in 1957	1975 Survey
Mother's education in 1957	1975 Survey
Father's 1970 Duncan SEI of 1957 job	1975 Survey
Head of household's 1970 occupational education of 1957 job	1975 Survey
Head of household's 1970 occupational earnings of 1957 job	1975 Survey
Parent's income 1957	1975 Survey
Farm background	1966 Tax data
Non-intact family	1975 Survey
Number of siblings	1975 Survey
Cognitive ability	
Henmon-Nelson test score from junior year	Wisconsin State Testing Service

Table 4.A.2: Traditional measures of the Wisconsin status attainment model in the WLS.

4.A.2 Additional analysis: Women and labour force participation

I use the educational attainment polygenic score, years of education and parental SES explain participating in the labour force among women. Labour force participation is measured as a binary variable (0 — woman is not employed at that time, 1 — is employed). I use logistic regression to model it. As was seen in Figure 4.2, a small proportion of women (similar to that of men) had not been employed full-time in or before 1964 (when the question was asked). My hypothesis is that such women were at the time still pursuing higher education. Thus, higher levels of educational attainment (measured later), the educational attainment and higher parental SES should all predict being at an educational institution in 1964 and, as a consequence, not being in the labour force at that time. Table 4.A.4 confirm this hypothesis: odds ratios are lower than 1. A much higher proportion of women were out of the labour force by 1970 (Figure 4.2). Years of education, controlling for parental SES, increases the odds of being in the labour force in 1970. Higher SES decreases such odds: it can be explained by the tendency of higher-status women to stay home when raising kids. For 1975 jobs only parental background has an effect on woman's propensity to not work. By 1992 most of the women have come back to the labour force (Figure 4.2 and in this context only years of education seems to be predictive of working: the higher the educational attainment, the more likely a woman is to be part of the labour market, controlling for her parental SES. Although statistically significant, these predictors are not that substantial. In other words, I do not find that two sets of predictors — education (measured as years of education and the polygenic score) and social background (which should be expected to be systemic differentiators between women who work and do not work in the 1970s) in our dataset are not strong explanatory factors.

Table 4.A.4: Estimated coefficients (odds ratios) from the logistics regressions associating women's propensity of not being in the labour force with explanatory variables.

	First job			1970 job			1975 job			1992 job		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Constant	13.741***	3355***	2445	0.817***	0.279***	0.167***	2.640***	2.260**	1.540	15.300***	2.530	2.020
	(1.050)	(1716)	(1332)	(0.032)	(0.080)	(0.053)	(0.111)	(0.708)	(0.533)	(1.210)	(1.560)	(1.360)
EA3 PGS	0.743***	0.961	0.974	1.020	0.978	0.988	0.973	0.968	0.975	1.160*	1.090	1.100
	(0.054)	(0.075)	(0.076)	(0.039)	(0.039)	(0.039)	(0.040)	(0.042)	(0.042)	(0.090)	(0.088)	(0.089)
Years of education		0.670***	0.687***		1.090***	1.130***		1.010	1.040		1.140**	1.160**
		(0.024)	(0.026)		(0.024)	(0.027)		(0.024)	(0.027)		(0.054)	(0.060)
Parental SES			0.884†			0.844***			0.882**			0.924
			(0.061)			(0.037)			(0.041)			(0.083)
N	2,788	2,788	2,788	2,710	2,710	2,710	2,814	2,814	2,814	2,750	2,750	2,750

Notes: EA3 PGS = educational attainment polygenic score. Parental SES = parental socio-economic status.
Data source: the Wisconsin Longitudinal Study.

4.A.3 Additional analysis: residuals

Model 1 (purple) boxplots represent residuals obtained when regressing occupational education on years of education. Model 2 (green) boxplots represent residuals obtained when regressing Model 1 residuals on the educational attainment polygenic score. We can see that the additional benefit from using the polygenic score on top of the educational attainment is very limited.

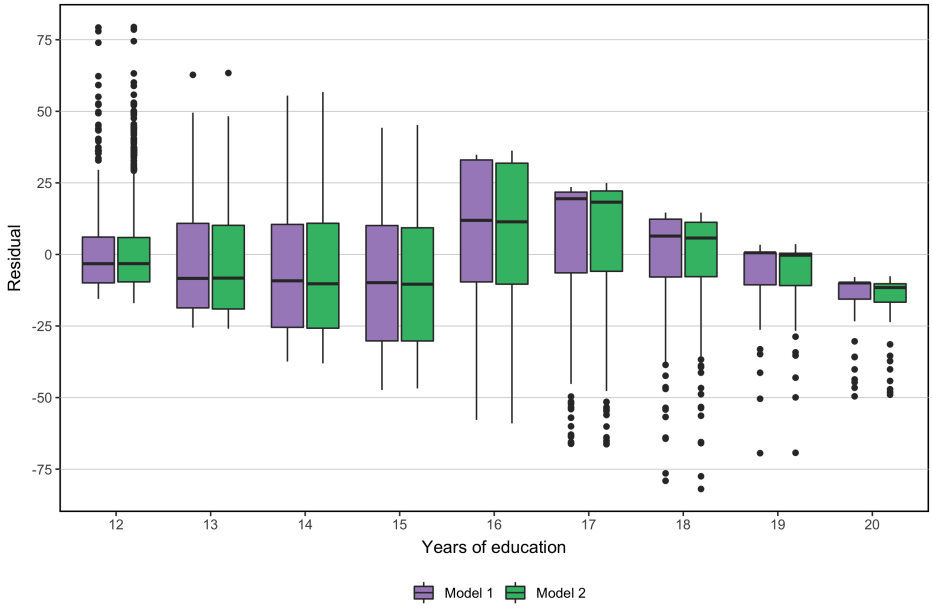


Figure 4.A.3: Boxplots of residuals from the OLS regressions explaining differences in the occupational education scores with educational attainment (purple) and explaining differences in the occupational education score with educational attainment and the educational attainment polygenic score (green).

Chapter 5. Conclusion

My thesis aimed at better explaining disparities in educational attainment in the US by using an individual's genetic predisposition to education, measured as the educational attainment polygenic score. In particular, this dissertation tried to connect the educational genetic propensity with parental involvement and the development of non-cognitive skills. During the COVID-19 era, these topics have gained special relevance. With many children across the globe now studying remotely, the time resources of their parents are more important than ever. Non-cognitive skills, such as delayed gratification, grit, or perseverance, are also essential in the present context of little teacher supervision and face-to-face engagement. The conclusion of the thesis will outline its main contributions, summarise the findings, enumerate the most important limitations, and discuss possible directions for future research.

5.1 Main contributions

This thesis makes several contributions to the existing literature on both educational stratification and social science genetics. First, I demonstrated that genetics is another important mechanism that produces inequalities in educational attainment. Unlike much of the published work in the research area, I did not start with the polygenic score and assign some sociological meaning to it but instead started from the rich sociological tradition, exemplified by the Wisconsin status attainment model, and embedded the genetic aspect of attainment within its existing frameworks. Second, I explored novel intermediate social channels which connect genetic propensity to education and educational attainment: parental involvement and non-cognitive skills. Social scientists often justifiably avoid such "soft" measures because of an inherent difficulty of operationalising them. My approach was to create different dimensions of such measures instead of trying to produce a single best measurement. My results, which were indeed sensitive to how I chose to measure key constructs, confirm the multidimensionality of the social world and demonstrate that the success of any endeavour in social science genetics depends not only on the quality of the genetic methodology but also the sociological one. Third, I complemented the analysis between different individuals with within-family modelling. It served as both a sensitivity check for the between-family models and also provided new knowledge on the dynamics of genetic predisposition in family units. Fourth, I

showed that educational attainment is a continuous experience by taking a life-course perspective to its association with genetic propensity to education. I started from childhood where I looked at the relationship between genetics and parental involvement and, in turn, educational outcomes. I also looked at how non-cognitive skills are produced in early life. Moreover, I investigated occupational attainment at different points in time as a product of a complex interplay between genetics, education, and social background which continues throughout one's life. Fifth, I highlighted the different experiences women and men have with regards to their genetic predisposition to education both in developing the skills and turning those skills into status outcomes. It enriches our knowledge on the heterogeneity of social and genetic mechanisms operating in the context of educational and occupational attainment.

5.2 Summary of findings

In this thesis I sought to address three main sets of questions. The first question is the following: To what extent can parental involvement in adolescence explain eventual educational attainment? Intensive forms of parenting amongst the middle classes have become very popular and are often prescribed as a necessity for child's success (Doepke et al., 2019). However, the causal link between the specific every-day actions parents take to be involved and their children's academic success is questionable as it might be confounded by the family's socioe-conomic background. Moreover, genetics could also be at play as child's genetic predispositions might inform the nature of parental involvement she receives (Harris, 1998). To explore these nuances I use the Add Health dataset which has a rich set of measures on parental involvement. Drawing on the existing literature, I operationalise it as having three dimensions: school-based activities which parents can choose to participate in, academic socialisation pertaining to future expectations and school-work discussions with a mother and a father separately, and parental control. I find that the association between different dimensions of parental involvement and educational attainment are greatly reduced after taking into account child-level characteristics such as age, gender, grades, cognitive ability, self-efficacy and behavioural problems as well as family-level factors such as parental education, income, or father's absence. The predictive power of parental involvement after controlling for

these measures is then lower than many of the above-mentioned variables, calling into question the importance of the particular parental activities studied in this chapter for child's later academic outcomes. I also find little evidence that child's genetic propensity to education in any way evokes parental responses, at least those with which my thesis is concerned with.

The second set of question I attempted to answer is this: Do non-cognitive skills explain some part of the association between genetic propensity to education and educational attainment and does this mechanism differ depending on the individual's social circumstances and personal characteristics? We already know that cognitive skills are to some extent genetically predisposed and that they, in turn, have an impact on educational outcomes (Lee et al., 2018). However, much less is know about another set of determinants — non-cognitive skills. To investigate their role in mediating the association between the educational attainment polygenic score and the educational attainment I use the Wisconsin Longitudinal Study which has an extensive questionnaire on personality and other non-cognitive measures. I find that among five non-cognitive skills studied in this chapter (conscientiousness, openness, delayed gratification, perseverance, self-efficacy, self-esteem), only openness seems to be a relatively important channel of genetic predisposition to education. All five non-cognitive measures together could explain 20% of the variance in educational attainment — half of what can be explained by cognitive skills measured as a single measure of IQ. Furthermore, I discover that for women the mediation channel through non-cognitive skills is more significant than for men. When it comes to socio-economic background, individuals from both lower and higher SES tend to derive benefits from non-cognitive skills at similar levels.

The third set of research questions in this dissertation is the following: Is the power of genetic propensity to education limited to the realm of education or does it also extend to occupational attainment? Do men and women experience it differently over the life course? It has been shown that the explanatory potential of the educational attainment polygenic score can go beyond strictly educational outcomes to be predictive of one's life chances and success more broadly (Belsky et al., 2018). What has been missing is the different aspects of occupational attainment and its potentially differential relationship with the genetic predisposition to education. To fill this gap I use the Wisconsin Longitudinal Study and open the famous Wisconsin status attainment model. Instead

of focusing on a single measure of occupational attainment I operationalise it with three separate indicators available in the dataset: occupational education score, occupational income score, and Siegel prestige index. My results show that one's genetic propensity to education is related the most to the occupational education score of a job compared to other two measures. Moreover, I also find that while for men the association between the polygenic score and occupational attainment is entirely explained by the intermediate mechanism of educational attainment, for women, as their careers progress, the genetic propensity maintains some independent power over their job status net of educational credentials. This finding shows how the genetic propensity, fixed at birth, is turned into a dynamic factor as it interacts with the social environment.

5.3 Limitations

The conclusions drawn in this dissertation should be evaluated with several limitations in mind. The first limitations stems from the difficult task of properly conceptualising and operationalising parental involvement and non-cognitive skill measures. The literature on these topics span multiple disciplines — economics, psychology, sociology — with each having its own traditional ways of approaching these rather abstract concepts. Even if the Add Health and the WLS serve as probably the best datasets for the respective topics, they are still fundamentally limited in what questions get asked during the interviews and how the different items are measured. Both the number and the choice of dimensions to be included in my chapters, although informed by the existing literature, can be seen as arbitrary. Furthermore, the items to be included in each indicator and how each of the items is to be weighted (or not weighted) within a given indicator also pose unique challenges. The results reported in this thesis, thus, should not be extrapolated to many other aspects of parental involvement or non-cognitive skills.

The second limitation concerns the fact that strong causal claims cannot be made in this thesis because for many important measures the data is not longitudinal. Parental involvement is measured only in adolescence and, thus, is by definition contaminated by earlier parental involvement and unmeasured child characteristics and behaviours. Non-cognitive skills are also measured at only one

point in time and quite late in one's life making it hard to disentangle the effects of early life versus later career towards the development of these skills. As a consequence, strong assumptions, in most cases supported by the existing literature but still debatable, had to be made about the stability or temporal ordering of these measures. The longitudinal career data are also not as extensive as one would like them to be. Measured at four times in one's life, it prevents me to track all career moves, especially for women who can leave and reenter job market multiple times.

It is also worth mentioning the third limitation that all three empirical chapters have in common — the sample selection bias. My samples in both the Add Health and the WLS datasets include only the individuals who agreed to have their genetic information collected. It is likely that the respondents who opted into the collection are on average different from the ones who did not, for example, on their education and socio-demographic characteristics. In Chapter 2 my sample includes only individuals whose parents or the individuals themselves completed questionnaires about parenting. We can expect such individuals to be more involved and engaged in general. The same concern surrounds the completion of the non-cognitive skill questionnaire in the WLS: the individuals who agreed to complete the questionnaire are likely to have higher non-cognitive skills. When studying occupational attainment over the life course the sample selection issue is even more pressing as at several points in time we observe only women who chose to stay in the labour market. More generally, because I included the educational attainment polygenic score, in my analysis I was restricted to the sample of European ancestry. The characterisation of the WLS as having a very homogeneous, rural and geographically constrained profile also calls into question the generalisability of the findings. It is also worth mentioning that my research questions and the hypotheses are inspired by the sociological literature focusing on the US in addition to the data I use coming from this country. The specifics of educational and occupational attainment as well as the related pressure on the intensity of parenting might not translate neatly in other contexts.

Finally, we should be aware of the limitations of the polygenic scores in their current form despite the attempts to make them more transparent. To begin with, polygenic scores serve as direct measures of genotypes but for now they capture only a small portion of presumed genetic effects. The estimates of GWAS or SNP heritability are nowhere close to the heritability estimates pro-

duced using twin studies, This discrepancy is known as missing heritability (Tropf et al., 2017). For example, GWASs might be unable to identify genetic variants which are rare in a given study population (Cesarini, 2016). Moreover, GWASs might be detecting only what is most common across different populations while disregarding population-specific causal variants (Conley, 2016). A related problem is, thus, a lack of diversty in the samples which are incorporated into GWASs (Mills and Rahal, 2019b; Mills and Tropf, 2020; Popejoy and Fullerton, 2016). GWASs also often suffer from flawed measurement and harmonisation of metrics (Mills and Tropf, 2020). All in all, polygenic scores likely capture an averaged out and potentially diluted version of genetic propensity. The low portability of the polygenic scores across populations of different ancestries and in different geographies is a concern (Privé et al., 2021). Understanding both the individual- and the population-level heterogeneity is essential in order to ascribe polygenic scores a true causal meaning.

5.4 Areas of future research

This dissertation could be extended in several ways. Social scientists are always on the lookout for natural experiments to strengthen their causal claims. The COVID-19 pandemic that we are still living through is in itself an unprecedented social experiment which has already altered the educational landscape. One avenue of research could be to look at whether the role of genetic propensity to education changes due to the pandemic-induced changes in the teaching model. Do the genetic disparities get amplified by the lack of face-to-face teacher-student contact or, on the contrary, do the genetic advantages get suppressed? Differences between countries could also be explored as in each the educational system was impacted in different ways. The transition period out of the pandemic might also serve as an experimental ground which social scientists can use to test various hypotheses. Even within the same country such as the US it is probable that different states will choose a different teaching model at schools for the next academic year: some will offer only an in-class option while others will choose a hybrid model. Consequently, the role of parental involvement and non-cognitive skills of a student and how they connect with the genetic propensity and academic outcomes could be explored in a quasi-experimental way. Another interesting line

of research would be to go beyond educational attainment and study the relationship between the genetic propensity to education and other educational outcomes such as achievement at various points in an academic career. Moreover, the disparities in the progression of learning both at school and in a professional setting could be connected to the educational attainment polygenic score which likely encapsulates the aspect of brain plasticity. Relatedly, not only the outcome variables could be richer but also the set of polygenic scores developed, used, and compared could be broader. Throughout this dissertation I used a specific polygenic score — the educational attainment polygenic score derived using the summary statistics from the third educational attainment GWAS. However, there already exist GWASs and corresponding polygenic scores for cognitive and non-cognitive skills. GWASs for outcomes other than educational attainment might also become available. Rather than seeing them as competing with each other for higher prediction or occupying separate knowledge domains, these different polygenic scores could be studied together to explore their interplay and build a more detailed picture of the genetic transmission of advantages. Finally, researchers in the future should aim at building multigenerational family datasets as they would allow those working in the area of social science genetics to have much better research designs. Having information not only of children but of their parents would allow to better understand the genetic origins of parental involvement. Moreover, the genetic nurture, or the phenomenon of non-transmitted alleles having an effect on the family environment and, in turn, children's outcomes, could be rigorously empirically studied. All in all, genetic and social information of family members from multiple generations could also help us to further understand the interaction of genetics and the ever-changing social aspects of the world.

Bibliography

Acker, J. (1973). Women and Social Stratification: A Case of Intellectual Sexism. *American Journal of Sociology 78*(4), 936–945.

Ackerman, P. L. and E. D. Heggestad (1997). Intelligence, Personality, and Interests: Evidence for Overlapping Traits. *Psychological Bulletin 121*(2), 219–245.

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020). It Takes a Village: The Economics of Parenting with Neighborhood and Peer Effects. *Working Paper*.

Aizer, A. and F. Cunha (2012). The Production of Human Capital: Endowments, Investments and Fertility. *NBER Working Paper Series*.

Alan, S., T. Boneva, and S. Ertac (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Allen, M. L. et al. (2016). Effective Parenting Interventions to Reduce Youth Substance Use: A Systematic Review. *Pediatrics 138*(2), e20154425.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. D. Kautz (2011). Personality Psychology and Economics. *NBER Working Paper Series*.

Almond, D. and J. Currie (2009). Human Capital Development Before Age Five. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, pp. 1315–1486. Amsterdam: North Holland.

Alon, S. and T. A. DiPrete (2014). Gender Differences in the Formation of a Field of Study Choice Set. *Sociological Science 2*, 50–81.

Anderson, R. G. (1967). On Genetics and Sociology. *American Sociological Review 32*(6), 997–999.

Andrisani, P. J. and G. Nestel (1976). Internal-External Control as Contributor to and Outcome of Work Experience. *Applied Psychology 61*(2), 156–165.

Auerbach, S. (2007). Visioning Parent Engagement in Urban Schools. *Journal of School Leadership 17*(6), 699–734.

Autor, D. H., L. F. Katz, and M. S. Kearny (2008). Trends in US Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Baker, S., M. R. Sanders, K. M. T. Turner, and A. Morawska (2017). A Randomized Controlled Trial Evaluating a Low-Intensity Interactive Online Parenting Intervention, Triple P Online Brief, with Parents of Children with Early Onset Conduct Problems. *Behaviour Research and Therapy 91*, 78–90.

Baldwin, A. L., C. Baldwin, and R. E. Cole (1990). Stress-Resistant Families and Stress-Resistant Children. In J. E. Rolf, A. S. Masten, K. H. Cicchetti, K. H. Nuechterlein, and S. Weintraub (Eds.), *Risk and Protective Factors in the Development of Psychopathology*. Cambridge: Cambridge University Press.

Bandura, A. (1982). Self-Efficacy Mechanism in Human Agency. *American Psychologist 37*(2), 122–147.

Bandura, A. (1989). Social Cognitive Theory. In R. Vasta (Ed.), *Annals of Child Development*, pp. 1–60. Greenwich, CT: JAI Press.

Barban, N. et al. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behaviour. *Nature Genetics 48*, 1462–1472.

Baron, R. M. and D. A. Kenny (1986). The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology 51*(6), 1173–1182.

Barone, C., M. Lucchini, and A. Schizzoretto (2011). Career Mobility in Italy. *European Societies 13*(3), 377–400.

Bartholomew, D. J. (2004). *Measuring Intelligence: Facts and Fallacies*. New York, NY: Cambridge University Press.

Baselmans, B. M. L. et al. (2019). Multivariate Genome-Wide Analyses of the Well-Being Spectrum. *Nature Genetics 51*(3), 445–451.

Batty, G. D., I. J. Deary, and L. S. Gottfredson (2007). Premorbid (Early Life) IQ and Later Mortality Risk: Systematic Review. *Annals of Epidemiology 17*(4), 278–288.

Baumeister, R. F. (1993). *Self-Esteem: The Puzzle of Low Self-regard*. Boston, Massachusetts: Springer.

Baumrind, D. (1966). Effects of Authoritative Parental Control on Child Behavior. *Child Development 37*(4), 887–907.

Beauchamp, J., D. Cesarini, and M. Johannesson (2017). The Psychometric and Empirical Properties of Measures of Risk Preferences. *Journal of Risk and Uncertainty 54*, 203–237.

Becker, G. S. (1964). *Human Capital*. New York, NY: National Bureau of Economic Research.

Becker, G. S. and N. Tomes (1976). Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy 84*, S143–S162.

Becker, G. S. and N. Tomes (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy 87*(6), 1153–1189.

Behrman, J. R., R. Pollak, and P. Taubman (1982). Parental Preferences and Provision for Progeny. *Journal of Political Economy 90*, 52–73.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994). Endowments and the Allocation of Schooling in the Family and in the Marriage Market. *Journal of Political Economy 102*(6), 1131–1174.

Belley, P. and L. Lochner (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital 1*(1), 37–89.

Belsky, D. W. et al. (2018). Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies. *PNAS 115*(31), E7275–E7284.

Belsky, D. W. and S. Israel (2014). Integrating Genetics and Social Science: Genetic Risk Scores. *Biodemography and Social Biology 60*(2), 137–155.

Belsky, D. W., T. E. Moffitt, D. L. Corcoran, B. W. Domingue, H. Harrington, S. Hogan, R. Houts, S. Ramrakha, K. Sugden, B. S. Williams, R. Poulton, and A. Caspi (2016). The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-Course Development. *Psychological Science 27*(7), 957–972.

Benjamin, D. J. et al. (2012). Promises and Pitfalls of Genoeconomics. *Annual Review of Economics 4*, 627–662.

Benner, A. D., A. E. Boyle, and S. Salder (2016). Parental Involvement and Adolescents' Educational Success: the Roles of Prior Achievement and Socioeconomic Status. *Journal of Youth and Adolescence 45*, 1053–1064.

Berg, I. (1970). *Education and Jobs*. New York: Praeger.

Bharadwaj, P., J. Eberhard, and C. Neilson (2018). Health at Birth, Parental Investments, and Academic Outcomes. *Journal of Labor Economics 36*(2), 349–394.

Bielby, W. T. and J. N. Baron (1986). Men and Women at Work: Sex Segregation and Statistical Discrimination. *American Journal of Sociology 91*(4), 759–799.

Billis, D. B. (1988). Educational Credentials and Hiring Decisions: What Employers Look For in New Employees. *Research in Social Stratification and Mobility 7*, 71–97.

Bjorklund, A., M. Jantti, and G. Solon (2005). Influence of Nature and Nurture on Earnings Variation. In S. Bowles, H. Gintis, and M. Osborne Groves (Eds.), *Unequal Chances: Family Background and Economic Success*, pp. 145–164. Princeton, NJ: Princeton University Press.

Black, S., P. Devereux, P. Lundborg, and K. Majlesi (2019). Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviors. *NBER Working Paper Series*.

Black, S. E., E. Gronqvist, and B. Ockert (2017). Born to Lead? The Effect of Birth Order on Non-Cognitive Abilities. *IZA Discussion Paper Series*.

Blanchet, T., L. Chancel, and A. Gethin (2019). How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017.

Blank, R. M. (2011). *Changing Inequality*. Berkeley: University of California Press.

Blau, F. and L. M. Kahn (2000). Gender Differences in Pay. *Journal of Economic Perspectives 14*(4), 75–99.

Blau, F. D., P. Brummund, and A. Y.-H. Liu (2013). Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System. *Demography 50*(2), 471–494.

Blau, P. M. and O. D. Duncan (1967). *The American Occupational Structure*. New York: Wiley.

Blossfeld, H. P. (1986). Career Opportunities in the FRG: A Dynamic Approach to the Study of Life Course, Period and Cohort Effects. *European Sociological Review 3*, 123–155.

Boardman, J. D., C. L. Blalock, and F. C. Pampel (2010). Trends in the Genetic Influences on Smoking. *Journal of Health and Social Behavior 51*(1), 108–123.

Bodner, T. E. (2008). What Improves with Increased Missing Data Imputations? *Structural Equation Modeling: A Multidisciplinary Journal 15*(4), 651–675.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources 43*, 972–1059.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What Grades and Achievement Tests Measure. *PNAS 113*(47), 13354–13359.

Bourdieu, P. and J.-C. Passerson (1977). *Reproduction in Education, Society and Culture*. London & Beverly Hills: Sage Publications.

Bowles, S. and H. Gintis (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. New York, NY: Basic Books.

Bowles, S., H. Gintis, and M. Groves Osborne (2005). *Unequal chances: family background and economic success*. Princeton, NJ: Princeton University Press.

Bowles, S., H. Gintis, and M. Osborne (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature 39*(4), 1137–1176.

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

Breen, R. and J. O. Jonsson (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology 31*, 223–243.

Briley, D. A. and E. M. Tucker-Drob (2014a). Genetic and Environmental Continuity in Personality Development. a Meta-analysis. *Psychological Bulletin 104*(5), 1303–1331.

Briley, D. A. and E. M. Tucker-Drob (2014b). Genetic and Environmental Continuity in Personality Development: A Meta-Analysis. *Psychological Bulletin 140*(5), 1303–1331.

Bronte-Tinkew, J., K. A. Moore, and J. Carrano (2006). The Father-Child Relationship, Parenting Styles, and Adolescent Risk Behaviors in Intact Families. *Journal of Family Issues 27*(6), 850–881.

Brooks-Gunn, J. and L. B. Markman (2005). The Contribution of Parenting to Ethnic and Racial Gaps in School Readiness. *The Future of Children 15*(1), 139–168.

Brunello, G. and M. Schlotter (2011). Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems. *IZA Discussion Paper Series*.

Buchmann, C. and T. A. DiPrete (2006). The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement. *American Sociological Review 71*, 515–541.

Budig, M. J. and M. J. Hodges (2010). Differences in Disadvantage: Variation in the Motherhood

Penalty across White Women's Earnings Distribution. *American Sociological Review 75*(7), 705–728.

Bukodi, E. and J. H. Goldthorpe (2011). Class Origins, Education and Occupational Attainment in Britain. *European Societies 13*(3), 347–375.

Burgette, L. F. and J. P. Reiter (2010). Multiple Imputation for Missing Data via Sequential Regression Trees. *American Journal of Epidemiology 172*(9), 1070–1076.

Burkam, D., D. Ready, V. Lee, and L. LoGerfo (2004). Social-Class Differences in Summer Learning between Kindergarten and First grade: Model Specification and Estimation. *Sociology of Education 77*, 1–31.

Burkhauser, R. V., B. Nolan, and K. A. Couch (2009). Intragenerational Inequality and Intertemporal Mobility. In W. Salverda, B. Nolan, and T. M. Smeeding (Eds.), *The Oxford Handbook of Economic Inequality*, pp. 522–548. New York: Oxford University Press.

Cairns, R. B., B. D. Cairns, H. Xie, M. C. Leung, and S. Hearne (1998). Paths Across Generations: Academic Competence and Aggressive Behaviors in Young Mothers and Their Children. *Developmental Psychology 34*(6), 1162–1174.

Carbonaro, W. (2007). The Effects of Education and Cognitive Skill on Earnings: How Much Do Occupations and Jobs Matter? *Research in Social Stratification and Mobility 25*, 57–71.

Carneiro, P., C. Crawford, and A. Goodman (2006). Which Skills Matter. *CEE DP 59*.

Carr, K. A. (2020). A Cohort Longitudinal Study of Individual Level Social Capital and Depressive Symptoms in the Wisconsin Longitudinal Study. *SSM - Population Health 10*, 100544.

Carroll, G. and K. Mayer (1987). Jobs and Classes: Structural Constraints on Career Mobility. *European Sociological Review 3*, 14–38.

Catsambis, S. (2001). Expanding Knowledge of Parental Involvement in Children's Secondary Education: Connections with High School Seniors' Academic Success. *Social Psychology of Education 5*(2), 149–177.

Catsambis, S. (2016). Parental Involvement in Education. In *The Blackwell Encyclopedia of Sociology*. John Wiley & Sons, Ltd.

Cawley, J., J. Heckman, and E. Vytlacil (2001). Three Observations on Wages and Measured Cognitive Ability. *Labor Economics 8*(4), 419–442.

Cesarini, D. (2016). Rare Mutations and Educational Attainment. *Nature Neuroscience 19*(12), 1538–1539.

Cesarini, D., M. Johannesson, P. Lichtenstein, O. Sandewall, and B. Wallace (2010). Genetic Variation in Financial Decision-Making. *The Journal of Finance 65*(5), 1725–1754.

Cesarini, D., M. Johannesson, and S. Oskarsson (2014). Pre-Birth Factors, Post-Birth Factors, and Voting: Evidence from Swedish Adoption Data. *American Political Science Review 108*(1), 71–87.

Cesarini, D. and P. M. Visscher (2017). Genetics and Educational Attainment. *Science of Learning 2*(4).

Chang, C. C. et al. (2015). Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience 4*, 7.

Cheng, S. (2014). A Life Course Trajectory Framework for Understanding the Intracohort Pattern of Wage Inequality. *American Sociological Review 120*(3), 633–700.

Cheng, S. and X. Song (2019). Linked Lives, Linked Trajectories: Intergenerational Association of Intragenerational Income Mobility. *American Sociological Review 84*(6), 1037–1068.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics 129*(4), 1553–1623.

Chevalier, A., K. Denny, and D. McMahon (2003). A Multi-Country Study of Intergenerational Educational Mobility. *ISSC Discussion Paper, University College Dublin*.

Chin, T. and M. Phillips (2004). Social Reproduction and Child-Rearing Practices: Social Class, Children's Agency, and the Summer Activity Gap. *Sociology of Education 77*, 85–210.

Clark, A. E., S. Flèche, R. Layard, N. Powdthavee, and G. Ward (2018). *The Origins of Happiness: The Science of Well-Being over the Life Course*. Princeton University Press.

Coan, R. W. (1974). *The Optimal Personality: An Empirical and Theoretical Analysis*. New York, NY: Columbia University Press.

Coleman, J. S. (1966). Equality of Educational Opportunity [Summary Report]. Technical report, U.S. Dept. of Health, Education, and Welfare, Office of Education, Washington.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology 94*, S95–S120.

Conley, D. (2016). Socio-Genomic Research Using Genome-Wide Molecular Data. *Annual Review of Sociology 42*, 275–299.

Conley, D. and B. W. Domingue (2016). The Bell Curve Revisited: Testing Controversial Hypotheses with Molecular Genetic Data. *Sociological Science 3*, 520–539.

Conley, D., B. W. Domingue, D. Cesarini, C. Dawes, C. A. Rietveld, and J. D. Boardman (2015). Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype? *Sociological Science 2*, 82–105.

Conti, G. and J. J. Heckman (2010). Understanding the Early Origins of the Education-Health Gradient: A Framework that Can Also be Applied to Analyze Gene-Environment Interactions. *Perspectives on Psychological Science 5*(5), 585–605.

Cox, J. E. et al. (2019). A Parenting and Life Skills Intervention for Teen Mothers: A Randomized Controlled Trial. *Pediatrics 143*(3), e20182303.

Cronqvist, H. and S. Siegel (2015). The Origins of Savings Behavior. *Journal of Political Economy 123*(1), 123–169.

Cunha, F. and J. Heckman (2008). Formulating, Identifying, and Estimating the Technology of Cognitive and Non-Cognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2007). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

Darling, N. and L. Steinberg (1993). Parental Style as Context: An Integrative Model. *Psychological Bulletin 113*(3), 487–496.

Datar, A., M. R. Kilburn, and D. S. Loughran (2010). Endowments and Parental Investments in Infancy and Early Childhood. *Demography 47*(1), 145–162.

Davey Smith, G. (2005). What Can Mendelian Randomisation Tell Us About Modifiable Behavioural and Environmental Exposures. *BMJ 330*(7499), 1076–1079.

de Graaf, N. D., P. M. de Graaf, and G. Kraaykamp (2000). Parental Cultural Capital and Educational Attainment in the Netherlands: A Refinement of the Cultural Capital Perspective. *Sociology of Education 73*(2), 92–111.

de Zeeuw, E. L., C. E. van Beijsterveldt, T. J. Glasner, M. Bartels, E. A. Ehli, G. E. Davies, J. J. Hudziak, S. S. G. A. Consortium, C. A. Rietveld, M. M. Groen-Blokhuis, J. J. Hottenga, E. J. de Geus, and D. I. Boomsma (2014). Polygenic Scores Associated with Educational Attainment in Adults Predict Educational Achievement and ADHD Symptoms in Children. *American Journal of Medical Genetics 165*(6), 510–520.

Deary, I. J. et al. (2009). Age-Associated Cognitive Decline. *British Medical Bulletin 92*, 135–152.

deCastro, B. S., S. Catsambis, N. E. Hill, and R. K. Chao (2009). Parents Still Matter: Parental Links to High School Seniors' Behaviors and Future Outlook. In *Family-School Relations during Adolescence Linking Research, Policy, and Practice*. New York, NY: Teachers College Press.

Deke, J. and J. Haimson (2006). Valuing Student Competencies: Which Ones Predict Postsecondary Educational Attainment and Earnings, and for Whom? *Mathematica Policy Research Inc.*

Del Bono, E., J. Ermisch, and M. Francesconi (2012). Intrafamily Resource Allocations: A Dynamic Structural Model of Birth Weight. *Journal of Labor Economics 30*(3), 657–706.

Demange, P. A. et al. (2021). Investigating the Genetic Architecture of Noncognitive Skills Using GWAS-by-Subtraction. *Nature Genetics 53*(1), 35–44.

Dendir, S. (2014). Children's Cognitive Ability, Schooling and Work: Evidence from Ethiopia. *Journal of Educational Development 38*, 22–36.

Devlin, B., M. Daniels, and K. Roeder (1997). The Heritability of IQ. *Nature 388*(6641), 468–471.

DeYoung, C. G., J. B. Peterson, and D. M. Higgins (2005). Sources of Openness/Intellect: Cognitive and Neuropsychological Correlates of the Fifth Factor of Personality. *Journal of Personality 73*(4), 825–858.

Digman, J. M. and N. K. Takemoto-Chock (1981). Factors in the Natural Language of Personality: Re-Analysis, Comparison, and Interpretation of Six Major Studies. *Multivariate Behavioral Research 16*(2), 149–170.

DiPrete, T. A. (2020). The Impact of Inequality on Intergenerational Mobility. *Annual Review of Sociology 46*, 379–398.

DiPrete, T. A. and C. Buchmann (2006). Gender-Specific Trends in the Value of Education and the Emerging Gender Gap in College Completion. *Demography 43*(1), 1–24.

DiPrete, T. A. and C. Buchmann (2013). *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools*. New York, New York: Russell Sage Foundation.

Dishion, T. J., J. McCord, and F. Poulin (1999). When Interventions Harm: Peer Groups and Problem Behavior. *American Psychologist 54*(9), 755–764.

Dittman, C., K. Burke, and J. Hodges (2020). Brief Parenting Support for Parents of Teenagers Dealing with Family Conflict: A Pilot Randomized Controlled Trial. *Child & Youth Care Forum 49*, 799–816.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The Economics of Parenting. *Annual Review of Economics 11*, 55–84.

Doepke, M. and F. Zilibotti (2017). Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission. *Econometrica 85*(5), 1331–1371.

Domina, T., A. Conley, and G. Farkas (2011). The Link between Educational Expectations and Effort in the College-for-all Era. *Sociology of Education 84*(2), 93–112.

Domingue, B. W., D. W. Belsky, D. Conley, K. M. Harris, and J. D. Boardman (2015). Polygenic Influence on Educational Attainment: New Evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open 1*(3), 1–13.

Domingue, B. W., D. W. Belsky, A. Harrati, D. Conley, D. R. Weir, and J. D. Boardman (2017). Mortality Selection in a Genetic Sample and Implications for Association Studies. *International Journal of Epidemiology 46*(4), 1285–1294.

Domingue, B. W. and J. M. Fletcher (2020). Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes. *Behavior Genetics 50*(5), 301–309.

Dornbusch, S. M., P. L. Ritter, P. H. Leiderman, D. F. Robert, and M. J. Fraleigh (1987). The Relation of Parenting Style to Adolescent School Performance. *Child Development 58*(5), 1244–1257.

Downey, D., P. von Hippel, and B. Broh (2004). Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year. *American Sociological Review 69*, 613–635.

Dubow, E. F., P. Boxer, and L. R. Huesmann (2009). Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations. *Merrill Palmer Quarterly 55*(3), 224–249.

Duckworth, A. L. and M. E. P. Seligman (2005). Self-discipline Outdoes IQ in Predicting Academic Performance of Adolescents. *Psychological Science 16*(12), 939–944.

Dumais, S. A. (2002). Cultural Capital, Gender, and School Success: The Role of Habitus. *Sociology of Education 75*, 44–68.

Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, E. M. Feinstein, L., J. Brooks-Gunn, H. Sexton, K. Duckworth, and C. Japel (2007). School Readiness and Later Achievement. *Developmental Psychology 43*(6), 1428–1446.

Duncan, G. J., A. Kalil, S. Mayer, R. Tepper, and M. R. Payne (2005). The Apple Does Not Fall Far from the Tree. In S. Bowles, H. Gintis, and M. Groves Osborne (Eds.), *Unequal Chances: Family Background and Economic Succes*, pp. 23–79. Princeton, NJ: Princeton University Press.

Duncan, G. J., K. A. Magnusson, and J. Ludwig (2004). The Endogeneity Problem in Developmental Studies. *Research in Human Development 1*(1&2), 59–80.

Duncan, O. D. (1961). A Socioeconomic Index for All Occupations. In A. J. J. Reiss (Ed.), *Occupational and Social Status*, pp. 109–138. New York: Free Press.

Eccles, J. S. and A. Wigfield (1995). In the Mind of the Achiever: The Structure of Adolescents' Academic Achievement Related Beliefs and Self-Perceptions. *Personality and Social Psycholgy Bulletin 21*, 215–225.

Ellwood, D. and T. Kane (2000). Who is Getting a College Education? Family Background and the Growing Gaps in Enrollment. In S. Danziger and J. Waldfogel (Eds.), *Securing the Future: Investing in Children from Birth to College*. New York: Russell Sage.

Elster, J. (2012). *Nuts and Bolts for the Social Sciences*. Cambridge University Press.

Epstein, J. L. (2001). *School, Family and Community Partnerships: Preparing Educators and Improving Schools*. Boulder, CO: Westview Press.

Erikson, R. and J. H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon.

Erikson, R., J. H. Goldthorpe, and L. Portocarero (1979). Intergenerational Class Mobility in Three Western Societies: England, France, and Sweden. *British Journal of Sociology 30*(4), 415–441.

Erikson, R. and J. O. Jonsson (1996). Introduction. Explaining Class Inequality in Education: The Swedish Test Case. In R. Erikson and J. O. Jonsson (Eds.), *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, CO: Westview Press.

Ermisch, J. (2008). Origins of Social Mobility and Inequality: Parenting and Early Child Development. *National Institute Economic Review 205*, 62–71.

Euesden, J., C. M. Lewis, and P. F. O'Reilley (2015). PRSice: Polygenic Risk Score software. *Bioinformatics 31*(9), 1466–1468.

Evertsson, M. and M. Nermo (2004). Dependence within Families and the Division of Labour: Comparing Sweden and the United States? *Journal of Marriage and Family 66*(5), 1272–1286.

Farkas, G. (2003). Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. *Annual Review of Sociology 29*, 541–562.

Farkas, G., P. England, K. Vicknair, and B. S. Kilbourne (1997). Cognitive Skill, Skill Demands of Jobs, and Earnings among Young European American, African American, and Mexican American Workers. *Social Forces 75*(3), 913–938.

Featherman, D. L. (1971). A Research Note: A Social Structural Model for the Socioeconomic Career. *American Journal of Sociology 77*, 293–304.

Featherman, D. L. and R. M. Hauser (1978). *Opportunity and Change*. New York, NY: Academic Press.

Feinstein, L. (2000). The relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed in Childhood. *Centre for Economic Performance, LSE*.

Fergusson, D. M., L. J. Horwood, and E. M. Ridder (2005). Show Me the Child at Seven: The Consequences of Conduct Problems in Childhood for Psychosocial Functioning in Adulthood. *Journal of Child Psychology and Psychiatry 46*(8), 837–849.

Finn, A. S., T. Lee, A. Kraus, and C. L. Hudson Kam (2014). When It Hurts (and Helps) to Try: The Role of Effort in Language Learning. *PLOS One 9*(7), e101806.

Fish, N. J., B. S. Russell, R. J. Watson, and S. T. Russell (2020). Parent-Child Relationships and Sexual Minority Youth: Implications for Adult Alcohol Abuse. *Journal of Youth and Adolescence 49*, 2034–2046.

Flynn, K. E. and M. A. Smith (2007). Personality and Health Care Decision-Making Style. *Journal of Gerontology: Psychological Sciences 62*(B), 261–267.

Ford, M. (1992). *Motivating Humans: Goals, Emotions, and Personal Agency Beliefs*. Newbury Park, CA: Sage.

Freese, J. and S. Shostak (2009). Genetics and Social Inquiry. *Annual Review of Sociology 35*, 107–128.

Frijsters, P., D. W. Johnston, M. Shah, and M. A. Shields (2013). Intrahousehold Resource Allocation: Do Parents Reduce or Reinforce Child Ability Gaps? *Demography 50*, 2187–2208.

Fuller, S. (2008). Job Mobility and Wage Trajectories for Men and Women in the United States. *American Sociological Review 73*(1), 158–183.

Gamoran, A. (2001). American Schooling and Educational Inequality: A Forecast for the 21st Century. *Sociology of Education 74*, 135–153.

Ganzeboom, H. B. G., P. M. de Graaf, and D. J. Treiman (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research 21*(1), 1–56.

Ganzeboom, H. B. G. and D. J. Treiman (2003). Three Internationally Standardised Measures for Comparative Research on Occupational status. In J. H. P. Hoffmeyer-Zlotnik and C. Wolf (Eds.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, pp. 159–193. New York: Klewer Academic.

Ganzeboom, H. B. G., D. J. Treiman, and W. C. Ultee (1991). Comparative Intergenerational Stratification Research: Three Generations and Beyond. *Annual Review of Sociology 17*, 277–302.

Gensowski, M. (2018). Personality, IQ, and Lifetime Earnings. *Labour Economics 51*, 170–183.

Gintis, H. (1971). Education, Technology, and the Characteristics of Worker Productivity. *American Economic Review 61*(2), 266–279.

Glass, D. V. (Ed.) (1954). *Social Mobility in Britain*. London: Routledge.

Goldin, C. and L. F. Katz (2007). The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005. Working Paper 12984, NBER.

Goldthorpe, J. H. (2015). *Sociology as a Population Science*. Cambridge: Cambridge University Press.

Gonzalez, M. R. (2004). International Perspectives on Families, Schools and Communities: Educational Implications for Family-School-Community Partnerships. *International Journal of Educational Research 41*, 3–9.

Gratz, M. (2018). Competition in the Family: Inequality Between Siblings and the Intergenerational Transmission of Educational Advantage. *Sociological Science 5*, 246–269.

Gratz, M. and F. Torche (2016). Compensation or Reinforcement? The Stratification of Parental Responses to Children's Early Ability. *Demography 53*, 1883–1904.

Greven, C. U. et al. (2009). More Than Just IQ School Achievement Is Predicted by Self-Perceived Abilities—But for Genetic Rather Than Environmental Reasons. *Psychological Science 20*(6), 753–762.

Griliches, Z. (1979). Sibling Models and Data in Economics: Beginning of a Survey. *Journal of Political Economy 87*, S37–S64.

Grolnick, W. S. and M. L. Slowiaczek (1994). Parents' Involvement in Children's Schooling: A Multidimensional Conceptualization and Motivational Model. *Child Development 65*(1), 237–252.

Grotzinger, A. D. et al. (2019). Genomic Structural Equation Modelling Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *Nature Human Behavior 3*(5), 513–525.

Groves, M. O. (2005). How Important Is Your Personality? Labor Market Returns to Personality for Women in the US and UK. *Journal of Economic Psychology 26*(6), 827–841.

Grusky, D. B. and J. Hill (2018). *Inequality in the 21st Century: A Reader*. New York: Routledge.

Grusky, D. B. and J. B. Sorensen (1998). Can Class Analysis be Salvaged? *American Journal of Sociology 103*(5), 1187–1234.

Grusky, D. B. and K. A. Weeden (2006). Does the Sociological Approach to Studying Social Mobility Have a Future? In S. L. Morgan, D. B. Grusky, and G. S. Fields (Eds.), *Mobility and Inequality*. Stanford, CA: Stanford University Press.

Grönqvist, E., B. Öckert, and J. Vlachos (2010). The Intergenerational Transmission of Cognitive and Non-Cognitive Abilities. *IFAU Working Paper Series*.

Guo, G. et al. (2015). Peer Influence, Genetic Propensity, and Binge Drinking: A Natural Experiment and a Replication. *American Journal of Sociology 121*(3), 914–954.

Guo, G. and E. Stearns (2002). The Social Influences on the Realization of Genetic Potential for Intellectual Development. *Social Forces 80*(3), 881–910.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental Education and Parental Time with Children. *Journal of Economic Perspectives 22*(3), 23–46.

Gutman, L. M. and J. S. Eccles (2007). Stage-environment fit during adolescence: Trajectories of family relations and adolescent outcomes. *Developmental Psychology 43*(2), 522–537.

Gutman, L. M. and I. Schoon (2013). The Impact of Non-Cognitive Skills on Outcomes for Young People. *Institute of Education Literature Review*.

Haller, A. O. (1982). Reflections on the Social Psychology of Status Attainment. In *Social Structure and Behavior: Essays in Honor of William Hamilton Sewell*, pp. 3–28. New York: Academic Press.

Haller, A. O. and A. Portes (1973). Status Attainment Processes. *Sociology of Education 46*(1), 51–91.

Hampden-Thompson, G., L. Guzman, and L. Lippman (2013). A Cross-National Analysis of Parental Involvement and Student literacy. *International Journal of Comparative Sociology 54*(3), 246–266.

Hango, D. (2007). Parental Investment in Childhood and Educational Qualifications: Can Greater Parental Involvement Mediate the Effects of Socioeconomic Disadvantage? *Social Science Research 36*(4), 1371–1390.

Harris, J. R. (1998). *The Nurture Assumption: Why Children Turn Out the Way They Do*. New York, NY: Free Press.

Harris, K. M., C. T. Halpern, E. A. Whitsel, Hussey, L. A. J. M. Killeya-Jones, J. Tabor, and S. C. Dean (2019). Cohort profile: The national longitudinal study of adolescent to adult health (add health). *International Journal of Epidemiology 48*(5), 1415–1415k.

Harris, M. A., C. E. Brett, W. Johnson, and I. J. Deary (2016). Personality Stability From Age 14 to Age 77 Years. *Psychology and Aging 31*(8), 862–874.

Hart, B. and T. R. Risley (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing.

Haug, M. R. (1977). Measurement in Social Stratification. *Annual Review of Sociology 3*, 51–77.

Hauser, R. M., D. H.-H. Kuo, and R. Cartmill (2003). Birth Order and Personality Among Adult Siblings: Are There Any Effects? *CDE Working Paper No. 2003-16*.

Hauser, R. M. and A. Palloni (2011). Adolescent IQ and Survival in the Wisconsin Longitudinal Study. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences 66B*(S1), i91–i101.

Hauser, R. M., J. T. Sheridan, and J. B. Warren (1999). Socioeconomic Achievements of Siblings in the Life Course: New Findings from the Wisconsin Longitudinal Study. *Research on Aging 21*, 338–378.

Hauser, R. M. and J. B. Warren (1997). Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology 27*(1), 177–298.

Haveman, R. and T. M. Smeeding (2006). The Role of Higher Education in Social Mobility. *The Future of Children 16*(2), 125–150.

Haveman, R. and B. Wolfe (1995). The Determinants of Children's Attainment: A Review of Methods and Findings. *Journal of Economic Literature 33*(4), 1829–1878.

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*, 734–736.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science 312*(5782), 1900–1902.

Heckman, J. J. (2007). The Economics, Technology, and Neuroscience of Human Capability Formation. *PNAS 104*(33), 13250–13255.

Heckman, J. J. (2013). *Giving Kids a Fair Chance*. MIT Press.

Heckman, J. J. and T. Kautz (2012a). Hard Evidence on Soft Skills. *Labour Economics 19*(4), 451–464.

Heckman, J. J. and T. Kautz (2012b). Hard Evidence on Soft Skills. *Labour Economics 19*(4), 451–464.

Heckman, J. J. and S. Mosso (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics 6*, 689–733.

Heckman, J. J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. and Y. Rubinstein (2001). The Benefits of Skill: The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review 91*(2), 145–149.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *NBER Working Paper Series*.

Hedstrom, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Hedstrom, P. and P. S. Bearman (2009). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.

Henmon, V. A. C. and M. J. Nelson (1946). *Henmon-Nelson Tests of mental ability, high school examination - Grades 7 to 12-forms A, B, C. Teacher's Manual.* Boston, MA: Houghton-Miffin Company.

Henry, C. S., S. W. Plunkett, and T. Sands (2011). Family Structure, Parental Involvement, and Academic Motivation in Latino Adolescent. *Journal of Divorce and Remarriage 52*(6), 370–390.

Herd, P. et al. (2019). Genes, Gender Inequality, and Educational Attainment. *American Sociological Review 84*(6), 1069–1098.

Herd, P., D. Carr, and C. Roan (2014). Cohort Profile: Wisconsin Longitudinal Study (WLS). *International Journal of Epidemiology 43*(1), 34–41.

Herd, P., J. Freese, K. Sicinski, B. W. Domingue, K. M. Harris, C. Wei, and R. M. Hauser (2019). Genes, Gender Inequality, and Educational Attainment. *American Sociological Review 84*(6), 1069–1098.

Hewitt, J. K. (2012). Editorial Policy on Candidate Gene Association and Candidate Gene-by-Environment Interaction Studies of Complex Traits. *Behavior Genetics 42*, 1–2.

Hill, C. R. and F. Stafford (1974). Allocation of Time to Preschool Children and Educational Opportunity. *Journal of Human Resources 9*, 323–341.

Hill, N. E. and L. C. Taylor (2004). Parental School Involvement and Children's Academic Achievement: Pragmatics and Issues. *Current Directions in Psychological Science 13*(4), 161–164.

Hill, N. E. and D. F. Tyson (2009). Parental Involvement in Middle School: A Meta-Analytic Assessment of the Strategies that Promote Achievement. *Developmental Psychology 45*(3), 740–763.

Hill, W. D. et al. (2019). Genome-Wide Analysis Identifies Molecular Systems and 149 Genetic Loci Associated with Income. *Nature Communications 10*.

Hjalmarsson, R. and M. Lindqvist (2013). The Origins of Intergenerational Associations in Crime: Lessons from Swedish Adoption Data. *Labour Economics 20*(C), 68–81.

Hogan, R. T. (1986). *Manual for the Hogan Personality Inventory.* Minneapolis: National Computer Systems.

Hoover-Dempsey, K. V. and H. M. Sandler (1997). Why Do Parents Become Involved in Their Children's Education? *Review of Educational Research 67*(1), 3–42.

Hornby, G. and R. Lafaele (2011). Barriers to Parental Involvement in Education: An Explanatory Model. *Educational Review 63*(1), 37–52.

Hout, M. (1988). More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s. *American Journal of Sociology 93*(6), 1358–1400.

Hout, M. (2015). A Summary of What We Know About Social Mobility. *The Annals of the American Academy 657*, 27–36.

Hsin, A. (2012). Is Biology Destiny? Birth Weight and Differential Parental Treatment. *Demography 49*, 1385–1405.

Härkönen, J., A. Manzoni, and E. Bihagen (2016). Gender Inequalities in Occupational Prestige Across the Working Life: An Analysis of the Careers of West Germans and Swedes Born from the 1920s to the 1970s. *Advances in Life Course Research 29*, 41–51.

Inglehart, R. (1977). *The Silent Revolution: Changing Values and Political Styles Among Western Publics.* Princeton University Press.

Jackson, M., J. H. Goldthorpe, and C. Mills (2005). Education, Employers and Class Mobility. *Research in Social Stratification and Mobility 23*, 3–33.

Jang, K. L., W. J. Livesley, and P. A. Vemon (1996). Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study. *Journal of Personality 64*(3), 577–592.

Jencks, C. (1972a). *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York, NY: Basic Books.

Jencks, C. (1972b). *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York, New York: Basic Books.

Jencks, C. (1979). *Who Gets Ahead? The Determinants of Economic Success in America.* New York, NY: Basic Books.

Jencks, C. (1980). Heredity, Environment, and Public Policy Reconsidered. *American Sociological Review 45*(5), 723–736.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability.* Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student. *Urban Education 40*(3), 237–269.

Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education 42*(1), 82–110.

John, O. P., E. M. Donahue, and R. L. Kentle (1991). *The "Big Five" inventory – version 4a and 5a.* Berkeley, CA: Institute of Personality and Social Research, University of California, Berkeley.

Jones, I. and S. White (2000). Family Composition, Parental Involvement, and Young Children's Academic Achievement. *Early Child Development and Care 161*(1), 71–82.

Jonsson, J. O. (1987). Class Origin, Cultural Origin, and Educational Attainment: The Case of Sweden. *European Sociological Review 3*(3), 229–242.

Kalil, A. (2014). Addressing the Parenting Divide and Children's Life Chances ". The Hamilton Project Discussion Paper (Chapter 2 in The Hamilton Project's "Policies to Address Poverty in America.". *Policies to Address Poverty in America*.

Kalil, A., R. Ryan, and M. Corey (2012). Diverging Destinies: Maternal Education and Investments in Children. *Demography 49*, 1361–1383.

Kalil, A., K. M. Ziol-Guest, R. Ryan, and A. J. Markowitz (2016). Changes in Income-Based Gaps in Parent Activities with Young Children From 1988 to 2012. *AERA Open 2*(3), 1–17.

Karlsson Linner, R. et al. (2019). Genome-Wide Association Analyses of Risk Tolerance and Risky Behaviors in over 1 Million Individuals Identify Hundreds of Loci and Shared Genetic Influences. *Nature Genetics 51*(2), 245–257.

Kelley, J. (1973). Causal Chain Models for the Socioeconomic Career. *American Sociological Review 38*, 481–493.

Kerckhoff, A. C. (1976). The Status Attainment Process: Socialization or Allocation? *Social Forces 55*(2), 368–381.

Knudsen, E. I. et al. (2006). Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce. *PNAS 103*(27), 10155–10162.

Kong, A. et al. (2018a). The Nature of Nurture: Effects of Parental Genotypes. *Science 359*(6374), 424–428.

Kong, A. et al. (2018b). The Nature of Nurture: Effects of Parental Genotypes. *Science 359*(6374), 424–428.

Krapohl, E. and R. Plomin (2016). Genetic Link between Family Socioeconomic Status and Children's Educational Achievement Estimated from Genome-Wide SNPs. *Molecular Psychiatry 21*, 437–443.

Krapohl, E., K. Rimfeld, N. G. Shakeshaft, M. Trzaskowski, A. McMillan, J. B. Pingault, K. Asbury, N. Harlaar, Y. Kovas, P. S. Dale, and R. Plomin (2014). The High Heritability of Educational Achievement Reflects Many Genetically Influenced Traits, Not Just Intelligence. *PNAS 111*(42), 15273–15278.

Laidra, K., H. Pullmann, and J. Allik (1981). Cultivating Competence, Self-efficacy, and Intrinsic Interest through Proximal Self-Motivation. *Personality and Individual Differences 41*(3), 586–559.

Laidra, K., H. Pullmann, and J. Allik (2007). Personality and Intelligence as Predictors of Academic Achievement: A Cross-sectional Study from Elementary to Secondary School. *Personality and Individual Differences 42*, 441–451.

Lareau, A. (1989). *Home Advantage: Social Class and Parental Intervention in Elementary Education*. London: Falmer.

Lareau, A. (2002). Invisible Inequality: Social Class and Childrearing in Black Families and White Families. *American Sociological Review 67*(5), 747–776.

Lareau, A. (2003). *Unequal Childhoods: Class, Race, and Family Life*. Los Angeles, CA: University of California Press.

Leary, M. and R. Baumeister (2000). The Nature and Function of Self-Esteem: Sociometer Theory. In *Advances in Experimental Social Psychology*, Volume 32, pp. 1–62. San Diego, CA: Academic Press.

Lee, J. 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*, 1112–1121.

Legewie, J. and T. A. DiPrete (2014). The High School Environment and the Gender Gap in Science and Engineering. *Sociology of Education 87*(4), 259–280.

Leight, J. and E. M. Liu (2018). Maternal Education, Parental Investment and Non-Cognitive Characteristics in Rural China. *IZA Discussion Paper Series*.

Lichtenstein, P., S. L. Hershberger, and N. L. Pedersen (1995). Dimensions of Occupations: Genetic and Environmental Influences. *Journal of Biosocial Science 27*(2), 193–206.

Lindquist, M. J., J. Sol, and M. Van Praag (2015). Why Do Entrepreneurial Parents Have Entrepreneurial Children? *Journal of Labor Economics 33*(2), 269–296.

Lipset, S. M. and R. Bendix (1959). *Social Mobility in Industrial Societies*. Berkeley: University of California Press.

Liu, H. (2018). Social and Genetic Pathways in Multigenerational Transmission of Educational Attainment. *American Sociological Review 83*(2), 278–304.

Liu, M. et al. (2019). Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use. *Nature Genetics 51*(2), 237–244.

Loehlin, J. C., R. R. McCrae, and P. T. Costa, Jr. (1998). Heritabilities of Common and Measure-Specific Components of the Big Five Personality Factors. *Journal of Research in Personality 32*, 431–453.

Lynch, J. L. and R. Brooks (2013). Low Birth Weight and Parental Investment: Do Parents Favor the Fittest Child? *Journal of Marriage and Family 75*, 533–543.

Maccoby, E. E. and C. N. Jacklin (1978). *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.

Maccoby, E. E. and J. Martin (1983). Socialization in the Context of the Family: Parent-Child Interaction. In *Handbook of Child Psychology*, pp. 1–101. New York: Wiley.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty Impedes Cognitive Function. *Science 30*, 976–980.

Manzoni, A., J. Härkönen, and K. U. Mayer (2014). Moving On? A Growth-Curve Analysis of Occupational Attainment and Career Progression Patterns in West Germany. *Social Forces 92*(4), 1285–1312.

Mare, R. D. (1993). Educational Stratification on Observed and Unobserved Components of Family Background. In Y. Shavit and H. P. Blossfeld (Eds.), *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview Press.

Marsh, H. W. and R. Craven (1997). Academic Self-Concept: Beyond the Dustbowl. In *Educational Psychology Series. Handbook of Classroom Assessment: Learning, Achievement, and Adjustment*, pp. 131–198. San Diego, CA: Academic Press.

Mathieson, I. et al. (2020). Genome-Wide Analysis Identifies Genetic Effects on Reproductive Success and Ongoing Natural Selection at the FADS Locus. *bioRxiv*.

Mayer, S. E., A. Kalil, and S. Gallegos (2015). Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together (PACT) Intervention.

McCarthy, M. I., G. R. Abecasis, L. R. Cordon, D. B. Goldstein, J. Little, J. P. Ioannidis, and J. N. Hirschhorn (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics 9*(5), 356–369.

McCrae, R. R. and P. T. Costa, Jr. (1990). *Personality in Adulthood*. New York, NY: Guilford Press.

McCrae, R. R. and O. P. John (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality 60*(2), 175–2015.

McGue, M. et al. (2007). The Environments of Adopted and Non-Adopted Youth: Evidence on Range Restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics 37*, 449–462.

McLanahan, S. (2004). Diverging Destinies: How Children Are Faring Under the Second Demographic Transition. *Demography 41*, 607–627.

McNeal Jr., R. B. (2001). Differential Effects of Parental Involvement on Cognitive and Behavioral Outcomes by Socioeconomic Status. *Journal of Socio-Economics 30*, 171–179.

Meijers, H., L. Borghans, and B. ter Weel (2006). The Role of Noncognitive Skills in Explaining Cognitive Test Scores. *MERIT Working Papers 44*.

Meyer, J., J. Fleckenstein, J. Retelsdorg, and O. Koller (2019). The Relationship of Personality Traits and Different Measures of Domain-Specific Achievement in Upper Secondary Education. *Learning and Individual Differences 69*, 45–59.

Mills, M. C. et al. (2021). Identification of 370 Genetic Loci for Age at First Sex and Birth Linked to Externalising Behaviour. *bioRxiv*.

Mills, M. C., N. Barban, and F. C. Tropf (2018). The Sociogenomics of Polygenic Scores of Reproductive Behavior and their Relationship to other Fertility Traits. *RSF: The Russell Sage Foundation Journal of the Social Sciences 4*(4), 122–136.

Mills, M. C., N. Barban, and F. C. Tropf (2020a). *An Introduction to Statistical Genetic Data Analysis*. Boston: The MIT Press.

Mills, M. C., N. Barban, and F. C. Tropf (2020b). *An Introduction to Statistical Genetic Data Analysis*. Boston, Massachusetts: MIT Press.

Mills, M. C., A. D. Johnston, and T. A. DiPrete (2006). Globalization and Men's Job Mobility in the United States. In H. P. Blossfeld, M. C. Mills, and F. Bernardi (Eds.), *Globalization,*

Uncertainty, and Men's Careers: An International Comparison, pp. 335–372. Chelthenham, UK: Edward Elgar.

Mills, M. C. and C. Rahal (2019a). A Scientometric Review of Genome-Wide Association Studies. *Communications Biology 2*.

Mills, M. C. and C. Rahal (2019b). A Scientometric Review of Genome-Wide Association Studies. *Nature Communications Biology 2*.

Mills, M. C. and F. C. Tropf (2020). Sociology, Genetics, and the Coming of Age of Sociogenomics. *Annual Review of Sociology 46*, 31.1–31.29.

Mischel, W., Y. Shoda, and M. I. Rodriguez (1989). Delay of Gratification in Children. *Science 244*(4907), 933–938.

Moffitt, T. E. et al. (2011). A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *PNAS 108*(7), 2693–2698.

Morgan, S. L. and M. W. McKerrow (2007). Human Capital Development and the Frontier of Research in the Sociology of Education. In G. Picot, R. Saunders, and A. Sweetman (Eds.), *Fulfilling Potential, Creating Success: Perspectives on Human Capital Development*, pp. 185–225. Montreal: McGill-Queen's University Press.

Mueller, G. and E. Plug (2006). Estimating the Effect of Personality on Male and Female Earnings. *Industrial and Labor Relations Review 60*(1), 3–22.

Murnane, R. J. and F. Levy (2006). Why the Changing American Economy Calls for Twenty-First Century Learning: Answers to Educators' Questions. *New Directions for Youth Development* (110), 53–62.

Mustillo, S. (2012). The Effects of Auxiliary Variables on Coefficient Bias and Efficiency in Multiple Imputation. *Sociological Methods & Research 41*(2), 335–361.

Nakao, K. and J. Treas (1991). Bringing Occupational Prestige Scores Up to Date: Construction and Evaluation of the New Scale.

Neisser, U. et al. (1996). Intelligence: Knowns and Unknowns. *American Psychologist 51*(2), 77–101.

Neuenschwander, M. P., M. Vida, J. L. Garrett, and J. S. Eccles (2007). Parents' Expectations and Students' Achievement in Two Western Nations. *International Journal of Behavioral Development 31*(6), 594–602.

Okbay, A. et al. (2016). Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment. *Nature 533*, 539–542.

Parsons, J. E., T. F. Adler, and C. M. Kaczala (1982). Socialization of Achievement Attitudes and Beliefs: Parental Influences. *Child Development 53*, 310–321.

Paulhus, D. L. et al. (1998). Egoistic and Moralistic Biases in Self-perception: The Interplay of Self-deceptive Styles with Basic Traits and Motives. *Journal of Personality 66*(6), 1025–1060.

Pearce, A., A. C. P. Sawyer, C. R. Chittleborough, M. N. Mittinty, C. Law, and J. W. Lynch (2016). Do Early Life Cognitive Ability and Self-Regulation Skills Explain Socio-Economic Inequalities in Academic Achievement? An Effect Decomposition Analysis in UK and Australian cohorts. *Social Science & Medicine 165*, 108–118.

Penke, L., J. J. A. Denissen, and G. F. Miller (2007). The Evolutionary Genetics of Personality. *European Journal of Personality 21*, 549–587.

Perrin, E. (1904). On the Contingency Between Occupation in the Case of Fathers and Sons. *Biometrika 3*, 467–469.

Pfeffer, F. T. (2008). Persistent Inequality in Educational Attainment and Its Institutional Context. *European Sociological Review 24*(5), 543–565.

Phillips, M. (2011). Parenting, Time Use, and Disparities in Academic Outcomes. In G. J. Duncan and R. J. Murnane (Eds.), *Whither Ppportunity: Rising Inequality, Schools, and Children's Life Chances*, pp. 207–228. New York, NY: Russell Sage Foundation and Spencer Foundation.

Pitt, M. M., M. R. Rosenzweig, and N. Hassan (1990). Productivity, Health and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries. *American Economic Review 80*(5), 1139–1156.

Plenty, M. S. and J. O. Jonsson (2021). Students' Occupational Aspirations: Can Family Relationships Account for Differences Between Immigrant and Socioeconomic Groups? *Child Development 92*(1), 157–173.

Plomin, R., M. J. Owen, and P. McGuffin (1994). The genetic basis of complex human behaviors. *Science 264*(5166), 1733–1739.

Polderman, T. J. C., B. Benyamin, C. A. de Leeuw, P. F. Sullivan, A. van Bochoven, P. M. Visscher, and D. Posthuma (2015). Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies. *Nature Genetics 47*, 702–709.

Pomerantz, E. M., E. A. Moorman, and S. D. Litwack (2007). The How, Whom and Why of Parents' Involvement in Children's Academic Lives: More is Not Always Better. *Review of Educational Research 77*(3), 373–410.

Poole, L. A. et al. (2018). A Randomized Controlled Trial of the Impact of a Family-Based Adolescent Depression Intervention on both Youth and Parent Mental Health Outcomes. *Journal of Abnormal Child Psychology 46*, 169–181.

Popejoy, A. B. and S. M. Fullerton (2016). Genomics is Failing on Diversity. *Nature 538*, 161–164.

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin 135*(2), 322–338.

Poropat, A. E. (2014). Other-rated Personality and Academic Performance: Evidence and Implications. *Learning and Individual Differences 43*, 24–32.

Portes, A. (1998). Social Capital: Its Origins and Applications in Modern Sociology. *Annual Review of Sociology 24*, 1–24.

Price, J. (2008). Parent-Child Quality Time Does Birth Order Matter? *Journal of Human Resources 43*(1), 240–265.

Price, J. (2010). The Effect of Parental Time Investments: Evidence from Natural Within-Family Variation. *Working Paper*.

Privé, F. et al. (2021). High-Resolution Portability of 240 Polygenic Scores When Derived and Applied in the Same Cohort. *medRxiv*.

Quadlin, N. Y. (2016). Gender and Time Use in College: Converging or Diverging Pathways? *Gender & Society 30*(2), 361–385.

Ramey, G. and V. A. Ramey (2009). The Rug Rat Race. *NBER Working Paper Series*.

Ream, R. K. and G. J. Palardy (2008). Reexamining Social Class Differences in the Availability and the Educational Utility of Parental Social Capital. *American Educational Research Journal 45*(2), 238–273.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. In G. J. Duncan and R. J. Murnane (Eds.), *Rising Inequality, Schools, and Children's Life Chances*, pp. 91–116. New York, NY: Russell Sage Foundation.

Reeves, R. V. and K. Howard (2013). The Parenting Gap. *Center on Children & Families at Brookings*.

Reiss, A. J. J. (1961). *Occupations and Social Status*. New York: Free Press of Glencoe.

Restrepo, B. J. (2016). Parental Investment Responses to a Low Birth Weight Outcome: Who Compensates and Who Reinforces? *Journal of Population Economics 29*, 969–989.

Richardson, M., C. Abraham, and R. Bond (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin 138*(2), 353–387.

Riemann, R., A. Angleitner, and J. Strelau (1997). Genetic and Environmental Influences on Personality: A Study of Twins Reared Together Using the Self- and Peer Report NEO-FFI Scales. *Journal of Personality 65*(3).

Rietveld, C. A. et al. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science 340*(6139), 1467–1471.

Rietveld, C. A. et al. (2014). Replicability and Robustness of Genome-Wide Association Studies for Behavioral Traits. *Psychological Science 25*(11), 1975–1986.

Rimfeld, C. A. et al. (2018). Genetic Influence on Social Outcomes During and After the Soviet Era in Estonia. *Nature Human Behavior 2*(4), 269–275.

Rimfeld, K., Z. Ayorech, P. S. Dale, Y. Kovas, and R. Plomin (2016). Genetics Affects Choice of Academic Subjects as well as Achievement. *Nature Scientific Reports 6*.

Rimfeld, K., Y. Kovas, P. S. Dale, and R. Plomin (2016). True Grit and Genetics: Predicting Academic Achievement from Personality. *Journal of Personality and Social Psychology 111*(5), 780–789.

Roberts, B. W. and W. F. DelVecchio (2000). The Rank-Order Consistency of Personality Traits from Childhood to Old Age: A Quantitative Review of Longitudinal Studies. *Psychological Bulletin 126*(1), 3–25.

Robison, M. R. et al. (2017). Genetic Evidence of Assortative Mating in Humans. *Nature Human Behavior 1*(0016).

Robling, M. et al. (2016). Effectiveness of a Nurse-Led Intensive Home-Visitation Programme for First-time Teenage Mothers (Building Blocks): A Pragmatic Randomised Controlled Trial. *The Lancet 387*(10014), 146–155.

Rockeach, M. (1960). *The Open and Closed Mind.* Oxoford, England: Basic Books.

Rogers, C. R. (1961). *On Becoming a Person.* Boston: Houghton Mifflin.

Rosales-Rueda, M. F. (2014). Family Investment Responses to Childhood Health Conditions: Intrafamily Allocation of Resources. *Journal of Health Economics 37*, 41–57.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image.* Princeton, NJ: Princeton University Press.

Rosin, H. (2012). *The End of Men: And the Rise of Women*. New York, NY: Penguin.

Rowe, D. C., K. C. Jacobson, and E. J. C. G. Van den Oord (1999). Genetic and Environmental Influences on Vocabulary IQ: Parental Education Level as Moderator. *Child Development 70*(5), 1151–1162.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc.

Ryan, R. M. and J. Stiller (1991). The Social Contexts of Internalization: Parent and Teacher Influences on Autonomy, Motivation and Learning. In M. L. Maehr and P. R. Pintrich (Eds.), *Advances in Motivation and Achievement*, pp. 115–149. Greenwich, CT: JAI Press.

Ryff, C. D. (1989). Happiness is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being. *Journal of Personality and Social Psychology 57*(6), 1069–1081.

Sacchi, S., I. Kriesi, and M. Buchmann (2016). Occupational Mobility Chains and the Role of Job Opportunities for Upward, Lateral and Downward Mobility in Switzerland. *Research in Social Stratification and Mobility 44*, 10–21.

Sanz-de Galdeano, A. and A. Terskaya (2019). Sibling Differences in Educational Polygenic Scores: How Do Parents Reacts? *IZA Discussion Paper Series*.

Sayer, L., A. Gauthier, and F. Furstenberg (2004). Educational Differences in Parents' Time with Children: Cross-National Variations. *Journal of Marriage and Family 66*, 1152–1169.

Scarr, S. and K. McCartney (1983). How People Make Their Own Environments: A Theory of Genotype →Environment Effects. *Child Development 54*(2), 424–435.

Scarr-Salapatek, S. (1971). Race, social class, and iq. *Science 174*(4016), 1285–1289.

Schaub, M. (2010). Parenting for Cognitive Development from 1950 to 2000: The Institutionalization of Mass Education and the Social Construction of Parenting in the United States. *Sociology of Education 83*(1), 46–66.

Segal, C. (2008). Motivation, Test Scores, and Economic Success. *Universitat Pompeu Fabra Working Papers*.

Selzam, S. et al. (2019). Comparing Within- and Between-Family Polygenic Score Prediction. *The American Journal of Human Genetics 105*, 351–363.

Selzam, S., P. S. Dale, R. K. Wagner, J. C. DeFries, M. Cederlof, P. F. O'Reilly, E. Krapohl, and R. Plomin (2017). Genome-Wide Polygenic Scores Predict Reading Performance Throughout the School Years. *Scientific Studies of Reading 21*(4), 334–349.

Selzam, S., E. Krapohl, S. von Stumm, P. F. O'Reilly, K. Rimfield, Y. Kovas, P. S. Dale, J. J. Lee, and R. Plomin (2017). Predicting Educational Achievement from DNA. *Molecular Psychiatry 22*, 267–272.

Sewell, W. A., A. O. Haller, and A. Portes (1969). The Educational and Early Occupational Attainment Process. *American Sociological Review 34*(1), 82–92.

Sewell, W. H. and R. M. Hauser (1972). Causes and Consequences of Higher Education: Models of the Status Attainment Process. *American Journal of Agricultural Economics 54*(5), 851–861.

Sewell, W. H., R. M. Hauser, and W. C. Wolf (1980). Sex, Schooling, and Occupational Status. *American Journal of Sociology 86*, 551–583.

Shah, A., S. Mullainathan, and E. Shafir (2012). Some Consequences of Having Too Little. *Science 339*, 682–685.

Shoda, Y., W. Mischel, and P. K. Peake (1990). Predicting Adolescent Cognitive and Self-Regulatory Competencies from Preschool Delay of Gratification: Identifying Diagnostic Conditions. *Developmental Psychology 26*(6), 978–986.

Siegel, P. S. (1971). *Prestige in the American Occupational Structure*. Unpublished Ph.D. dissertation, University of Chicago, Chicago.

Slatkin, M. (2008). Linkage Disequilibrium - Understanding the Evolutionary Past and Mapping the Medical Future. *Nature Reviews Genetics 9*, 477–485.

Smith-Whooley, E., S. Selzam, and R. Plomin (2019). Polygenic Score for Educational Attainment Captures DNA Variants Shared between Personality Traits and Educational Achievement. *Journal of Personality and Social Psychology 117*(6), 1145–1163.

Smithers, L. G. et al. (2018). A Systematic Review and Meta-Analysis of Effects of Early Life Non-Cognitive Skills on Academic, Psychosocial, Cognitive and Health Outcomes. *Nature Human Behavior 2*, 867–880.

Sorokin, P. (1927). *Social Mobility.* New York: Harper and Brothers.

Spera, C. (2005). A Review of the Relationship Among Parenting Practices, Parenting Styles, and Adolescent School Achievement. *Educational Psychology Review 17*(2), 125–146.

Spilerman, S. (1977). Careers, Labor Market Structure, and Socioeconomic Achievement. *American Journal of Sociology 83*(3), 551–593.

Sullivan, T. R., A. B. Salter, P. Ryan, and K. J. Lee (2015). Bias and Precision of the "Multiple Imputation, Then Deletion" Method for Dealing With Missing Outcome Data. *American Journal of Epidemiology 182*(6), 528–534.

Svalastoga, K. (1959). *Prestige, Class, and Mobility.* Copenhagen: Gyldendal.

Teachman, J. D. (1987). Family Background, Educational Resources, and Educational Attainment. *American Sociological Review 52*(4), 548–557.

Teachman, J. D. (1996). Intellectual Skill and Academic Performance: Do Families Bias the Relationship? *Sociology of Education 69*(1), 35–48.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective.* New York: Academic Press.

Tropf, F. et al. (2017). Hidden Heritability Due to Heterogeneity Across Seven Populations. *Nature Human Behavior 1*, 757–765.

Tucker-Drob, E., D. A. Briley, L. E. Engelhardt, F. D. Mann, and K. Paige Harden (2016). Genetically-Mediated Associations Between Measures of Childhood Character and Academic Achievement. *Journal of Personality and Social Psychology 111*(5), 790–815.

Tucker-Drob, E. M., D. A. Briley, L. E. Engelhardt, F. D. Mann, and K. P. Harden (2016). Genetically-Mediated Associations between Measures of Childhood Character and Academic Achievement. *Journal of Personality and Social Psychology 111*(5), 790–815.

Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science 9*(5), 160–164.

Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio, and I. I. Gottesman (2003). Socioeconomic Status Modifies Heritability of IQ in Young Children. *Psychological Science 14*(6), 623–628.

Tyre, P. (2008). *The Trouble with Boys: A Surprising Report Card on Our Sons, Their Problems at School, and What Parents and Educators Must Do*. New York, NY: Three Rivers Press.

van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Boca Raton, Florida: CRC Press.

van Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software 45*(3), 1–67.

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annual Review of Public Health 37*, 17–32.

Vilhjamsson, B. J. et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics 97*(4), 576–592.

Visscher, P. M., W. G. Hill, and N. R. Wray (2008). Heritability in the Genomics Era - Concepts and Misconceptions. *Nature Reviews Genetics 9*, 255–266.

von Culin, K. R., E. Tsukayama, and A. L. Duckworth (2014). Unpacking Grit: Motivational Correlates of Perseverance and Passion for Long-Term Goals. *The Journal of Positive Psychology 9*(4), 306–312.

von Hippel, P. T. (2007). Regression with missing y's: An improved strategy for analyzing multiply imputed data. *Sociological Methodology 37*, 83–117.

von Otter, C. (2014). Family Resources and Mid-Life Level of Education: A Longitudinal Study of the Mediating Influence of Childhood Parental Involvement. *British Educational Research Journal 40*(3), 555–574.

von Stumm, S. and P. L. Ackerman (2013). Investment and Intellect: A Review and Meta-Analysis. *Psychological Bulletin 139*(4), 841–869.

von Stumm, S., B. Hell, and T. Chamorro-Premuzic (2011). The Hungry Mind: Intellectual Curiosity is the Third Pillar of Academic Performance. *Perspectives on Psychological Science 6*(6), 574–588.

Waldfogel, J. and E. Washbrook (2011). Early Years Policy. *Child Development Research*, 1–12.

Wang, M.-T. and J. S. Eccles (2012). Social Support Matters: Longitudinal Effects of Social Support on Three Dimensions of School Engagement From Middle to High School. *Child Development 83*(3), 877–895.

Wang, M.-T. and S. Sheikh-Khalil (2014). Does Parental Involvement Matter for Student Achievement and Mental Health in High School. *Child Development 85*(2), 610–625.

Ware, E. B., L. L. Schmitz, J. D. Faul, A. Gard, C. Mitchell, J. A. Smith, W. Zhao, D. Weir, and S. L. R. Kardia (2017). Heterogeneity in Polygenic Scores for Common Human Traits. *bioRxiv*, 106062.

Warner, J. (2005). *Perfect Madness: Motherhood in the Age of Anxiety Paperback*. New York, NY: Riverhead Books.

Warren, J. B. (2001). Changes with Age in the Process of Occupational Stratification. *Social Science Research 30*, 264–288.

Warren, J. B., J. T. Sheridan, and R. M. Hauser (1998). Choosing a Measure of Occupational Standing: How Useful Are Composite Measures in Analyses of Gender Inequality in Occupational Attainment. *Sociological Methods & Research 27*(1), 3–76.

Warren, J. R., J. T. Sheridan, and R. M. Hauser (2002). Occupational Stratification across the Life Course: Evidence from the Wisconsin Longitudinal Study. *American Sociological Review 67*(3), 432–455.

Watanabe, K. et al. (2019). A Global Overview of Pleiotropy and Genetic Architecture in Complex Traits. *Nature Genetics 51*(9), 1339–1348.

Wertz, J., T. E. Moffitt, J. Agnew-Blais, L. Arseneault, D. W. Belsky, D. L. Corcoran, R. Houts, T. Matthews, J. A. Prinz, L. S. Richmond-Rakerd, K. Sugden, B. Williams, and A. Caspi (2019).

Using DNA from Mothers and Children to Study Parental Investment in Children's Educational Attainment. *Child Development*, 1–17.

West, R. (2007). *Marriage, Sexuality, and Gender*. Routledge.

Woelfel, J. and A. O. Haller (1971). Significant Others, The Self-Reflexive Act and the Attitude Formation Process. *American Sociological Review 36*(1), 74–87.

Yi, J., J. J. Heckman, J. Zhang, and G. Conti (2015). Early Health Shocks, Intra-Household Resource Allocation, and Child Outcomes. *Economic Journal 125*, F347–F371.