THE INTERGENERATIONAL TRANSMISSION OF EDUCATIONAL ATTAINMENT REVISITED: THE EFFECTS OF SOCIOECONOMIC BACKGROUND, GENETIC INHERITANCE, AND COHORT

Meng-Jung Lin

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Sociology in the College of Arts and Sciences.

Chapel Hill
2017

Approved by:

Guang Guo

Ted Mouw

Francois Nielsen

ABSTRACT

Meng-Jung Lin: The Intergenerational Transmission of Educational Attainment Revisited: The Effects of Socioeconomic Background, Genetic Inheritance, and Cohort
(Under the direction of Guang Guo)

This research revisits the intertwined social and biological pathways of the intergenerational transmission of educational attainment. By estimating the effects of the whole-genome genetic variants by the continuation ratio logit regressions using 8,251 samples from the Health and Retirement Study (HRS), and considering for socioeconomic status in childhood on education at the same time, I first examine the relative individual impacts of biological and social influences. Then, I consider how parental education shapes the expression of the genetic potential by including moderating effects between the two. Finally, I explore the curvilinear trend of genetic effects over time, and use cohort separated models to investigate the decline in the moderating effects of parental education on educational attainment. The findings suggest the influences are from both genes and family socioeconomic background. Also, the genetic effects were not only negatively moderated by socioeconomic background, but changed curvilinearly over time corresponding to the expansion of higher education in the mid-twentieth century in the U.S. The pattern indicates the educational opportunities equalized at first but saturated after higher education became more accessible. This study furthers the understanding of the social mobility process and provides suggestions for policymakers on education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In the U.S., children born to the families in the poorest 20 percent of the income distribution have barely about 9 percent of chances to rise to the richest 20 percent as adults for the last two generations (Chetty et al. 2014). While the figure remains stable for the 1970-80s birth cohorts, recognition of the unequal intergenerational mobility of the U.S. society has increased lately. The upward and downward intergenerational mobility occurs depends largely on both parental and children's educational attainment (Blau and Duncan 1967). However, although social scientists interpret the transmission of educational attainment mainly through social inheritance, it is arguable that genetic heritability also plays a role in the mobility process (Eckland 1967; Duncan 1968; Behrman and Taubman 1989; Jencks and Tach 2006; Nielsen 2006; Nielsen and Roos 2015). Failure to discern social and biological pathways leads to a weak standpoint for sociologists' belief that the transmission of educational achievement operates socially from family backgrounds, and it may simply represent the effects of a genetic predisposition underlying the process. Past studies using twins and siblings to address the issue were not able to account for the specific genetic effects due to the unclear identification of the shared genes, and the difficulties to distinguish genetic effects from environmental effects because the identical twins are more likely to pursue alike environment more than fraternal twins do. Therefore, by incorporating the polygenic score of education, a measure that summarizes the effects of specific genetic variants that are associated with education, this study attempts to answer the question directly.

To sociologists, distinguishing both social and genetic pathways could illuminate familial influences in the intergenerational transmission of educational attainment. Furthermore, the integration of genetic effects not only can reduce the bias of the past

sociological research, but it also helps to facilitate the examination of how environment shapes the expression of the genes. The realization of the genetic potential, which refers to the actualization of the innate ability, can be considered as a signal for the equalization of opportunities to attain higher education. When the environment provides individuals with appropriate resources, their genetic potential can actualize more fully than when the resources are barren. For example, in terms of proximate surroundings, family backgrounds might limit or encourage the full achievement of the genetic predisposition. Several hypotheses have been proposed for the gene-environment interaction pattern on status-related outcomes (Nielsen 2016). The directions of the moderating effects of the environment can be positive (Scarr-Rowe hypothesis), curvilinear (Pareto hypothesis), or negative (Saunders hypothesis) under different circumstances. When families with high status are able to encourage the genetic effects of their offspring and it is not the case for the low status families, the Scarr-Rowe hypothesis is supported, whereas when these high status families are merely able to protect their offspring from downward mobility and the genetic effects express more fully for the low status families, the Saunders hypothesis is evidenced. Somewhere in between, the Pareto hypothesis would be true if both the highest and the lowest status are crystallized, and the middle class is the only class to mobile. Empirically, studies mostly support the Scarr-Rowe hypothesis. A case would be the children from disadvantaged backgrounds often suffer from the constrained chances to fulfill their potentials (Guo and Stearns 2002). However, since studies have seldom tested these alternative hypotheses, this paper will examine them using the interaction terms between the family backgrounds and the genetic polygenic score.

Also, with regard to the macro environment, historical changes could suppress or enhance the genetic effects on attainment. Research has shown that as educational policies became liberal in the second half of the twentieth century, genetic potential turned out to be a prominent factor for educational attainment (Heath et al. 1985). While on the other hand, the

universal access to higher education might contribute to the weaker genetic effects for the later born cohorts (Okbay et al. 2016). The inconsistent findings can be reconciled by considering the genetic effects changed curvilinearly through the expansion of higher education. If the curvilinear genetic effects have synced with the declining or the inverted U-shaped effects of family socioeconomic status and their downward moderating effects on genes over time, these would suggest an equalization process during the time. In contrast, if the family socioeconomic status effects increase, the reducing genetic effects and growing family backgrounds effects depict an unequal society in development.

In this study, I will examine both the moderating effects of family backgrounds and historical changes on the relationship between genetic variants and educational attainment. Since the United States has undergone the expansion of higher education gradually and stably in the twentieth century, and the sample I use constituted of adults above age 50, they could have experienced the growing opportunities to obtain higher education to different extents if they were born in different periods of time. The loosening constraints of attaining higher education might thus result in the better likelihood of realizing their genetic potential for education in the younger cohorts within these old adults. At the same time, the saturation of higher education which refers to the lower thresholds to entering college education would also be possible to reduce the effect of genes.

This study uses the nationally representative data of older adults from the Health and Retirement Study (HRS) to examine three related issues on educational attainment by applying the polygenic scores constructed from the recently GWAS results (Okbay et al. 2016). First, I examine the relative magnitude of the effects of social inheritance and genetic heritability on offspring's education. Second, I investigate how socioeconomic status moderates the realization

of genetic potentials and which gene-environment interaction hypothesis on educational attainment is supported. Third, to understand whether the U.S. society equalized or became unequal in the mid-twentieth century, by considering cohort differences further, I test the argument that historical changes influence the structural opportunities for individuals to achieve their genetic potentials, and the changing effects of family socioeconomic status over time. The study can contribute both academically and publicly by providing new insights into to the long-lasting issue of social mobility, and advising how public policies can help to equalize the opportunities for educational attainment.

BACKGROUND

Educational Attainment and Social Mobility

Individuals possess two statuses, ascribed status and achieved status, which also describes the processes through which one obtains position in society. Ascribed status refers to the status individuals are born with. For example, gender, ethnicity, genetic predisposition, parental education, and family socioeconomic status, are all determined before their birth and can rarely be changed. In contrast, achieved status is the status that is achieved by the individual. Achievements such as educational and occupational attainment are considered achieved statuses. In this distinction, sociologists often consider the realization of the latter as an indicator of social mobility in the society. Along the same lines, Blau and Duncan's (1967) seminal status attainment model demonstrates the paths among ascribed statuses and achieved statuses. Their analysis illustrates how a father's education and occupation are influential factors in the respondent's educational and occupational attainments, and thereby highlights the effects of the ascribed status on social mobility. However, Blau and Duncan still argued that, "(self-)education operates *primarily* to induce variation in occupational status that is independent of initial status (pp.203)," which maintains the role of education as an equalizer that ameliorates the reproduction of social status. While the "vicious cycle" of reproducing social status across generations might not be true, and education might be an equalizer, it is confirmed that parental education and occupation are important factors in their offspring's attainments. Nevertheless, it remains unclear how the ascribed status, in this case parental education, affects children's educational attainment.

Studies have been done to understand this black box. Soon after Blau and Duncan (1967), the Wisconsin Model tried to explain the intergenerational transmission of educational attainment by taking social psychological variables into account, such as the influence and aspirations of significant others (Sewell, Haller, and Portes 1969). Using a broader framework, Jonsson et al. (2011) provides a set of mechanisms analyzing intergenerational reproduction of occupation, which is tightly linked with education. Under their framework, four kinds of resources underlie the mechanisms: human capital, cultural capital, social network, and economic resources. Human capital includes the cognitive skills and abilities the class members share and the families have. Cultural capital refers to the culture and taste enjoyed by the class, and the aspirations from parents. Social network indicates the social ties that the neighborhoods and family members own, which can possibly connect to better resources. And economic resources are the incomes and businesses the class and the families have. By transmitting all of these resources to children, parents reproduce their advantage or disadvantage in the next generation.

Although these studies have established abundant accounts for the reproduction of education, their explanations are often built upon the assumption that ascribed statuses transmit effects socially. That is, only through resources can ascriptive characteristics other than genetic predisposition affect achieved statuses. However, biological heritability between parents and children also connects parents' achievements and children's. The overlapping pathways entangle the social and biological mechanisms together, and henceforth, genetic pathway may confound the social influences. In the next section, I will summarize the studies that attempt to integrate biological factors to solve the intertwined explanations.

The Integration of Biological Accounts

Sociologists have long recognized the possibility to incorporate genetics into accounts to explain the intergenerational mobility. Forty years ago, Eckland (1967) stated that although environmental components are relatively obvious, genetic factors are not ignorable for IQ performance. And since it was infeasible at that time to have data and methods to discern between hereditary and environmental components, researchers tended not to untwine the two (Duncan 1968). A sociological work using the method closer to the one used now was Scarr and Weigberg's (1987) study on the IQ of adoptees and biological children. They reported the strong effects of the biological parent's IQ rather than the social parental IQ on the adoptees and thus suggested the genetic effects account for a large portion of the effects of family background. However, their estimates were still crude since the biological parental IQ only played as a proxy for genes in the study.

Interests in the issue were resurgent as quantitative genetics developed. Researchers began to collect twins and siblings' data to analyze the social and biological influences on social mobility. Using data of U.S. male twins who were born between 1917 and 1927, economists Behrman and Taubman (1989) found that above 80 percent of the observed variation in schooling can be attributed to genetics than to the environment. Also, by implementing the ACE models, which decompose the total variance in the outcome variables into heritability, shared environment (i.e., environment that siblings share and differ between families), and nonshared environment (i.e., measurement errors and individual-specific differences), in a large sibling sample, Nielsen's (2006) results showed that for adolescents' verbal IQ, grade point average, and college plans, genetic component explains about 50 to 70 percent of the variances, unshared environmental component accounts for 30 to 40 percent, while shared environment explains only a 0 to 10 percent. Both these results indicate a relatively strong genetic inheritance of educational

attainment in the offspring.

Researchers can consider the genetic component as an indicator of the opportunity for success. Although genes cannot be changed after conception, and it is fair to consider it as an ascriptive characteristic, the expression of it can be shaped by the environment (Bronfenbrenner and Ceci 1994; Perry 2002; Shanahan and Hofer 2005). Especially under the circumstances where barriers are small and resources are adequate, maximizing the potential of the genes is more likely. Therefore, when opportunities to realize one's genetic potential become higher in the society, the influence of genetic components would also become more prominent. According to this argument, the results from Behrman and Taubman (1989) and Nielsen (2006) indicating the society in the twentieth century was relatively equalized, which allowed individuals to realize their innate potential to a larger extent, since the genetic component explained more and shared environment explained less of the variances in the educational attainment.

Furthermore, to untangle the gene-environment interaction patterns further, Nielsen (2016) summarized three alternative hypotheses on the gene-environment interaction on status-related outcomes. First and the most popular one is the Scarr-Rowe hypothesis (Tucker-Drob and Bates 2016) which argues that genes express more thoroughly when the family socioeconomic status becomes better. Another possible pattern of this hypothesis is the initial increment of the gene expression when socioeconomic status is low, but it slows down after the environment reaches a threshold. So, the relationship between the socioeconomic status and the expression of genes is a positive linear line or at least a positive relationship at first.

The second hypothesis was argued by Pareto (Pareto 1909). In his hypothesis, genes express to a peak when the individuals are from middle class families, but genes express weakly in both the poorest and the wealthiest families for the reason that the environments are too harsh and

suppressed for the poorer children to mobile up, and too protective and abundant for the rich children to mobile down. Hence, Pareto hypothesized a curvilinear model between gene expression and socioeconomic background.

Finally, the Saunders hypothesis (Saunders 2010) suggests a "reverse Scarr-Rowe hypothesis." In Saunders' analyses on British data, he found that social mobility in Britain depends on meritocracy to a large extent. When considering the effect of intelligence as a measure of meritocracy, and assuming it is inheritable via genes, the predicted intergenerational social mobility pattern is almost the same as the actual pattern. However, although Saunders describes the British society as a more open society than expected, he does claim that the middle class families still have the advantages of preventing their offspring from falling into working class. The "stickiness" (Saunders 2010: 36) of the middle class is shown by the fact that children from working class are required to have higher IQ scores than their counterparts from middle class to enter the service occupations. Therefore, according to Saunders findings, the extended hypothesis maintains that the gene expression is constrained by the high status families since they preserve the opportunities for their children to obtain higher education and positions irrespective of their innate abilities. However, unlike Pareto's hypothesis, Saunders did not hold that the low status families restrict the gene expression. As a consequence, the Saunders hypothesis is a negative linear line between socioeconomic status and gene expression.

Empirically, the interaction terms between the genetic polygenic score and the family socioeconomic status should behave in a certain way if any of the above arguments are true. To support the Scarr-Rowe hypothesis, the interaction term should be positive, meaning that the effects of the genetic component become larger in the higher status families. In contrast, if the Saunders hypothesis is true, the effects of the genetic polygenic score would be weaker across

different socioeconomic statuses, and the interaction term is negative. And in the middle ground, if Pareto's hypothesis is correct, the interaction term would be positive for the middle class families, but be negative or less positive for the lowest and the highest status families.

Research also has tested the environmental influences on the realization of genetic potential, and most of them support the Scarr-Rowe hypothesis. For example, Guo and Stearns (2002) used a large sibling sample to study the heritability and the social influences on intelligence. Their results showed that for children who live under the disadvantaged environments, the realization of the genetic potential will be limited. Other studies also showed that genes only explain a little variation in IQ for children raised in low socioeconomic status families when they are at age 7 or even 2-year-old, while it accounts for 50% or more for children from affluent families (Turkheimer et al. 2003; Tucker-Drob et al. 2010).

The estimates of genetic effects and heritability provided by twin studies paved the way for furthering the understanding of the effects of both genetics and environment. However, although these studies have attempted to solve the interwoven pathways, where genetics confound the social pathways, these analyses from twins and siblings did not take the genetic effects into account precisely. Since the method could not identify the specific genes and the overlapped genes within pairs, and it also fails to distinguish genetic effects from environmental effects, which might become problematical as the equal environments assumption (EEA) could be violated when twins sharing the same genes tend to seek similar environment, it is unclear what genes are being considered when comparing identical twins with fraternal twins or between sibling pairs (Freese 2008). Comparing the outcome differences among paired samples therefore could not solve the issue directly.

In the recent decade, genome-wide association studies (GWAS) have provided new

opportunities for social scientists to incorporate the results into studies, providing researchers with chances to solve this interwoven issue. The GWAS is a hypothesis-free method used to identify the single nucleotide polymorphisms (SNPs) among the whole genome (around 1 to 2 million SNPs) that associate with the phenotype or the trait significantly (Belsky and Israel 2014). A SNP is a base difference on the specific position of a gene that may vary across individuals. It is a form of mutation that might result in individual differences in traits or diseases. The method corrects the potential statistical artifacts by implementing stringent significance level, where the p-value is required to be lower than 5×10^{-8}. And therefore, the GWAS study needs large sample size, usually above tens of thousands of individuals, to maximize the statistical power (Belsky and Israel 2014). In some cases, loci might be reported along with SNPs because the SNPs are too small and can be correlated with other variants in the same region, studies often times also report the associated region (i.e., loci) where the SNPs situated (Wray et al. 2014).

Purcell et al. (2009) suggests that researchers can combine GWAS results into their studies by using the polygenic scores that generated from the significant SNPs. To construct the score, researchers need to sum the risk alleles of the SNPs the individual has. Usually, there are only zero to two risky variations (i.e., nucleotides) for each SNP. The number of the risk alleles an individual has can be related to the degrees of the expression of the disease or the trait. There are two approaches to construct the score. One is the top-hits approach which only includes the SNPs with p-values lower than 5×10^{-8} that contribute more to the phenotype, the other approach is the whole-genome approach which assumes the infinitesimal contributions of a large number of SNPs and uses the whole-genome genetic variants that are significant at a higher level (e.g., p<.1). The score can also incorporate weights from the effect sizes resulted from the

ordinary least squares (OLS) regressions of individual SNPs in the GWAS, and thereby takes the contribution of each SNP into account (Belsky and Israel 2014).

The GWAS on educational attainment have shown that there are several SNPs significantly related to it. Using data from 126,559 individuals, Rietveld et al. (2013) identified three independent SNPs (rs9320913, rs11584700, and rs4851266) that relate to either years of education or college completion. However, the effect sizes of the SNPs are only about one month of schooling for each allele. And the linear polygenic score of these SNPs can only account for two percent of the variation in educational attainment. Nevertheless, the results were also replicated later (Rietveld et al. 2014). More recently, Okbay et al. (2016) found 74 loci with $p<5 \times 10^{-8}$ that are associated with educational attainment by using a sample of 293,723 individuals. The estimated effects of these 74 loci range from 2.7 to 9.0 weeks of schooling individually. And the highest increment in R^2 is up to 0.035%. In this study, I will construct the polygenic score by using the recently reported whole-genome effect sizes from Okbay et al. (2016) to measure genetic effects on educational attainment directly.

Using the earlier effect sizes reported by Rietveld et al. (2013), social scientists have made some progress in the field of social mobility. Conley et al. (2015) used the polygenic score based on the whole-genome SNPs with the relaxed significance threshold from the Rietveld et al.'s study to predict education in the Framingham Heart Study (FHS) and Health Retirement Study (HRS). They found that one-sixth of the correlation between parental and children's education can be explained by genetic inheritance, and the genetic effect does not vary by maternal education once children's genetic score is controlled. They concluded by suggesting that the policies focusing on equalizing educational opportunities might have a trivial impact on intergenerational mobility, since parental education could not moderate the genetic effects.

Besides adults' educational attainment, studies have shown that the polygenic scores of these three SNPs are positively associated with adolescent's educational achievement (Benjamin et al. 2015), can explain about at least three percent of the variance in children's educational achievement (Krapohl and Plomin 2015), has an interaction effect with fathers' social class when predicting education, and even is strongly associated with income at age 46 (Davies et al. 2015). However, as new loci identified and the better-powered genetic risk scores are developed, studies are needed to confirm or challenge the previous results. Henceforth, to compare with the past studies, this study will not only use the whole-genome SNPs with the effect sizes from the Okbay et al.'s (2016) study to test the genetic and social pathways, but also examine which gene-environment hypothesis is true for the older U.S. adults.

In light of the theoretical review above, using the new method, I will test the following two sets of hypotheses:

Hypothesis 1: Both socioeconomic status and genetic predisposition have positive impacts on educational attainment.

Hypothesis 2a (Scarr-Rowe Hypothesis): Socioeconomic status positively moderates the genetic influences on education. Individuals from advantaged backgrounds will have better opportunities to reach their potential of their genetic predisposition, whereas the opposite might be true for their disadvantaged counterparts. In this case, the interaction term would be positive.

Hypothesis 2b (Pareto Hypothesis): The genetic influences peak at the middle level of socioeconomic status, but depress at both the lowest and the highest ends of socioeconomic

status. The interaction term would be positive for the middle class, but be negative or less positive in the highest status.

Hypothesis 2c (Saunders Hypothesis): Socioeconomic status negatively moderates the genetic influences on education. The most advantaged families are capable of protecting their offspring mobile downward, so genes do not matter much for them. However, genes would be the key for the poor to mobile upward. For this hypothesis, the interaction terms would be negative.

Historical Changes and Genetic Effects on Educational Attainment

The gene by environment interaction (G×E) covers the impacts of the macro historical changes in addition to the influences from the proximal surroundings on the individuals. The expression of genes can be suppressed or encouraged by the external or policy changes. For example, Branigan et al.'s (2013) meta-analysis of thirty-four cohorts on educational attainment across countries found that genetic component can explain more variance in education for men and those who born in the latter half of the twentieth century, and vice versa for women and individuals born earlier. As for the United States, Nielsen and Roos (2015) used the recent sibling data to estimate the fractions of heritability, shared environment, and nonshared environment components in educational attainment, and compared their results to other studies. They found that the variance in education explained by genetic component declined, whereas the portion explained by shared environment increased. Since genetic potential expresses more fully when the society provides appropriate opportunities, the decline impacts of genetic component indicates the opportunity to attain higher education has become more unequal over the last six decades in the United States.

Another impressive case of the macro environmental effect is Heath et al.'s (1985) study of Norwegian twins. They found that family background had larger impacts on the educational attainment of Norwegians born before 1940 than after. Furthermore, the patterns varied between genders across time periods. While the variances accounted by genetic predisposition increased for males after the World War II, it remained relatively stable for females in the same periods. The authors maintained that the main explanation for the general increase in the fraction of heritability was due to the adoption of the liberal social and educational policies of the Norwegian government after the WWII, as well as the fact that more opportunities were available for males than females at that time.

Although the above studies suggest that the liberalization of the society would encourage the expression of genes because of the greater opportunities but vice versa when the society becomes conservative, the universality of the chances to enter higher education might obscure the effects of genes in the liberalized society. Under this circumstance, the effect of genes declines over time. For example, several studies below have shown the decreasing genetic effects across cohorts. But their results do not necessarily suggest the more unequal society is developing.

In a recent work using the whole-genome polygenic score from Rietveld et al. (2013), Conley and Domingue (2016) found that the effect of polygenic score becomes weaker in the later birth cohort. In addition, if separated the sample into different educational transition stages as Mare (1980) did, the negative interaction term between the polygenic score and the birth cohort in the full sample is contributed by the lower educational transitions, while it is positive in the highest educational transition. The authors explained the results by the maximally maintained inequality theory (MMI) which maintains that as the lower levels of education expand, the

entrances into them become less unequal. And since the highest educational institutions have expanded relatively slowly, the unequal opportunity of entrance remained at the highest level. Also, using the Swedish Twins Registry data in 1929-1958, Okbay et al. (2016) reported the decreasing effect of their all-SNP score throughout the birth cohorts. They interpreted their results as a consequence of the liberal reform of the educational system undergone in the 1950s and 1960s, which extended the compulsory education and postponed the educational tracking.

However, it is possible that at the beginning of the liberalization process, those who are talented innately would be able to grip the marginally increased chances to enter higher education. But as higher education becomes nearly universal, and almost everyone can access it, both the selectivity of higher education and the variation of the educational attainment drops, and therefore the genetic effects decline. This process suggests a curvilinear change in the genetic effects which means the effect of genes might increase when the expansion of educational institutions begins, and decreases after the higher levels of education become saturated.

The expected trend stated above corresponds to the saturation argument Raftery and Hout (1993) theorized within their maximally maintained inequality (MMI) hypothesis. The MMI hypothesis claims that the expansion of higher education, although aims at equalizing the impacts of family origins on educational attainment by increasing educational opportunities, as the supply of the targeted level of education surpasses the demand in the society, the familial influences decrease at the particular level, but transfer to the next level. Thus, the inequality persists at the maximum level of education whenever there is at least a higher level that is not saturated. Saturation here refers to the likelihood that all the offspring from the advantaged families attain the certain level of education. For example, when all the children from the wealthy families obtain a high school diploma, the high school level of education is saturated,

the odds ratio of attaining secondary education decreases for the group, meaning that the inequality to attend it diminishes from then on if the given level keeps expanding.

From this perspective, if the effect of genes is regarded as meritocratic ability, as one level of education is saturated, its influence might also decline as the effects of family resources do. As shown in Roksa et al.'s (2007) study, although the U.S. higher education has never reached saturation before, and parental education has been influential over time, even greater in the recent cohort than ever, parental education is slightly weaker for the post-World War II cohort who are born before 1970s. Furthermore, their results also suggested that father's occupation has a smaller impact on students who go to college in the 1980s. And this particular group who benefits most from the educational expansion is the youngest group surveyed by the data used in this study, the Health and Retirement Study.

Therefore, to examine whether the educational opportunity in the U.S. society became unequal or equal in the earlier decades, two hypotheses can be tested. If the unequal transformation was true, the genetic effects would decline over time, but the effect of family socioeconomic status would increase across cohorts. On the other hand, if the U.S. society turned more equal, the effects of genes should increase and the effects of family backgrounds would decrease. An alternative of this second statement is the saturation argument that the effect of genes might raise first, but declines latter, and the family background effect should decline or have a similar curvilinear trend because children from resourceful families tend to seize the chances first.

In this study, I will use cohorts to capture the effects of historical changes on the relationships between genetic transmission, social inheritance, and educational attainment. The sample I use consists of middle-aged to older individuals ranging from those born before 1924 to

those born in 1959, a group that exposed to federal policies encouraging youths to pursue further education and the expansion of higher education (Trow 1972, 2007; Mumper et al. 2011). Therefore, it is likely that the opportunities to realize the genetic potential would be better in the younger cohorts; hence the genetic effects on education will be greater among the younger cohorts. Or, the universal accessibility of higher education would lead to an inverted U-shaped trend of the genetic effects. At the same time, the effects of family backgrounds might become smaller for the younger cohorts, or undergo the same curvilinear trend as resourceful individuals enjoyed more advantages at the beginning of the expansion, but the benefits declined for the later cohorts. On the contrary, the rising college tuition in the latter half of the twentieth century might result in the decreasing genetic effect, but increasing family background effects across birth cohorts.

Along with the same reasoning, if the educational opportunities were truly equalized, the conditioning effects of family backgrounds would decline in the latter cohorts irrespective of whether higher education is saturated or not. However, there is no explicit pattern of interaction hypothesized if the society became unequal. So, while the declining conditioning effects cannot rule out the more unequal society hypothesis, it strengthens the equalization hypothesis if the main effects of genes and socioeconomic status support the hypothesis. Although past studies have examined the changing impacts of genetic components over time, they seldom tested the changing moderating effects of family background on genetic expression over birth cohorts. Therefore, in this study, I will investigate the changes of the effects of genes, family socioeconomic status, and the interaction between them across cohorts.

In consideration of the above reviews, the third hypotheses set and the fourth hypothesis are:

Hypothesis 3a (Equalization): The genetic effects on educational attainment are greater for the younger cohorts, while the effects of socioeconomic status are smaller for the younger cohorts or have a stronger impact for the middle cohorts but smaller for the younger cohorts.

Hypothesis 3b (Equalization with saturation): The genetic effects on educational attainment increase at first, and decrease for the younger cohorts, whereas the socioeconomic status effects decline over time or become strong at first but decline later.

Hypothesis 3c (Became unequal): The genetic effects on educational attainment decrease across birth cohorts, and the effects of socioeconomic status increase in the meanwhile.

Hypothesis 4: The moderating effects of socioeconomic status on education became weaker for the younger cohorts.

DATA AND METHODS

Data

I use data from the Health and Retirement Study (HRS) (http://hrsonline.isr .umich.edu/) for the analysis. HRS is a national representative survey of adults over age of 50 in the U.S. It is a longitudinal study which has been continuously administered since 1992 with data collected every two years. The National Institute on Aging (NIA) sponsors the study and the Institute for Social Research (ISR) at the University of Michigan collects the data. Information on social, economic, and other factors related to the antecedents and consequences of retirement were included in the data collection. The study also collected genetics data by asking respondents

to provide saliva specimens in 2006 and 2008, thus allowing me to test the effect of genes on education. Among the sample cases, 13,129 samples were put into genotype using the Illumina HumanOmni2.5-4v1 array at the Center for Inherited Disease Research, and 12,507 passed the Quality Control process at the Genetics Coordinating Center of the University of Washington. However, since the GWAS results for educational attainment was based on Caucasians (Rietveld et al. 2013; Okbay et al. 2016), to align with the GWAS analysis, after excluding other races and ethnic groups, 9,215 self-reported non-Hispanic whites remain in the analytic sample. In addition, there are 964 cases with missing values on parental education, by using list wise deletion, the final sample size is 8,251.

Variable Measurement

Educational attainment

The outcome variable in this study is educational attainment. The respondents were asked "What is the highest grade of school or year of college you completed?" The answer ranged from 0 to 17 and above. Given individuals are required to pass one educational level to the other, and must decide whether entering the next stage or not, I used the categories reconstructed from the HRS 2014 Tracker file instead of the continuous years to measure educational attainment. These categories were then recombined into: No degree, GED/High school diploma, Two year college degree/Degree unknown/Some college, Four year college degree, and Master/Professional degree (Ph.D., M.D., J.D.).

Childhood socioeconomic status

Childhood socioeconomic status is captured by parental education. Although earlier

research often uses father's education as an indicator of parental education, mother's education is also taken into consideration in this study because mothers are closer to the child and usually the caretakers of the child. I use the highest years of education of father and mother as parental educational attainment measures. Parental educational attainment is based on the following two questions: "What is the highest grade of school your mother completed?" "And what is the highest grade of school your father completed?" The answers were also ranged from 0 to 17 and above. Parental education is also standardized according to the cohort the respondent belongs to reflect the differential distribution of parental educational attainment across cohorts.

Polygenic score

I construct the polygenic score by using the GWAS results without HRS and 23andme from Okbay et al. (2016). The whole-genome single-nucleotide polymorphisms (SNPs) from the HRS imputed genetics data were extracted. I use the PRSice program to detect the best predicting polygenic score threshold, which reports the threshold as 0.56 significance level. Therefore, I calculate the weighted average of risk alleles by weighting the risk alleles of 99,239 SNPs by their effect sizes (betas) which were significant at 0.56 level reported in Okbay et al.'s study.

The polygenic score can also be calculated from the 74 top hits. I show the results in Appendix B and compare them with the whole-genome polygenic score in the conclusion and discussion section. Although 74 single-nucleotide polymorphisms (SNPs) were significant at 5×10^{-8} level in Okbay et al.'s study, only 73 SNPs (rs12772375 unavailable) from the HRS imputed genetics data are available after substituting six SNPs (rs17824247, rs2964197, rs2431108, rs261591, rs13294439, and rs17119973) with proxy SNPs. Therefore, the results in Appendix show the effects of the 73-SNPs polygenic score.

Cohorts

Both the continuous ages and the categorical cohorts are used to examine the historical changes hypotheses. I calculate age by subtracting respondent's birth year from 2006, the year when HRS first collected their genetics data. As for cohorts, six cohorts are classified according to the HRS survey design. These cohorts are mainly born between certain years and are named as: Aging & Health Dynamics cohort (AHEAD), born between 1905 and 1924; Children of the Depression cohort (CODA), born between 1924 and 1930; Health and Retirement cohort (HRS), born between 1931 and 1941; War Babies cohort (WB), born between 1942 and 1947; Early Boomers cohort (EBB), born between 1948 and 1953; and Mid Boomers cohort (MBB), born between 1954 and 1959. These cohorts were entered and surveyed in different years: AHEAD entered in 1993 and was surveyed in 1995, and 1998 to 2012; CODA was surveyed from 1998 to 2010; HRS was surveyed from 1992 to 2012; WB was surveyed from 1998 to 2012; EBB was surveyed from 2004 to 2012; and MBB entered in 2010 and was also surveyed in 2012.

Control variables

Control variables include gender, region (where the respondent was born), and degree of urbanization (depends on where the respondent lived at age 10).

Population stratification will also be considered since the allele frequency differences due to systematic ancestry differences can result in the spurious associations between SNPs and traits. For example, one of the identified height related SNPs is strongly associated with the European ancestry (Campbell et al. 2005). To deal with this population stratification issue, researchers conducted principal components analyses to identify the potential ancestral differences in SNPs

in genotype data (Price et al. 2006). Usually, at least 10 largest principal components were controlled in the studies. Therefore, I will report the findings with and without controlling the largest 10 principal components to compare the results.

Analytic Strategy

The continuation ratio logistic model will be used in the analysis. Individuals make decisions along the way throughout their educational career. The process requires them to pass through the ladder to enter the next stage. Based on the continuation ratio model, researcher can study the determinants of individuals' transition between stages when they fulfill the requirement or complete the previous stage. The Mare model, which relies on the continuation ratio logit model to estimate the odds of completion of a certain level of degree, is widely used in educational stratification field (Mare 1980; Ganzeboom, Treiman, and Ultee 1991). The conditional probability is defined as given in a certain stage, the probability of advancing to the next stage, which for the j th category in J categories is

$$\Pr(y > j | y \geq j)$$

Let y equals five levels of education (LHS: less than high school; HS: high school; SC: some college; CO: 4-year college; GR: graduate school), the above probability provides the base for the four logit equations as below:

$$\ln \frac{\Pr(y > LHS | y \geq LHS)}{1 - \Pr(y > LHS | y \geq LHS)} = \theta_{LHS} + x'\beta$$

$$\ln \frac{\Pr(y > HS | y \geq HS)}{1 - \Pr(y > HS | y \geq HS)} = \theta_{HS} + x'\beta$$

$$\ln \frac{\Pr(y > SC | y \geq SC)}{1 - \Pr(y > SC | y \geq SC)} = \theta_{SC} + x'\beta$$

$$\ln \frac{\Pr(y > CO | y \geq CO)}{1 - \Pr(y > CO | y \geq CO)} = \theta_{CO} + x'\beta$$

where xs are the covariates, θs are the cut points for each category compares to the lowest category, and βs are the coefficients of the covariates which are assumed to be the same across contrasts. So the general continuation ratio logit model is

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)} = \theta_j + x'\beta$$

When plug in the original probability, the model is

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)} = ln \frac{\sum_{k=j+1}^{J} \Pr(y = j)}{\Pr(y = j)}$$

The continuation ratio model is similar to the ordered logit model in the sense that they only have one set of coefficients. That is, both require the proportional odds assumption. However, the continuation ratio model allows the interaction between the dummy stage variables and the other independent variables of interest, and therefore relaxes the assumption (Allison 2012: 186). If the totally unconstrained model is in request, separate ordinary binary logistic regression models which take the separated conditional samples into consideration can provide the same results with different parameters (Agresti 2006: 192). For example, the Mare model for educational attainment (Mare 1980) is a combination of separated binary logit models which uses the conditional samples. In this study, I will discuss the results of relaxing the effects of genetics across stages later in the discussion.

To examine my four hypotheses, I first include the parental education in the model and add the polygenic score to check the relative effects of social and genetic inheritance on educational attainment. The model with the main effects and control variables (x_s) is

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)} = \theta_j + \beta_1(Parental\ Education) + \beta_2(Polygenic\ Score) + \beta_s x_s$$

Second, I add the interaction terms of polygenic score and socioeconomic status in childhood to test the moderating hypothesis. To examine whether the data supports the Scarr-

Rowe hypothesis or the Saunders hypothesis, the model includes a single interaction term between parental education and polygenic score:

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)} = \theta_j + \beta_1(Parental\ Education) + \beta_2(Polygenic\ Score) + \beta_3(Parental\ Education \times Polygenic\ Score) + \beta_s x_s$$

If β_3 is positive, the result supports the Scarr-Rowe hypothesis, whereas the Saunders hypothesis is supported when β_3 is negative. As for the Pareto hypothesis, I include two more interaction terms in the model to assess the U-shaped relationship between family backgrounds and the expression of genes. The model is

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)} = \theta_j + \beta_1(Parental\ Education) + \beta_2(Polygenic\ Score) + \beta_3(Parental\ Education \times Polygenic\ Score) + \beta_4(Parental\ Education^2) + \beta_5(Parental\ Education^2 \times Polygenic\ Score) + \beta_s x_s$$

In the above equation, β_4 is the coefficient for parental education-squared, and β_5 indicates the moderating effect of parental education on genes when parental education is extremely high or low. As argued by Pareto hypothesis, β_5 should be negative to represent the crystallizing of the class structure at the highest and lowest ends of social class, and β_3 would be positive to support the hypothesis that children from the middle class families are more likely to realize their genetic potentials.

Third, the interaction terms between the polygenic score, age (x_3), and age-squared will be tested for the changing genetic effects over time. Also, I will examine the interaction terms between the cohort and the socioeconomic status at the same time to understand whether the U.S. society became equal or unequal in the earlier decades. The model will thus be

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)}$$

$$= \theta_j + \beta_1(Parental\ Education) + \beta_2(Polygenic\ Score) + \beta_3(Parental\ Education \times Polygenic\ Score) + \beta_6(Age) + \beta_7(Age^2) + \beta_8(Polygenic\ Score \times Age) + \beta_9(Polygenic\ Score \times Age^2) + \beta_{10}(Parental\ Education \times Age) + \beta_{11}(Parental\ Education \times Age^2) + \beta_s x_s$$

To support the equalization hypothesis, it is hypothesized that both β_8 and β_9 are negative, meaning that the genetic effects are smaller for the older cohorts but greater for the younger cohorts. As for the equalization with saturation hypothesis, only β_9 would be negative, whereas β_8 should be positive. These show that the middle cohorts enjoyed the benefits of the initial expansion of higher education, the older cohorts had no such opportunities, while there were too many chances for the younger cohorts to enter higher education so only a little variation left in educational attainment. Alongside with these two hypotheses, β_{10} and β_{11} can both be positive, suggesting that the effects of parental education become weaker in the younger cohorts, or β_{10} can be positive and β_{11} is negative, implying that children with higher parental education were more likely to enter higher education at the beginning of educational expansion, but the advantages diminished later on as accesses to higher education became universal.

The model also tests the hypothesis of became unequal. If β_8 is positive, and β_{10} is negative, regardless of the directions of β_9 and β_{11}, this hypothesis will be supported. To illustrate, the positive β_8 means the younger cohorts are less likely to actualize their genetic potential. Although this is in line with the equalization with saturation hypothesis, when combining with the negative β_{10}, which suggests a weaker impacts of parental education in the middle cohorts, but stronger impacts in the younger cohorts, the overall pattern would support the hypothesis that the U.S. society became unequal during the mid-twentieth century.

Finally, to test the changing moderating effects of socioeconomic status on the genetic

effects, I add a 3-way interaction term between polygenic score, parental education, and age in the model:

$$\ln \frac{\Pr(y > j | y \geq j)}{1 - \Pr(y > j | y \geq j)}$$
$$= \theta_j + \beta_1(Parental\ Education) + \beta_2(Polygenic\ Score) + \beta_3(Parental\ Education \times Polygenic\ Score) + \beta_6(Age) + \beta_7(Age^2) + \beta_8(Polygenic\ Score \times Age) + \beta_9(Polygenic\ Score \times Age^2) + \beta_{10}(Parental\ Education \times Age) + \beta_{11}(Parental\ Education \times Age^2) + \beta_{12}(Parental\ Education \times Polygenic\ Score \times Age) + \beta_s x_s$$

According to the declining moderating effects hypothesis, β_{12} should be negative in this case.

In addition to the analyses above, the sample will also be separated into three aggregated cohorts (1905~30, 1931~41, and 1942~59) to illustrate the declines of the moderating effects.

RESULTS

Descriptive Statistics

Table 1 shows the descriptive statistics of the variables used in this example. The table reveals some characteristics of the HRS dataset. First, fifty-seven percent of the sample earned their highest degree as GED or high school, which is in concordance with the requirement of compulsory education in the U.S. In the older cohorts, more individuals have no degree than the younger cohorts do, while a higher portion of the younger cohort own secondary and postsecondary degrees. These figures agree with the burgeoning opportunities for the younger cohorts to pursue higher education as a result of the expansion of the secondary educational institution in the twentieth century. Second, the standardized whole-genome polygenic score is around 0.000 after standardized within the analytic sample. The score is slightly lower in the younger cohorts than in the older cohorts.

Third, the summary statistics for parental education suggest that the mean years of parental education of the sample is around 11. The number is about 2.7 years higher for the youngest cohorts than the oldest cohorts. After standardizing parental education by cohorts, the mean is 0.000 for each aggregated cohort and overall.

Finally, since the dataset focuses mainly on the old population, there are more females (57.8%) than males due to the longer life expectancy of females. Also, the mean birth year of 1937 means that the respondents were about 70 years old when the genetic data were collected in 2006 and 2008. Geographically, more individuals are from the Midwest, and about half of the total respondents lived in the urban area when they were young.

Table 1. Descriptive Statistics for Full Sample and Different Cohorts

	Total		AHEAD & CODA 1905~30		HRS 1931~41		WB, EBB, & MBB 1948~59	
Variables	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Degree								
No Degree	.112	.003	.173	.008	.125	.006	.053	.004
GED/High School	.568	.005	.570	.011	.591	.009	.541	.009
2-yr/some college	.054	.002	.034	.004	.042	.004	.081	.005
4-yr college	.156	.004	.144	.008	.138	.006	.184	.007
MA/PhD	.110	.003	.078	.006	.104	.005	.141	.006
Year of Education	13.319	.028	12.885	.058	13.120	.044	13.860	.042
Standardized Polygenic Score	.000	.011	.111	.022	-.042	.017	-.033	.019
Parental Education in Years	1.961	.035	9.437	.061	1.898	.056	12.138	.052
Standardized Parental Education	-.001	.011	.000	.022	.000	.017	.000	.019
Female	.574	.005	.570	.011	.545	.009	.609	.009
Cohort	3.188	.013	1.597	.011	3.000	.000	4.554	.009
Birth year	1937.409	.112	1924.071	.101	1936.332	.055	1948.304	.078
Age in 2006/10	6.859	.011	8.193	.010	6.967	.006	5.770	.008
Region								
Northeast	.211	.004	.209	.009	.210	.007	.214	.008
Midwest	.351	.005	.342	.010	.352	.008	.356	.009
South	.265	.005	.260	.010	.296	.008	.233	.008
West	.112	.003	.143	.008	.093	.005	.111	.006
Other	.024	.002	.006	.002	.046	.004	.011	.002
Missing	.038	.002	.040	.004	.003	.001	.076	.005
Rural								
Urban	.525	.005	.532	.011	.497	.009	.553	.009
Rural	.443	.005	.405	.011	.475	.009	.435	.009
Missing	.031	.002	.063	.005	.028	.003	.012	.002
N	8,251		2,093		3,272		2,886	

Figure 1 illustrates the correlation between parental education and individual's polygenic score. The Pearson correlation coefficient is .112 and significant at p<.000 level, with the higher the parental education, the higher the polygenic score for education. Since parents contribute genes to their offspring, the figure suggests that other than transmitted socially, the effects of parental education can also be genetically transmitted, which has often been ignored in the sociological studies. Although the evidence would be clearer if parental genetics data are available, this figure shows at least a crude picture of the intertwined relationship pathways of intergenerational transmission of educational attainment. However, the binary correlation does not control for other variables, the analyses below will address the issue.

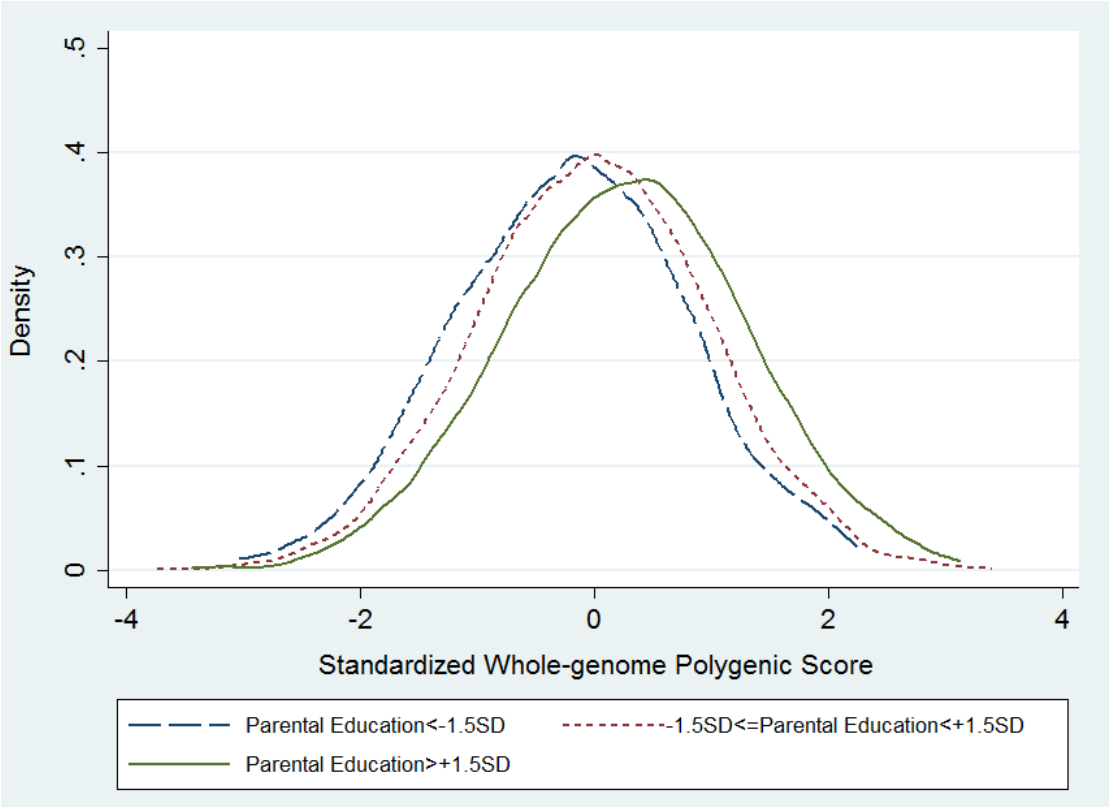

Figure 1. Correlation between the Standardized Parental Education and Offspring's Standardized Whole-genome Polygenic Score for Education. Parental education is standardized within the cohort the respondent belongs. Three levels of parental education are categorized according to their standard deviation away from the mean within the cohort. Of the three groups, parental education 1.5 standard deviation away from the mean in the negative direction is classified as the lowest group, and parental education 1.5 standard deviation away from the mean in the positive side is the highest group, while those have parental education between -1.5 and 1.5 standard deviation is the middle group. The X-axis is the standardized whole-genome polygenic score, and the Y-axis represents the probability density of it. In this figure, respondents with parental education higher than 1.5 standard deviation, on average, are more likely than those from the other two groups to possess a higher standardized whole-genome polygenic score.

Continuation Ratio Models Predicting Educational Attainment

Table 2 presents the results from the continuation ratio models. Model 1 and 2 are the traditional educational attainment models, which only include the demographic and socioeconomic background variables. The polygenic score was added to Model 3 to Model 8 in comparison to the two previous models. In Model 9 to Model 13, population stratification is controlled by entering ten principal components in the models.

The results from Model 1 and 2 support the traditional status attainment model. Model 1

considers the effects of parental education on offspring's educational attainment. In general, the more advantageous family the offspring was raised, the higher the education the offspring would attain. Individuals whose parents have one year more education are 1.78 times [$e^{1.775}$=1.775] as likely to advance into next stage as those whose parents have one year less. And, after other control variables are accounted for in Model 2, although the effect size declines slightly, the positive effect of parental education still holds.

The polygenic score is included in Model 3. The effect size of .329 indicates that a standard deviation increase in the polygenic score is associated with .39 times increase [$e^{0.329}$-1=0.389] in the likelihood to advance to the next educational level. Parental education enters in Model 4. On the one hand, the inclusion of parental education lowers the genetic effects to .294, while on the other hand, the effect of parental education also decreases from .574 in Model 1 to .554 in Model 4. After controlling other variables except for population stratification, the effects of both the polygenic score and parental education only fluctuate slightly, suggesting that other variables cannot explain away the main effects of both variables.

To examine the moderating effect of parental education, Model 7 tests the Scarr-Rowe hypothesis against the Saunders hypothesis by considering the interaction term between parental education and the polygenic score, while Model 8 further includes the interaction term between parental education-squared and the polygenic score to test the Pareto hypothesis. The significant negative interaction effect in Model 7 indicates that parental education negatively moderates the influences of genes. That is, the impact of genes on educational attainment decreases as parental education increases. This implies that children of highly educated parents are less likely to realize their genetic potential, but those with lower educated parents do. On the contrary, the interaction term of parental education and the polygenic score, and the interaction term between

Table 2. Continuation Ratio Model Predicting Educational Attainment

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
Standardized Whole-genome Polygenic Score (PGS)			.329***	.294***	.327***	.299***	.308***	.317***	.319***	.282***	.323***	.291***	.300***	.309***
			(.018)	(.018)	(.018)	(.019)	(.019)	(.022)	(.018)	(.019)	(.018)	(.019)	(.019)	(.022)
Standardized Parental Education	.574***	.543***		.554***		.526***	.531***	.534***		.559***		.532***	.537***	.540***
	(.019)	(.020)		(.019)		(.020)	(.020)	(.020)		(.019)		(.020)	(.020)	(.020)
Parental Education*PGS							-.065***	-.063***					-.065***	-.063***
							(.018)	(.019)					(.018)	(.019)
Parental Education2								-.001						-.002
								(.011)						(.011)
Parental Education2*PGS								-.009						-.010
								(.011)						(.011)
Female		-.380***			-.387***	-.377***	-.377***	-.377***			-.387***	-.375***	-.375***	-.375***
		(.036)			(.036)	(.037)	(.037)	(.037)			(.036)	(.037)	(.037)	(.037)
Age/10		-.248***			-.263***	-.268***	-.272***	-.272***			-.265***	-.270***	-.273***	-.273***
		(.018)			(.018)	(.018)	(.018)	(.018)			(.018)	(.018)	(.018)	(.018)
Region (0=Northeast)														
Midwest		-.147**			-.162***	-.141**	-.142**	-.142**			-.154**	-.107*	-.109*	-.109*
		(.049)			(.048)	(.050)	(.050)	(.050)			(.049)	(.051)	(.051)	(.051)
South		-.170**			-.257***	-.146**	-.142**	-.142**			-.245***	-.109*	-.105+	-.105+
		(.053)			(.052)	(.053)	(.053)	(.053)			(.053)	(.055)	(.055)	(.055)
West		.035			.098	.028	.030	.029			.105	.050	.051	.051
		(.065)			(.064)	(.066)	(.066)	(.066)			(.065)	(.066)	(.066)	(.066)
Other		-.046			-.057	-.042	-.039	-.040			-.037	-.042	-.040	-.040
		(.123)			(.121)	(.124)	(.124)	(.124)			(.122)	(.125)	(.125)	(.125)
Missing		-.119			-.218*	-.111	-.114	-.113			-.211*	-.089	-.091	-.091
		(.101)			(.099)	(.102)	(.102)	(.102)			(.100)	(.102)	(.102)	(.102)
Rural (0=Urban)														
Rural		-.435***			-.512***	-.423***	-.422***	-.422***			-.508***	-.407***	-.406***	-.406***
		(.037)			(.037)	(.038)	(.038)	(.038)			(.037)	(.038)	(.038)	(.038)
		-			-	-	-	-			-	-	-	-
Missing		1.678***			1.696***	1.616***	1.609***	1.610***			1.693***	1.622***	1.615***	1.616***
		(.118)			(.117)	(.119)	(.119)	(.119)			(.117)	(.119)	(.120)	(.120)
Stage (0=GED/HS vs. No Degree)														
	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2-yr/Some College	2.852***	3.025***	2.717***	2.912***	2.910***	3.087***	3.102***	3.103***	2.720***	2.916***	2.912***	3.089***	3.104***	3.105***
	(.045)	(.047)	(.043)	(.046)	(.046)	(.048)	(.049)	(.049)	(.043)	(.046)	(.046)	(.048)	(.049)	(.049)
						-	-	-				-	-	-
4-yr College	-.751***	-.954***	-.561***	-.828***	-.788***	1.034***	1.050***	1.051***	-.565***	-.835***	-.790***	1.038***	1.053***	1.054***
	(.065)	(.067)	(.063)	(.065)	(.065)	(.067)	(.068)	(.068)	(.063)	(.065)	(.065)	(.067)	(.068)	(.068)
	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MA/PhD	2.858***	3.130***	2.572***	2.983***	2.889***	3.259***	3.268***	3.270***	2.580***	2.996***	2.893***	3.266***	3.275***	3.276***
	(.060)	(.063)	(.057)	(.062)	(.061)	(.065)	(.065)	(.065)	(.057)	(.062)	(.061)	(.065)	(.065)	(.065)
Population Stratification														

PC1									-.557	1.364	-.033	1.088	11.018	1.927
									(8.811)	(9.274)	(8.988)	(9.415)	(9.483)	(9.492)
PC2									17.971	19.796	13.806	19.962	19.736	19.774
									(14.062)	(14.914)	(13.944)	(14.795)	(14.913)	(14.915)
PC3									3.408*	16.925	31.754*	2.033	19.603	19.406
									(13.654)	(14.355)	(13.523)	(14.233)	(14.323)	(14.326)
PC4									.258	-17.723	-13.902	-28.536	-31.956	-31.792
									(37.168)	(38.117)	(37.955)	(38.808)	(38.841)	(38.842)
PC5									54.304	52.797	42.490	44.519	45.098	44.575
									(36.515)	(37.475)	(37.259)	(38.131)	(38.151)	(38.156)
PC6									-37.877	-22.288	-35.179	-17.999	-15.990	-15.742
									(34.813)	(35.766)	(35.557)	(36.404)	(36.440)	(36.441)
PC7									3.242	13.653	24.867	1.147	9.837	1.530
									(28.453)	(29.150)	(29.084)	(29.721)	(29.734)	(29.747)
PC8									-14.167	3.337	-14.367	.864	4.026	3.598
									(26.647)	(27.419)	(27.206)	(27.909)	(27.943)	(27.949)
PC9									-1.779	-24.297+	7.401	-12.534	-12.125	-12.674
									(13.814)	(14.333)	(14.194)	(14.701)	(14.719)	(14.732)
PC10									-18.601	-31.532	-8.554	-25.789	-25.423	-25.139
									(26.419)	(27.214)	(26.980)	(27.682)	(27.689)	(27.695)
Constant	2.193***	4.554***	2.113***	2.230***	4.657***	4.715***	4.757***	4.757***	2.119***	2.284***	4.646***	4.740***	4.783***	4.784***
	(.036)	(.143)	(.035)	(.037)	(.142)	(.145)	(.146)	(.147)	(.048)	(.050)	(.146)	(.149)	(.150)	(.151)
-2 Log-Likelihood	19758	19136	20382	19492	19628	18872	18860	18860	20360	19462	19618	18856	18844	18844
Observations	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418	20,418
N	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,252

Standard errors in parentheses.

*** p<.001, ** p<.01, * p<.05, + p<.1

parental education-squared and the polygenic score have no sign of supporting the inverted U-shaped effect of family background on the expression of genes as Pareto hypothesis argues. Therefore, the Saunders hypothesis is supported by the evidence since resourceful parents have the abilities to retain their offspring within the higher levels of education, and therefore, only children from disadvantaged families have to rely on their natural talents to mobile up.

After controlling ancestral differences in SNPs by holding principal components constant, the findings of the main effects of the polygenic score, family background, and the interaction terms still hold from Model 9 to Model 14. The effect size of the polygenic score drops a small .01 in Model 9 compared to Model 4. As for the effects of parental education in Model 10 and 12, the coefficient increases only slightly for parental education after controlling for population stratification. Moreover, in Model 13, the significant negative effect of the interaction between parental education and the polygenic score stays the same, whereas the nonsignificant interaction effect in Model 14 also holds.

The results support my hypothesis 1 and hypothesis 2c. The polygenic scores and family socioeconomic status have significant positive impacts on the transition to the higher stages of education. Moreover, the main effects of parental education and polygenic scores are independent from each other to some extent. Only slight drops or increases are found in the analyses when including other variables into the models. Auxiliary analysis (analysis not shown) using the ordinary least square model also shows the whole-genome polygenic score alone can explain 4.9% of the variation in years of schooling when considering population stratification, while family socioeconomic backgrounds and other control variables accounts for 24%. When the polygenic score is added to the model which has already controlled the socioeconomic status and control variables, the R^2 increases to 26.2%. These results also support the independent main

effects of both genes and social factors on educational attainment.

As for the second hypothesis, the significant interaction terms between parental education and polygenic score suggests that it moderates the genetic influences on education. The negative results further support the Saunders' argument (hypothesis 2c) that individuals who are from advantaged families do not rely on their talent as much as their disadvantaged counterparts. From another perspective, if an individual is talented, parental education does not matter so much for them to earn a higher degree; but, if an individual is not as talented, parent's education is important to their transitions to higher education. This result suggests that children whose parents are highly educated are able to protect them against downward mobility.

Cohort Differences

The results for cohort differences are shown in Table 3. First of all, the effects of polygenic score show a curvilinear relationship over time. The interaction term between the polygenic score and age is positive, but the interaction term between the polygenic score and age-squared is negative in Model 2. This pattern suggests that although the influence of genes is larger for the middle cohort, the impact is smaller in both the younger and the older cohorts. In Model 3, the effect of parental education also shows a positive relationship with age but a negative relationship with age-squared, indicating the effect of parental education also has an inverted U-shaped relationship, with parental education substantially affects the middle cohort, but not so to the younger and older cohorts. These results support the hypothesis 3b (equalization with saturation): the genetic effects increase at first when educational policies begin to liberalize, but decline after higher education becomes universally accessible. And, the effects of socioeconomic backgrounds also increase at first but decrease later since children from resourceful families are

Table 3. Cohort Differences in Continuation Ratio Model Predicting Educational Attainment

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					1905~30 AHEAD & CODA		1931~41 HRS		1948~59 WB, EBB, & MBB	
Standardized Whole-genome Polygenic Score (PGS)	.300***	-1.735*	-1.393+	-1.503*	.210***	.222***	.336***	.347***	.305***	.312***
	(.019)	(.741)	(.743)	(.745)	(.039)	(.039)	(.031)	(.031)	(.032)	(.032)
Standardized Parental Education	.537***	.538***	-3.683***	-3.531***	.402***	.429***	.643***	.647***	.533***	.535***
	(.020)	(.020)	(.788)	(.791)	(.041)	(.042)	(.033)	(.033)	(.033)	(.033)
Parental Education*PGS	-.065***	-.067***	-.062***	.230+		-.111**		-.071*		-.043
	(.018)	(.018)	(.019)	(.135)		(.041)		(.029)		(.031)
Female	-.375***	-.373***	-.373***	-.375***	-.416***	-.423***	-.445***	-.447***	-.294***	-.292***
	(.037)	(.037)	(.037)	(.037)	(.077)	(.077)	(.060)	(.060)	(.062)	(.062)
Age/10	-.273***	-.135	-.212	-.187	-.300***	-.302***	-.249**	-.240*	-.055	-.059
	(.018)	(.220)	(.221)	(.221)	(.085)	(.085)	(.094)	(.094)	(.074)	(.074)
(Age/10)2		-.010	-.005	-.006						
		(.016)	(.016)	(.016)						
PGS*(Age/10)		.590**	.495*	1.191***						
		(.216)	(.217)	(.233)						
PGS*(Age/10)2		-.042**	-.035*	.524*						
		(.016)	(.016)	(.217)						
Parental Education*(Age/10)			1.243***	-.037*						
			(.232)	(.016)						
Parental Education*(Age/10)2			-.090***	-.085***						
			(.017)	(.017)						
Parental Education*PGS*(Age/10)				-.043*						
				(.020)						
Region (0=Northeast)										
Midwest	-.109*	-.111*	-.114*	-.116*	-.252*	-.255*	-.023	-.031	-.109	-.106
	(.051)	(.051)	(.051)	(.051)	(.106)	(.107)	(.083)	(.083)	(.083)	(.083)
South	-.105+	-.101+	-.093+	-.094+	.058	.063	-.093	-.091	-.227*	-.222*
	(.055)	(.055)	(.055)	(.055)	(.113)	(.113)	(.088)	(.088)	(.092)	(.092)
West	.051	.057	.061	.059	.147	.150	.049	.049	-.099	-.096
	(.066)	(.066)	(.066)	(.066)	(.127)	(.127)	(.115)	(.115)	(.110)	(.110)
Other	-.040	-.052	-.062	-.066	.210	.164	-.069	-.071	.036	.048
	(.125)	(.125)	(.126)	(.126)	(.449)	(.448)	(.152)	(.152)	(.303)	(.303)
Missing	-.091	-.087	-.092	-.093	-.487*	-.492*	-.181	-.189	-.090	-.091
	(.102)	(.102)	(.102)	(.102)	(.209)	(.209)	(.528)	(.531)	(.129)	(.129)
Rural (0=Urban)										
Rural	-.406***	-.408***	-.410***	-.408***	-.574***	-.569***	-.464***	-.460***	-.233***	-.235***
	(.038)	(.038)	(.038)	(.038)	(.081)	(.081)	(.062)	(.062)	(.062)	(.062)
Missing	-1.615***	-1.624***	-1.669***	-1.665***	-3.455***	-3.457***	-.356+	-.341+	-.911**	-.896**
	(.120)	(.120)	(.120)	(.120)	(.235)	(.236)	(.184)	(.185)	(.312)	(.312)
Stage (0=GED/HS vs. No Degree)										
2-yr/Some College	-3.104***	-3.109***	-3.120***	-3.121***	-2.965***	-2.981***	-3.169***	-3.187***	-3.508***	-3.525***
	(.049)	(.049)	(.049)	(.049)	(.092)	(.093)	(.077)	(.078)	(.098)	(.100)
4-yr College	-1.053***	-1.058***	-1.072***	-1.074***	-.379*	-.396**	-.833***	-.853***	-1.953***	-1.970***
	(.068)	(.068)	(.068)	(.068)	(.150)	(.150)	(.114)	(.115)	(.117)	(.118)

MA/PhD	-3.275***	-3.280***	-3.295***	-3.296***	-3.158***	-3.171***	-3.169***	-3.181***	-3.870***	-3.882***
	(.065)	(.065)	(.065)	(.065)	(.133)	(.133)	(.105)	(.106)	(.118)	(.119)
Population Stratification										
PC1	11.018	11.483	11.801	11.784	-68.940	-61.722	9.178	11.652	44.144**	44.111**
	(9.483)	(9.492)	(9.502)	(9.515)	(51.216)	(51.340)	(17.980)	(18.037)	(16.893)	(16.943)
PC2	19.736	19.951	2.913	2.505	38.438	33.738	14.313	15.255	67.654+	67.624+
	(14.913)	(14.943)	(14.871)	(15.014)	(3.940)	(3.937)	(34.857)	(34.896)	(36.310)	(36.342)
PC3	19.603	19.888	2.085	19.678	31.609	29.132	1.929	12.839	51.032	49.874
	(14.323)	(14.344)	(14.285)	(14.395)	(2.522)	(2.576)	(31.718)	(31.740)	(33.840)	(33.885)
PC4	-31.956	-34.809	-33.054	-33.054	-184.085*	-187.335*	68.975	65.480	-26.173	-28.079
	(38.841)	(38.870)	(38.900)	(38.901)	(86.664)	(86.788)	(63.415)	(63.471)	(62.568)	(62.599)
PC5	45.098	45.546	48.603	48.481	-17.686	-2.793	106.312+	108.114+	49.269	49.846
	(38.151)	(38.162)	(38.198)	(38.193)	(78.181)	(78.238)	(62.942)	(62.940)	(63.464)	(63.513)
PC6	-15.990	-19.163	-17.225	-16.352	-174.111*	-169.207*	41.243	43.396	4.901	41.153
	(36.440)	(36.471)	(36.503)	(36.502)	(82.202)	(82.283)	(57.261)	(57.285)	(61.085)	(61.133)
PC7	9.837	1.979	13.701	14.966	-28.726	-26.310	-37.312	-38.065	92.711+	91.478+
	(29.734)	(29.758)	(29.787)	(29.789)	(65.067)	(65.116)	(48.681)	(48.702)	(49.025)	(49.037)
PC8	4.026	3.364	1.619	.393	-116.401+	-114.382+	-46.788	-41.725	101.958*	103.440*
	(27.943)	(27.957)	(27.980)	(27.978)	(63.877)	(63.848)	(47.480)	(47.544)	(44.473)	(44.525)
PC9	-12.125	-13.116	-14.020	-13.407	-62.732*	-59.054+	-1.364	-9.377	34.793	33.876
	(14.719)	(14.742)	(14.733)	(14.777)	(31.068)	(31.035)	(27.771)	(27.750)	(27.526)	(27.515)
PC10	-25.423	-25.128	-21.474	-2.530	-4.365	-1.393	-21.161	-2.006	-4.851	-6.638
	(27.689)	(27.690)	(27.718)	(27.725)	(6.112)	(6.137)	(46.613)	(46.619)	(45.901)	(45.957)
Constant	4.783***	4.316***	4.602***	4.504***	4.891***	4.930***	4.521***	4.479***	4.045***	4.083***
	(.150)	(.759)	(.760)	(.762)	(.709)	(.710)	(.666)	(.666)	(.451)	(.452)
-2 Log-Likelihood	18844	18836	18808	18804	4490	4484	7162	7156	6830	6828
Observations	20,418	20,418	20,418	20,418	4,828	4,828	7,856	7,856	7,734	7,734
N	8,251	8,251	8,251	8,251	2,093	2,093	3,272	3,272	2,886	2,886

Standard errors in parentheses.
*** p<.001, ** p<.01, * p<.05, + p<.1

more likely to seize the opportunities first, and the impacts decrease when opportunities become widely accessible.

Secondly, the three-way interaction term between parental education, the polygenic score, and age is included in Model 4 to assess the moderating effect changes of parental education. The significant negative effect shows that the negative moderating effect of parental education on the expression of genes is stronger in the older cohort. In addition, the cohort separated analyses as presented in Model 5 to Model 10 also demonstrate the moderating impact of parental education on the polygenic score are greater for the oldest cohort in Model 6, less strong for the middle cohort in Model 8, and weaker for the youngest cohort in Model 10. The reducing effects of parental education on the realization of genetic potential further strengthens the equalization process in the U.S. between the 1920s and 1970s.

As for the other control variables, the negative effect of being female also shows a similar curvilinear pattern, indicating males benefit more at the beginning of the expansion of higher education, but the benefit for males diminishes later on when women in the younger cohorts enjoy more opportunities to earn higher degrees compared to their older counterparts. This depicts a declining gender gap in education. Besides the findings on gender, the negative effects of living in the Midwest and in the South are significant for the oldest cohort and the younger cohort respectively, which might imply the unequal distribution of educational resources across the country in the different periods.

Predicted Probability

Figure 2 shows the predicted probability for advancing to a four-year college for different levels of parental education and the polygenic score after estimating Model 4 in Table 3. The

standardized parental education is divided into three groups according to its standard deviation away from zero. I use 1.5 standard deviation as a cutoff point, so parental education lower than -1.5 standard deviation is categorized as the lowest group, -1.5 to +1.5 standard deviation is the middle group, and the +1.5 and above is the highest group. In the graph, parental education affects the probability of going to a four-year college the most when an individual's genetic potential for education is lower. However, as the polygenic score increases, the distances between the three lines shrink and converge in the end. This figure clearly illustrates the negative interaction term between parental education and the polygenic score as shown in the final model in Table 2.

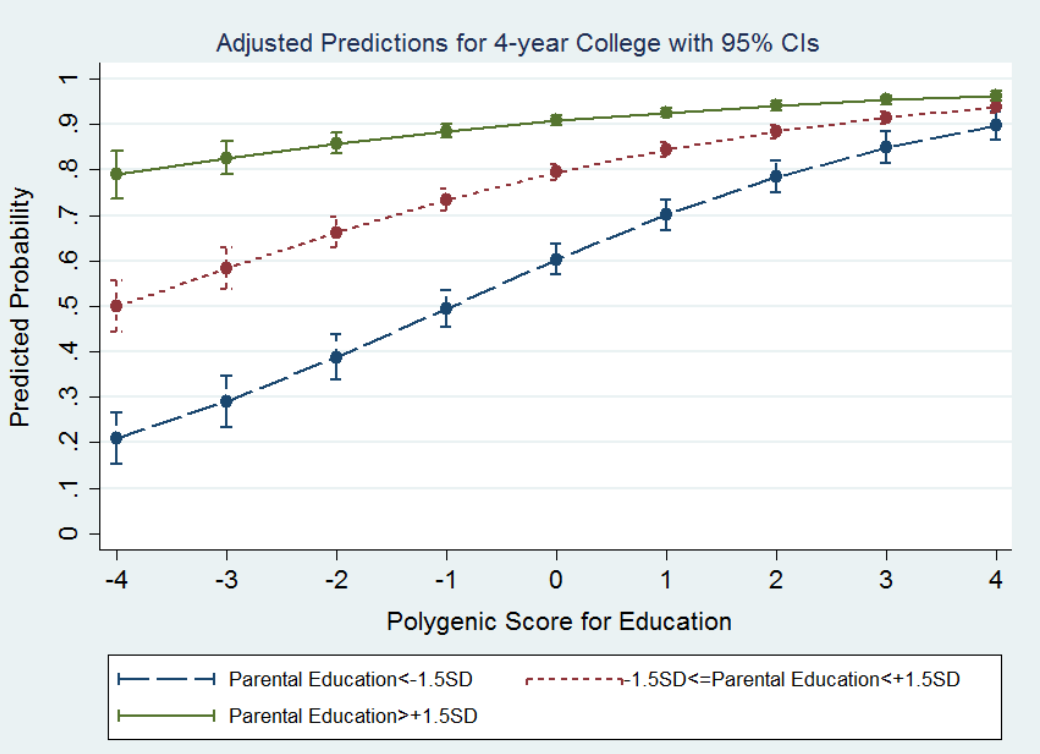

Figure 2. Predicted Probabilities for Advancing into a 4-year College by Holding Covariates at Means. Three levels of parental education are categorized according to their standard deviation away from the mean within the cohort the respondent belongs. These three groups are: parental education 1.5 standard deviation lower than the mean is the lowest group, between -1.5 and +1.5 is the middle group, and 1.5 standard deviation greater than the mean is the highest group. The X-axis is the standardized whole-genome polygenic score, and the Y-axis represents the predicted probability for attending a 4-year college when holding other variables at means. As shown in the figure, although children from the highest parental education group are more likely to attend a 4-year college overall, the differences in the predicted probabilities between the levels of parental education shrink as the polygenic score increases.

The cohort differences can be seen in Figure 3. Each plot in Figure 3 presents the predicted probability to attend a four-year college for the three levels of parental education at different ages. For the fifty years old group, the three lines are close and almost parallel to each other, but in the older age groups, not only the differences between the three levels of parental education at the lower end of the polygenic score widen as age, but the pattern of three lines converging in the high end of polygenic score becomes apparent in the older age groups. Moreover, the figure also shows a trend that the younger the age cohort, the higher the probability of going to a four-year college for individual with every level of parental education. Furthermore, besides visualizing the negative three-way interaction between parental education, the polygenic score, and age, which suggests a reducing moderating impacts of parental education on the expression of genetic potential over time, the figure further supports the equalization with saturation hypothesis by showing that it is because the less talented individuals with lower parental education have better chances to enter higher education in the younger cohorts, rather than the chances for the more talented individuals with lower parental education decline that results in the reducing predictive power of genes for the youngest cohort.

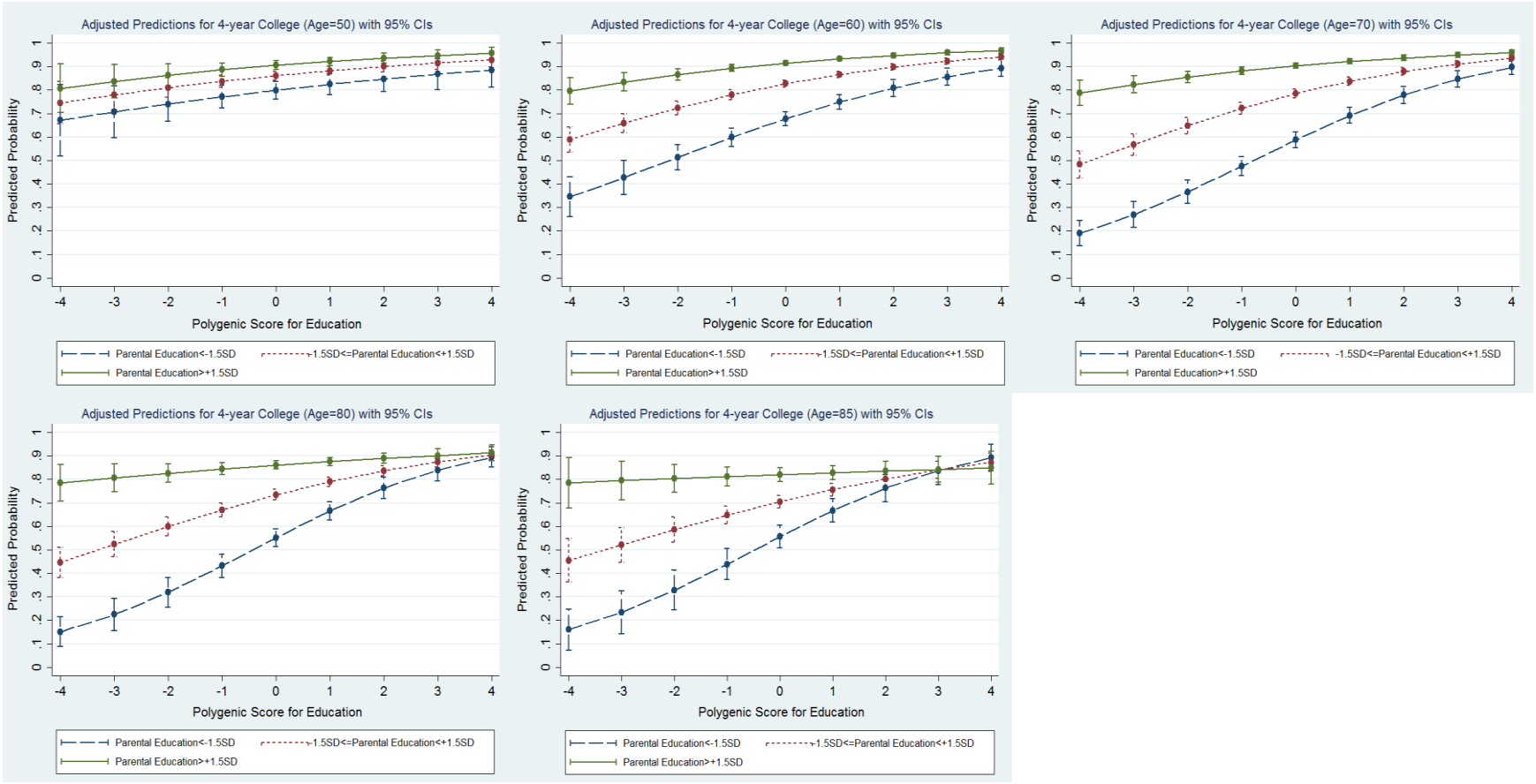

Figure 3. Cohort Differences in the Predicted Probabilities for Advancing into a 4-year College by Holding Covariates at Means. From left to right, top to bottom, these plots show the predicted probabilities for advancing into a 4-year college by holding covariates at means along with the standardized polygenic score for education for age at 50, 60, 70, 80, and 85, respectively. Three levels of parental education are categorized according to their standard deviations away from the mean within the cohort the respondent belongs. These groups are: parental education 1.5 standard deviation below the mean is the lowest group, between -1.5 and +1.5 is the middle group, and 1.5 standard deviation above the mean is the highest group. These plots show that the genetic effects are larger among the older, especially the middle cohorts, where the slopes of the curves are steeper. Also, the predicted probabilities differences between levels of parental education increases slightly for the middle cohorts but the overall distances between lines become smaller. Moreover, the trend of the higher the polygenic score, the smaller the differences between levels of parental education is replaced by three closely paralleled lines in the youngest cohort, meaning that the moderating effects of parental education diminishes at the later stage of the expansion of higher education in the U.S.

CONCLUSION AND DISCUSSION

This study uses the latest available genetic measurement to answer the longstanding questions on educational attainment in sociology. The analyses show that, first, both genetic components and traditional family socioeconomic background are positively related to the educational transitions. The main effects of these two pathways fluctuate slightly when controlling one another, indicating that they have independent effects on educational attainment. The inclusion of the genetic component does not really challenge the effect of socioeconomic background, thus solidifies the foothold of the sociological status attainment model.

Secondly, although the main effects of the genetic component and the socioeconomic background are independent and does not explain away the effects of each other, the significant gene-environment interaction effects between them further enrich the understanding of the mechanisms for educational transition. The negative interplay between the genetic polygenic score and the parental education implies the effectiveness of the resourceful parents to keep their offspring from falling from the advantaged levels of schooling. In other words, untalented individuals are able to earn higher degrees if their parents are well educated. Nevertheless, individuals from the disadvantaged backgrounds are still likely to obtain higher education if they are talented. This pattern therefore supports the Saunders hypothesis but not the Scarr-Rowe nor the Pareto hypothesis. Although past studies on recent birth cohorts suggest the Scarr-Rowe hypothesis which holds the genes express better in the abundant environment might be true (Guo and Stearns 2002; Turkheimer et al. 2003; Tucker-Drob et al. 2010) , for the older U.S. adults, the talented in poor situations could still succeed academically in the old days.

Thirdly, the findings of this study support the equalizing educational opportunity in the U.S. in the middle of the twentieth century and provide evidence for the effects of historical

changes on genetic expression. The effects of the polygenic score increased at the beginning of the implementing of the liberal policies, and declined later after higher education accepted students almost universally. At the meantime, not only the effect of family socioeconomic background declined in the younger cohorts, but the interaction term between the genetic component and the socioeconomic backgrounds also decreased over time, which also further strengthens the equalization with saturation hypothesis. This shows that the Maximally Maintained Inequality (MMI) hypothesis can apply to both the effects of parental education and the effects of genes, with the two show the same patterns over time during the expansion of higher education. Individuals are ascribed with family backgrounds and genes as their status. When opportunities for higher education expand, those with higher status, either family socioeconomic status or genetic predisposition, are able to grab them and increase the inequality at first, but as the given level of higher education saturated since almost everyone possesses higher status attains it, the impacts of ascribed status declines. However, although on the one hand, if genetic potential is considered as a measure of ability, the reducing selectivity of higher education might thus lead to less meritocracy in the latest cohort; on the other hand, the expansion of higher education, as discussed in Raftery and Hout (1993), decreased the inequality in educational opportunities to some extent.

The nonsignificant interaction terms between parental education and polygenic score in the younger cohort but significant in the older cohort imply the comparatively less protective ability of the advantaged families in the younger cohort. These are in line with those found in Branigan et al. (2013) and Heath et al. (1985), where the liberal policies in the latter half of the twentieth century encouraged the expression of genes of individuals born at that time. However, the later trend of the downward effects of the genes and the socioeconomic status over time might be specific to the cohorts analyzed here. Since the HRS respondents were born between 1905 and 1959, most of them entered higher education before 1980, the tuition

of the postsecondary education would not affect their decision to continue education as much as it may now be due to the dramatic rise in the recent decades. As National Center for Education Statistics (2016) estimated, the average 1983-84 tuition in constant 2013-14 dollars was 9,620, the number rose to 12,745 ten years after, and it became 21,003 in 2013-14. The growth was 32.5% in the 1983-94 period, but was 65% in 1993-14, which doubled the number of the earlier period. This huge rise in college tuition might thus lead to the decline of genetic expression because the poorer families cannot afford the college tuition which would result in the worse opportunities for the poor but talented to attain the education they deserve in studies focusing current teenagers and young adults. In addition, the moderating effects of socioeconomic status can turn positive due to the reason that the wealthy families have more advantages to enable their offspring to earn higher degrees regardless of their children's talents, but even the talented in the poor families have few chances to go to higher education now then decades before. It is possible that the evidence of educational attainment for the Scarr-Rowe hypothesis which argues that the rich environment is linked to higher expression of genes was the consequence of the rising tuition in the past decade.

Compared to Conley et al. (2015) and Conley and Domingue (2016), this study not only confirms the effect of genes on educational attainment again, it also provides evidence for the gene-environment interactions in both micro and macro ways. From the micro point of view, unlike the weak and slightly positive interaction between the polygenic score and the maternal education reported by Conley et al. (2015), I find the stronger but negative interaction effects of polygenic score and parental education. The result suggests that the Saunders hypothesis might be true and refutes the alternative hypotheses on gene-environment interaction. As for the macro viewpoint, although the historical changes was evidenced by the significant interaction between birth year and polygenic score in Conley and Domingue's (2016) work, this study discovers the concave curvilinear association between

genetic effects and educational transition over time. Combined with the findings of the similar inverted U-shaped effects of family socioeconomic status and the weakening moderating gene-environment interaction effects, the equalization with saturation hypothesis is further supported. Furthermore, although they found the increasing genetic effects on the graduate school level transition over time, the results from the unconstrained continuation ratio model (Appendix Table 1) show the marginally significant interaction term of the highest transition, age, and the polygenic score which suggest the genetic effects on advancing to graduate schools vary across cohorts weakly, and this can be attributed to the interaction effects of educational stages and age, implying that the cohort differences in the opportunities to higher education might be more important than individual talents alone. Henceforth, by employing the newly reported effect sizes and the better-powered polygenic score, this study advances the understanding of the genetic effects, the socioeconomic status' effects, and the gene-environment interaction on educational attainment.

With respect to the policy implication, first, since both genetics and socioeconomic status matter for educational attainment, resource redistribution policies might be able to improve the achievement of the children from the deprived families. The significant moderating effect of socioeconomic status also implies the welfare policies could boost the academic performance of the untalented from the poor families. Second, the empirical supports for the equalization with saturation hypothesis show the impacts of the liberal policies on educational opportunities in the mid-twentieth century. Although policies similar to these might raise the effects of both genes and family socioeconomic status at first, they ameliorate the influences of gene and family socioeconomic status by increasing the chances for the less talented and the poor to attend higher education later on. It should be insightful for policy making if further studies can investigate the genetic effect changes after the 1980s conservative Reagan revolution. The study by Roksa et al. (2007) is an example focusing on

the changing effects of social inheritance, which has shown the influence of having a college and above educated parent on entering four-year college increased significantly again in the 1990s. Future research can work on the genetic effect changes at the same period.

Several concerns should be mentioned in this study. First, it is essential to discuss the decision between using the polygenic score constructed by the top hits or by the whole-genome SNPs. The results from the top hits polygenic score are usually weaker than the whole-genome polygenic score. Appendix Table 2 to 4 show the results of the 73 top hits polygenic score. The score is not standardized, so the results can be interpreted as the effect of an allele. Although the main effects of polygenic score are significant in all the models, only the interaction term between parental education and the polygenic score term is significant in Appendix Table 3 and are in the same directions as the whole-genome results, but the interaction pattern between parental education, the polygenic score, and age is less significant in Appendix Table 4. Since the findings do not differ much, the results of the more powerful whole-genome polygenic score are reported in the main text.

Secondly, due to the large portion of missing cases on father's occupation and self-rated childhood SES in the younger cohort, I only consider parental education as the measure of childhood socioeconomic status in the analyses. However, there are still 964 cases who have missing values on parental education. Therefore, besides the listwise deletion applied to the main tables reported earlier, the multiple imputation method is also used to check the differences. As reported in Appendix Table 5 to 7, the effects of polygenic scores are still strong among all the models when using the multiple imputation method, and so do the negative interaction term between parental education and polygenic score, but the cohort differences are weaker in Appendix Table 7. But, in general, the patterns are similar to the one without imputing data, and to reduce the likelihood of making up data, I report the former results in the study.

Thirdly, I also estimate the heritability of educational attainment by the genome-relatedness-matrix restricted maximum likelihood (GREML) method. In this analysis, only individuals whose genetic relatedness lower than .025 are included to avoid artificial correlation. Using the genome-wide complex trait analysis (GCTA) software, the heritability of a trait is calculated by estimating how much of the variance in the phenotype is explained by the variance in the genotype. Appendix Table 8 reports the pairwise genome-wide relatedness values on years of schooling. The results of the full sample final model which include all variables but interaction terms in this study show the heritability of 18.3%. This figure is similar to the 20% reported by Conley et al. (2015) and thus confirms the previous results. Additionally, the cohort separated analyses report the heritabilities of 26.7% and 2.3% for the older and the younger cohort respectively. Though the estimate for the younger cohort is not significant, the results suggest the genetic components become less powerful in explaining the variation in educational attainment. These results thus agree with the findings of this paper since the genetic effects dropped in the younger cohort.

There are also several limitations of the paper. First, the HRS dataset does not include the parental genetics information and therefore a part of the puzzle remains unsolved in this study. Parent's genes might directly or indirectly influence their own status attainment and henceforth affect offspring's genetic expression. Although Conley et al. (2015) has attempted to include parent's genetic information by analyzing the Framingham Heart Study and showed that maternal genotype has no effects on offspring's education when offspring's genotype is controlled, further research using larger sample sizes are needed. Second, utilizing the sibling fixed effect model might control other unobserved variables better. However, sibling information is not available in the HRS data, so other datasets should be used to reconfirm the results in this study. Third, the family socioeconomic backgrounds in childhood, in this case, the parental education, was retrospectively surveyed. Since the

respondents are on average 70 years old, the chances of forgotten or subjectively modified might be relatively high and can thus distort the results. Fourthly, the samples analyzed in this study are restricted to non-Hispanic white due to the availability of the GWAS results. Future research can study the other racial groups when the information required is available. Finally, the betas used in the analyses were identified by the meta-analysis of several cross-national cohorts using the genetics data at one shot from each participant. However, it could be possible that the effects of each SNP express differently over time, and the SNPs that matter significantly at one time might not be significant at other periods. Therefore, cautions are needed to generalize the perhaps cohort-specific results in this paper. Studies can try to identify the specific SNPs for different historical periods when data and method are available.

In spite of these limitations, through the inclusion of genetic polygenic score into the analysis, this study strengthens the sociological explanation of educational attainment. As genetic data become widely available nowadays, together with the rapid development in molecular genetics, future studies can take advantage of the burgeoning opportunities to not only explore the genetic effects on sociological outcomes, but further solidify the sociological accounts of social behaviors and status-related outcomes.

APPENDIX A: UNCONSTRAINED CONTINUATION RATIO MODEL

Appendix Table 1. Unconstrained Continuation Ratio Model Predicting Educational Attainment

VARIABLES	Model 1	Model 2	Model 3	Model 4
Standardized Whole-genome Polygenic Score	.300***	-1.393+	.366***	-3.337+
	(.019)	(.743)	(.040)	(1.837)
Standardized Parental Education	.537***	-3.683***	.731***	-.235
	(.020)	(.788)	(.041)	(1.881)
Parental Education*PGS	-.065***	-.062***	-.005	.023
	(.018)	(.019)	(.020)	(.043)
Female	-.375***	-.373***	-.367***	-.356***
	(.037)	(.037)	(.037)	(.037)
Age	-.273***	-.212	-.276***	-2.338***
	(.018)	(.221)	(.019)	(.590)
Age2		-.005		.118**
		(.016)		(.040)
Stage (0=GED/HS vs. No Degree)				
2-yr/Some College	-3.104***	-3.120***	-3.224***	-8.970***
	(.049)	(.049)	(.053)	(2.410)
4-yr College	-1.053***	-1.072***	-1.130***	-19.554***
	(.068)	(.068)	(.071)	(3.090)
MA/PhD	-3.275***	-3.295***	-3.047***	-18.159***
	(.065)	(.065)	(.070)	(3.171)
Age*PGS		.495*		1.085*
		(.217)		(.508)
Age*Parental Education		1.243***		.415
		(.232)		(.539)
Age2*PGS		-.035*		-.077*
		(.016)		(.035)
Age2*Parental Education		-.090***		-.038
		(.017)		(.038)
Stage*PGS (0=GED/HS vs. No Degree)				
2-yr/Some College*PGS			-.067	1.829
			(.049)	(2.121)
4-yr College*PGS			.024	3.573
			(.070)	(2.914)
MA/PhD*PGS			-.283***	4.572+
			(.063)	(2.756)
Stage*Parental Education (0=GED/HS vs. No Degree)				
2-yr/Some College*Parental Education			-.090+	-4.233+
			(.051)	(2.243)
4-yr College*Parental Education			-.340***	3.324
			(.070)	(3.258)
MA/PhD*Parental Education			-.661***	-2.041
			(.061)	(2.820)
Stage*Parental Education*PGS (0=GED/HS vs. No Degree)				
2-yr/Some College*Parental Education*PGS				-.038
				(.053)
4-yr College*Parental Education*PGS				-.005
				(.074)
MA/PhD*Parental Education*PGS				-.004
				(.064)
Stage*Age (0=GED/HS vs. No Degree)				
2-yr/Some College*Age				1.190+
				(.670)
4-yr College*Age				4.534***
				(.886)
MA/PhD*Age				3.818***
				(.906)
Stage*Age*PGS (0=GED/HS vs. No Degree)				
2-yr/Some College*Age*PGS				-.513
				(.595)
4-yr College*Age*PGS				-.970
				(.845)
MA/PhD*Age*PGS				-1.438+
				(.793)
Stage*Age*Parental Education (0=GED/HS vs. No Degree)				
2-yr/Some College*Age*Parental Education				1.146+
				(.648)
4-yr College*Age*Parental Education				-1.203
				(.972)
MA/PhD*Age*Parental Education				.356
				(.822)
Stage*Age2 (0=GED/HS vs. No Degree)				

2-yr/Some College*Age2				-.054
				(.046)
4-yr College*Age2				-.265***
				(.063)
MA/PhD*Age2				-.234***
				(.064)
Stage*Age2*PGS (0=GED/HS vs. No Degree)				
2-yr/Some College*Age2*PGS				.033
				(.041)
4-yr College*Age2*PGS				.063
				(.060)
MA/PhD*Age2*PGS				.103+
				(.056)
Stage*Age2*Parental Education (0=GED/HS vs. No Degree)				
2-yr/Some College*Age2*Parental Education				-.079+
				(.046)
4-yr College*Age2*Parental Education				.093
				(.072)
MA/PhD*Age2*Parental Education				-.024
				(.059)
Region (0=Northeast)				
Midwest	-.109*	-.114*	-.109*	-.111*
	(.051)	(.051)	(.051)	(.051)
South	-.105+	-.093+	-.091+	-.084
	(.055)	(.055)	(.055)	(.055)
West	.051	.061	.050	.050
	(.066)	(.066)	(.066)	(.067)
Other	-.040	-.062	-.016	-.054
	(.125)	(.126)	(.125)	(.127)
Missing	-.091	-.092	-.085	-.094
	(.102)	(.102)	(.102)	(.103)
Rural (0=Urban)				
Rural	-.406***	-.410***	-.400***	-.396***
	(.038)	(.038)	(.038)	(.039)
Missing	-1.615***	-1.669***	-1.617***	-1.570***
	(.120)	(.120)	(.121)	(.123)
Population Stratification				
PC1	11.018	11.801	11.825	13.253
	(9.483)	(9.502)	(9.358)	(9.502)
PC2	19.736	2.913	2.648	21.805
	(14.913)	(14.871)	(14.706)	(14.952)
PC3	19.603	2.085	19.360	21.382
	(14.323)	(14.285)	(14.156)	(14.368)
PC4	-31.956	-33.054	-29.505	-26.073
	(38.841)	(38.900)	(38.881)	(39.116)
PC5	45.098	48.603	47.046	49.881
	(38.151)	(38.198)	(38.265)	(38.459)
PC6	-15.990	-17.225	-19.321	-14.908
	(36.440)	(36.503)	(36.563)	(36.819)
PC7	9.837	13.701	1.527	13.950
	(29.734)	(29.787)	(29.806)	(29.944)
PC8	4.026	1.619	2.067	1.769
	(27.943)	(27.980)	(27.899)	(28.166)
PC9	-12.125	-14.020	-13.953	-15.277
	(14.719)	(14.733)	(14.694)	(14.806)
PC10	-25.423	-21.474	-21.233	-15.251
	(27.689)	(27.718)	(27.730)	(27.846)
Constant	4.783***	4.602***	4.888***	13.543***
	(.150)	(.760)	(.153)	(2.159)
-2 Log-likelihood	18844	18808	18676	18442
Observations	20,418	20,418	20,418	20,418
N	8,251	8,251	8,251	8,251

Standard errors in parentheses
*** p<.001, ** p<.01, * p<.05, + p<.1

APPENDIX B: RESULTS FOR 73-SNPS POLYGENIC SCORE

Appendix Table 2. Descriptive Statistics for Samples Available for the 73-SNPs Polygenic Score

	Total		AHEAD & CODA 1905~30		HRS 1931~41		WB, EBB, & MBB 1948~59	
Variables	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Degree								
No Degree	.112	.003	.173	.008	.125	.006	.052	.004
GED/High School	.569	.005	.570	.011	.591	.009	.542	.009
2-yr/some college	.054	.002	.034	.004	.042	.004	.081	.005
4-yr college	.156	.004	.144	.008	.138	.006	.183	.007
MA/PhD	.110	.003	.078	.006	.104	.005	.141	.006
Year of Education	13.475	.049	12.885	.058	13.147	.051	14.271	.117
Polygenic Score (PGS)	75.821	.061	75.910	.121	75.817	.096	75.761	.102
Parental Education in Years	1.963	.035	9.437	.061	1.898	.056	12.138	.052
Standardized Parental Education	.000	.011	.000	.022	.000	.017	.000	.019
Female	.574	.005	.570	.011	.545	.009	.609	.009
Cohort	3.191	.013	1.597	.011	3.000	.000	4.556	.009
Birth year	1937.432	.111	1924.071	.101	1936.332	.055	1948.318	.078
Age in 2006/10	6.857	.011	8.193	.010	6.967	.006	5.768	.008
Region								
Northeast	.211	.004	.209	.009	.210	.007	.214	.008
Midwest	.351	.005	.342	.010	.352	.008	.357	.009
South	.265	.005	.260	.010	.296	.008	.232	.008
West	.112	.003	.143	.008	.093	.005	.111	.006
Other	.024	.002	.006	.002	.046	.004	.011	.002
Missing	.038	.002	.040	.004	.003	.001	.076	.005
Rural								
Urban	.525	.005	.532	.011	.497	.009	.552	.009
Rural	.444	.005	.405	.011	.475	.009	.436	.009
Missing	.031	.002	.063	.005	.027	.003	.012	.002
N	8,266		2,093		3,273		2,900	

Appendix Table 3. Continuation Ratio Model Predicting Educational Attainment with 73-SNPs Polygenic Score

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
73-SNPs Polygenic Score (PGS)			.013***	.014***	.014***	.015***	.016***	.013***	.012***	.012***	.013***	.014***	.015***	.012***
			(.003)	(.003)	(.003)	(.003)	(.003)	(.004)	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)
Standardized Parental Education	.573***	.543***		.573***		.543***	1.256***	1.266***		.579***		.550***	1.220***	1.231***
	(.019)	(.020)		(.019)		(.020)	(.251)	(.251)		(.019)		(.020)	(.252)	(.251)
Parental Education*PGS							-.009**	-.009**					-.009**	-.009**
							(.003)	(.003)					(.003)	(.003)
Parental Education2								-.175						-.176
								(.131)						(.129)
Parental Education2*PGS								.002						.002
								(.002)						(.002)
Female		-.378***			-.388***	-.377***	-.377***	-.379***			-.387***	-.374***	-.374***	-.376***
		(.036)			(.036)	(.036)	(.036)	(.036)			(.036)	(.036)	(.036)	(.037)
Age/10		-.246***			-.242***	-.248***	-.248***	-.250***			-.245***	-.251***	-.252***	-.254***
		(.018)			(.018)	(.018)	(.018)	(.018)			(.018)	(.018)	(.018)	(.018)
Region (0=Northeast)														
Midwest		-.148**			-.168***	-.146**	-.146**	-.146**			-.141**	-.095+	-.096+	-.096+
		(.049)			(.048)	(.049)	(.049)	(.049)			(.049)	(.050)	(.050)	(.050)
South		-.168**			-.280***	-.163**	-.161**	-.162**			-.244***	-.104+	-.103+	-.104+
		(.053)			(.051)	(.053)	(.053)	(.053)			(.053)	(.054)	(.054)	(.054)
West		.036			.117+	.043	.043	.040			.139*	.079	.078	.076
		(.065)			(.063)	(.065)	(.065)	(.065)			(.064)	(.066)	(.066)	(.066)
Other		-.045			-.061	-.045	-.047	-.044			-.023	-.029	-.030	-.027
		(.123)			(.120)	(.123)	(.123)	(.123)			(.121)	(.124)	(.124)	(.124)
Missing		-.112			-.220*	-.108	-.105	-.110			-.197*	-.070	-.068	-.073
		(.101)			(.098)	(.101)	(.101)	(.101)			(.099)	(.101)	(.101)	(.101)
Rural (0=Urban)														
Rural		-.435***			-.529***	-.434***	-.433***	-.434***			-.515***	-.409***	-.409***	-.410***
		(.037)			(.037)	(.037)	(.037)	(.038)			(.037)	(.038)	(.038)	(.038)
Missing		-1.671***			-1.762***	-1.674***	-1.675***	-1.683***			-1.756***	-1.677***	-1.677***	-1.686***
		(.118)			(.116)	(.118)	(.118)	(.118)			(.116)	(.118)	(.118)	(.118)
Stage (0=GED/HS vs. No Degree)														
2-yr/Some College	-2.856***	-3.028***	-2.653***	-2.860***	-2.846***	-3.034***	-3.038***	-3.038***	-2.661***	-2.870***	-2.850***	-3.039***	-3.043***	-3.044***
	(.045)	(.047)	(.043)	(.045)	(.045)	(.048)	(.048)	(.048)	(.043)	(.045)	(.045)	(.048)	(.048)	(.048)
4-yr College	-.758***	-.958***	-.482***	-.765***	-.706***	-.966***	-.971***	-.969***	-.493***	-.778***	-.711***	-.974***	-.978***	-.977***
	(.065)	(.066)	(.063)	(.065)	(.065)	(.067)	(.067)	(.067)	(.063)	(.065)	(.065)	(.067)	(.067)	(.067)
MA/PhD	-2.859***	-3.129***	-2.428***	-2.869***	-2.745***	-3.140***	-3.144***	-3.145***	-2.448***	-2.893***	-2.755***	-3.154***	-3.158***	-3.160***
	(.060)	(.063)	(.056)	(.060)	(.059)	(.063)	(.063)	(.064)	(.056)	(.061)	(.059)	(.064)	(.064)	(.064)
Population Stratification														
PC1									-8.991	3.241	-8.539	2.768	2.708	2.856

									(8.722)	(9.194)	(8.897)	(9.331)	(9.350)	(9.371)
PC2									16.642	18.803	11.368	18.149	17.656	18.433
									(13.891)	(14.744)	(13.726)	(14.554)	(14.635)	(14.686)
PC3									34.035*	19.637	34.577**	22.144	22.036	22.750
									(13.492)	(14.196)	(13.334)	(14.030)	(14.089)	(14.126)
PC4									-8.125	-25.272	-21.491	-35.737	-34.394	-35.051
									(36.914)	(37.908)	(37.693)	(38.571)	(38.565)	(38.586)
PC5									65.860+	63.419+	53.352	54.836	53.648	54.934
									(36.213)	(37.245)	(36.954)	(37.898)	(37.903)	(37.933)
PC6									-47.751	-3.270	-44.059	-24.920	-25.744	-26.496
									(34.495)	(35.494)	(35.227)	(36.128)	(36.134)	(36.177)
PC7									35.547	17.223	3.383	13.874	12.963	13.077
									(28.217)	(28.957)	(28.847)	(29.512)	(29.524)	(29.547)
PC8									.059	15.308	.289	13.284	12.872	11.975
									(26.421)	(27.221)	(26.973)	(27.709)	(27.716)	(27.734)
PC9									-29.007*	-4.165**	-11.005	-29.004*	-28.885*	-29.177*
									(13.584)	(14.112)	(13.951)	(14.456)	(14.471)	(14.495)
PC10									-28.987	-4.961	-18.562	-35.463	-33.816	-33.301
									(26.197)	(27.029)	(26.737)	(27.475)	(27.479)	(27.499)
Constant	2.195***	4.543***	1.055***	1.135***	3.425***	3.437***	3.367***	3.562***	1.134***	1.298***	3.426***	3.521***	3.454***	3.653***
	(.036)	(.143)	(.237)	(.243)	(.275)	(.282)	(.283)	(.312)	(.240)	(.247)	(.277)	(.284)	(.285)	(.313)
-2 Log-Likelihood	19786	19166	20740	19766	19968	19146	19138	19106	20694	19710	19946	19112	19106	19072
Observations	20,450	20,450	20,450	20,450	20,450	20,450	20,450	20,418	20,450	20,450	20,450	20,450	20,450	20,418
N	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251	8,251

Standard errors in parentheses.
*** p<.001, ** p<.01, * p<.05, + p<.1

Appendix Table 4. Continuation Ratio Model Predicting Educational Attainment Using 73-SNPs Polygenic Score

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					1905~30 AHEAD & CODA		1931~41 HRS		1948~59 WB, EBB, & MBB	
73-SNPs Polygenic Score (PGS)	.015***	.068	.042	.040	.018**	.019**	.013*	.014**	.013*	.014*
	(.003)	(.133)	(.133)	(.133)	(.007)	(.007)	(.005)	(.005)	(.005)	(.005)
Standardized Parental Education	1.220***	1.215***	-2.913***	-3.515*	.409***	1.433**	.650***	1.307**	.568***	1.229**
	(.252)	(.252)	(.820)	(1.792)	(.041)	(.511)	(.033)	(.413)	(.033)	(.415)
Parental Education*PGS	-.009**	-.009**	-.009**	-.001		-.013*		-.009		-.009
	(.003)	(.003)	(.003)	(.022)		(.007)		(.005)		(.005)
Female	-.374***	-.373***	-.373***	-.373***	-.442***	-.446***	-.441***	-.443***	-.275***	-.272***
	(.036)	(.037)	(.037)	(.037)	(.076)	(.077)	(.059)	(.059)	(.061)	(.061)
Age/10	-.252***	1.384	.643	.619	-.299***	-.307***	-.223*	-.221*	-.064	-.065
	(.018)	(2.974)	(2.970)	(2.970)	(.085)	(.085)	(.093)	(.093)	(.073)	(.073)
(Age/10)2		-.137	-.079	-.078						
		(.214)	(.214)	(.214)						
PGS*(Age/10)		-.019	-.011	1.311***						
		(.039)	(.039)	(.321)						
PGS*(Age/10)2		.002	.001	-.010						
		(.003)	(.003)	(.039)						
Parental Education*(Age/10)			1.227***	.001						
			(.231)	(.003)						
Parental Education*(Age/10)2			-.089***	-.089***						
			(.017)	(.017)						
Parental Education*PGS*(Age/10)				-.001						
				(.003)						
Region (0=Northeast)										
Midwest	-.096+	-.095+	-.098+	-.098+	-.246*	-.254*	.025	.024	-.120	-.119
	(.050)	(.050)	(.050)	(.050)	(.106)	(.106)	(.082)	(.082)	(.082)	(.082)
South	-.103+	-.103+	-.095+	-.095+	.068	.065	-.089	-.088	-.222*	-.218*
	(.054)	(.054)	(.054)	(.054)	(.112)	(.112)	(.087)	(.087)	(.091)	(.091)
West	.078	.083	.086	.086	.182	.179	.080	.076	-.075	-.070
	(.066)	(.066)	(.066)	(.066)	(.126)	(.126)	(.114)	(.114)	(.109)	(.109)
Other	-.030	-.038	-.048	-.048	.266	.266	-.051	-.055	.085	.090
	(.124)	(.124)	(.125)	(.125)	(.450)	(.448)	(.151)	(.151)	(.302)	(.302)
Missing	-.068	-.063	-.068	-.069	-.477*	-.488*	-.188	-.204	-.084	-.077
	(.101)	(.101)	(.101)	(.101)	(.208)	(.208)	(.524)	(.524)	(.128)	(.128)
Rural (0=Urban)										
Rural	-.409***	-.410***	-.412***	-.411***	-.559***	-.552***	-.468***	-.469***	-.245***	-.245***
	(.038)	(.038)	(.038)	(.038)	(.081)	(.081)	(.062)	(.062)	(.062)	(.062)
Missing	-1.677***	-1.675***	-1.721***	-1.721***	-3.511***	-3.518***	-.411*	-.409*	-.947**	-.936**
	(.118)	(.118)	(.118)	(.118)	(.234)	(.234)	(.182)	(.182)	(.305)	(.305)
Stage (0=GED/HS vs. No Degree)										

2-yr/Some College	-3.043***	-3.044***	-3.056***	-3.056***	-2.938***	-2.944***	-3.089***	-3.094***	-3.468***	-3.475***
	(.048)	(.048)	(.048)	(.048)	(.092)	(.092)	(.075)	(.075)	(.097)	(.098)
4-yr College	-.978***	-.979***	-.992***	-.992***	-.346*	-.352*	-.731***	-.737***	-1.896***	-1.902***
	(.067)	(.067)	(.067)	(.067)	(.149)	(.149)	(.113)	(.113)	(.116)	(.116)
MA/PhD	-3.158***	-3.159***	-3.173***	-3.173***	-3.111***	-3.117***	-3.016***	-3.020***	-3.754***	-3.760***
	(.064)	(.064)	(.064)	(.064)	(.132)	(.132)	(.103)	(.103)	(.116)	(.117)
Population Stratification										
PC1	2.708	2.787	3.189	3.221	-77.942	-7.077	-3.293	-3.285	34.159*	33.832*
	(9.350)	(9.350)	(9.359)	(9.363)	(51.234)	(51.468)	(17.833)	(17.834)	(16.739)	(16.752)
PC2	17.656	18.212	19.142	19.088	43.052	37.969	5.725	6.436	56.708	55.955
	(14.635)	(14.631)	(14.566)	(14.583)	(31.194)	(31.270)	(34.694)	(34.700)	(35.880)	(35.863)
PC3	22.036	22.610	22.795	22.701	35.458+	34.450+	8.277	9.683	46.377	46.086
	(14.089)	(14.088)	(14.035)	(14.050)	(2.562)	(2.559)	(31.550)	(31.568)	(33.593)	(33.574)
PC4	-34.394	-34.306	-33.544	-33.776	-181.750*	-181.268*	56.361	56.788	-39.382	-36.397
	(38.565)	(38.582)	(38.610)	(38.616)	(86.280)	(86.295)	(62.935)	(62.926)	(62.254)	(62.297)
PC5	53.648	53.120	57.044	56.954	-9.125	-9.959	117.413+	114.111+	59.247	6.224
	(37.903)	(37.925)	(37.962)	(37.962)	(78.010)	(78.046)	(62.453)	(62.470)	(63.003)	(63.054)
PC6	-25.744	-26.676	-25.003	-24.886	-182.728*	-182.476*	35.969	34.692	23.999	22.776
	(36.134)	(36.157)	(36.188)	(36.190)	(82.063)	(82.131)	(56.697)	(56.709)	(6.494)	(6.541)
PC7	12.963	12.267	15.332	15.285	-39.181	-42.730	-33.845	-32.899	103.514*	102.525*
	(29.524)	(29.534)	(29.566)	(29.566)	(65.033)	(65.103)	(48.193)	(48.199)	(48.648)	(48.683)
PC8	12.872	12.572	11.162	11.233	-111.484+	-109.114+	-21.911	-23.696	108.407*	108.892*
	(27.716)	(27.717)	(27.733)	(27.733)	(63.828)	(63.830)	(46.925)	(46.948)	(44.145)	(44.187)
PC9	-28.885*	-28.421*	-29.302*	-29.311*	-78.158*	-75.595*	-29.980	-29.874	17.216	17.381
	(14.471)	(14.485)	(14.478)	(14.482)	(3.978)	(3.961)	(27.350)	(27.336)	(27.172)	(27.177)
PC10	-33.816	-33.986	-3.185	-3.030	-1.949	2.142	-33.837	-31.994	-22.202	-21.125
	(27.479)	(27.482)	(27.510)	(27.514)	(59.988)	(6.038)	(46.130)	(46.145)	(45.547)	(45.566)
Constant	3.454***	-1.174	1.138	1.246	3.551***	3.501***	3.204***	3.117***	3.023***	2.949***
	(.285)	(1.181)	(1.164)	(1.167)	(.851)	(.851)	(.773)	(.775)	(.600)	(.601)
-2 Log-Likelihood	19106	19104	19076	19076	4514	4510	7280	7278	6946	6942
Observations	20,450	20,450	20,450	20,450	4,828	4,828	7,858	7,858	7,764	7,764
N	8,266	8,266	8,266	8,266	2,873	2,873	4,711	4,711	3,820	3,820

Standard errors in parentheses.
*** p<.001, ** p<.01, * p<.05, + p<.1

APPENDIX C: MULTIPLE IMPUTATION RESULTS

Appendix Table 5. Descriptive Statistics for Full Sample and Different Cohorts Using Multiple Imputation

	Total		AHEAD & CODA 1905~30		HRS 1931~41		WB, EBB, & MBB 1948~59	
Variables	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Degree								
No Degree	.131	.004	.196	.008	.143	.006	.069	.004
GED/High School	.569	.005	.565	.010	.590	.008	.549	.009
2-yr/some college	.052	.002	.032	.004	.041	.003	.079	.005
4-yr college	.144	.004	.130	.007	.129	.006	.171	.007
MA/PhD	.103	.003	.075	.005	.097	.005	.132	.006
Standardized Polygenic Score	.000	.010	.122	.020	-.045	.017	-.041	.018
Standardized Parental Education	-.012	.011	-.011	.022	-.012	.017	-.012	.019
Female	.578	.005	.577	.010	.548	.008	.612	.009
Cohort	3.183	.013	1.620	.010	3.000	.000	4.557	.009
Age in 2006/10	6.866	.011	8.177	.009	6.971	.005	5.768	.007
Region								
Northeast	.213	.004	.213	.008	.213	.007	.214	.007
Midwest	.342	.005	.336	.010	.344	.008	.345	.008
South	.271	.005	.261	.009	.302	.008	.242	.008
West	.110	.003	.136	.007	.091	.005	.111	.006
Other	.022	.002	.005	.001	.045	.003	.010	.002
Missing	.041	.002	.049	.004	.004	.001	.078	.005
Rural								
Urban	.521	.005	.531	.010	.494	.008	.545	.009
Rural	.447	.005	.406	.010	.478	.008	.443	.009
Missing	.031	.002	.063	.005	.028	.003	.011	.002
N	9,215		2,414		3,578		3,223	

Appendix Table 6. Continuation Ratio Model Predicting Educational Attainment Using Multiple Imputation

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
Standardized Whole-genome Polygenic Score (PGS)			.332***	.299***	.332***	.306***	.309***	.317***	.322***	.286***	.328***	.297***	.301***	.310***
			(.017)	(.017)	(.017)	(.018)	(.018)	(.022)	(.017)	(.018)	(.017)	(.018)	(.018)	(.022)
Standardized Parental Education	.547***	.516***		.527***		.498***	.492***	.499***		.532***		.503***	.497***	.504***
	(.018)	(.019)		(.019)		(.019)	(.020)	(.022)		(.019)		(.019)	(.020)	(.022)
Parental Education*PGS							-.036*	-.002					-.037*	-.002
							(.018)	(.003)					(.018)	(.003)
Parental Education2								-.012						-.012
								(.011)						(.011)
Parental Education2*PGS								-.009						-.009
								(.012)						(.012)
Female		-.349***			-.356***	-.345***	-.343***	-.346***			-.356***	-.343***	-.341***	-.344***
		(.034)			(.034)	(.035)	(.035)	(.035)			(.034)	(.035)	(.035)	(.035)
Age/10		-.240***			-.259***	-.263***	-.265***	-.263***			-.260***	-.265***	-.267***	-.265***
		(.017)			(.017)	(.017)	(.017)	(.017)			(.017)	(.017)	(.017)	(.017)
Region (0=Northeast)														
Midwest		-.109*			-.121**	-.100*	-.102*	-.103*			-.117*	-.072	-.075	-.075
		(.047)			(.046)	(.047)	(.047)	(.047)			(.047)	(.048)	(.048)	(.048)
South		-.172***			-.247***	-.143**	-.146**	-.145**			-.237***	-.110*	-.113*	-.111*
		(.050)			(.049)	(.050)	(.050)	(.050)			(.050)	(.051)	(.051)	(.051)
West		.070			.129*	.066	.063	.064			.133*	.085	.081	.083
		(.062)			(.061)	(.062)	(.062)	(.062)			(.061)	(.063)	(.063)	(.063)
Other		-.001			-.017	-.001	-.004	.002			.006	.007	.003	.009
		(.119)			(.118)	(.120)	(.120)	(.120)			(.118)	(.121)	(.121)	(.121)
Missing		-.132			-.228*	-.126	-.136	-.133			-.226*	-.108	-.119	-.115
		(.092)			(.090)	(.093)	(.093)	(.093)			(.091)	(.093)	(.093)	(.093)
Rural (0=Urban)														
Rural		-.459***			-.525***	-.445***	-.445***	-.445***			-.521***	-.430***	-.430***	-.430***
		(.036)			(.035)	(.036)	(.036)	(.036)			(.035)	(.036)	(.036)	(.036)
Missing		-1.617***			-1.636***	-1.564***	-1.563***	-1.566***			-1.633***	-1.569***	-1.567***	-1.570***
		(.111)			(.110)	(.112)	(.112)	(.112)			(.110)	(.112)	(.112)	(.112)
Stage (0=GED/HS vs. No Degree)														
2-yr/Some College	-2.722***	-2.886***	-2.606***	-2.784***	-2.792***	-2.950***	-2.954***	-2.953***	-2.610***	-2.788***	-2.794***	-2.953***	-2.956***	-2.956***
	(.041)	(.043)	(.040)	(.042)	(.042)	(.044)	(.044)	(.044)	(.040)	(.042)	(.042)	(.044)	(.044)	(.045)
4-yr College	-.588***	-.781***	-.419***	-.668***	-.639***	-.866***	-.869***	-.868***	-.423***	-.674***	-.641***	-.869***	-.872***	-.871***
	(.061)	(.063)	(.060)	(.062)	(.061)	(.063)	(.064)	(.064)	(.060)	(.062)	(.061)	(.063)	(.064)	(.064)
MA/PhD	-2.633***	-2.892***	-2.384***	-2.764***	-2.689***	-3.028***	-3.023***	-3.026***	-2.392***	-2.775***	-2.693***	-3.034***	-3.029***	-3.032***
	(.057)	(.059)	(.054)	(.058)	(.057)	(.061)	(.061)	(.062)	(.054)	(.058)	(.057)	(.061)	(.061)	(.062)
Population Stratification														
PC1									-7.853	1.510	-7.589	.873	1.593	1.487
									(7.856)	(8.162)	(7.966)	(8.249)	(8.430)	(8.301)

PC2									11.590	12.810	8.933	12.854	13.101	12.534
									(12.188)	(12.530)	(11.928)	(12.402)	(12.519)	(12.474)
PC3									22.138+	9.535	23.948*	12.003	12.445	11.395
									(11.962)	(12.246)	(11.721)	(12.148)	(12.241)	(12.227)
PC4									15.800	1.122	4.541	-7.534	-9.830	-7.616
									(35.058)	(35.981)	(35.780)	(36.605)	(36.783)	(36.601)
PC5									36.475	35.230	21.139	23.655	22.748	22.824
									(34.436)	(35.309)	(35.065)	(35.896)	(36.117)	(35.941)
PC6									-36.517	-21.225	-31.778	-15.786	-16.795	-17.273
									(32.798)	(33.701)	(33.415)	(34.243)	(34.274)	(34.273)
PC7									34.809	19.821	32.459	19.379	2.083	2.287
									(26.827)	(27.535)	(27.389)	(28.030)	(28.001)	(27.989)
PC8									-.723	14.094	-.338	12.726	14.218	13.221
									(24.632)	(25.449)	(25.083)	(25.822)	(25.999)	(25.797)
PC9									-6.560	-19.083	11.561	-6.999	-6.172	-7.308
									(12.889)	(13.318)	(13.251)	(13.647)	(13.656)	(13.658)
PC10									-33.895	-46.684+	-28.406	-44.033+	-42.743+	-41.545
									(24.317)	(25.059)	(24.822)	(25.480)	(25.703)	(25.620)
Constant	2.001***	4.277***	1.931***	2.036***	4.403***	4.454***	4.470***	4.468***	1.916***	2.063***	4.377***	4.457***	4.475***	4.473***
	(.032)	(.135)	(.031)	(.033)	(.133)	(.137)	(.137)	(.137)	(.043)	(.045)	(.136)	(.140)	(.141)	(.141)
-2 Log-Likelihood	22120	21444	22720	21812	21903	21137	21151	21139	22698	21782	21892	21122	21136	21125
Observations	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264	22,264
N	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215

Standard errors in parentheses

*** p<.001, ** p<.01, * p<.05, + p<.1

Appendix Table 7. Cohort Differences in Continuation Ratio Model Predicting Educational Attainment Using Multiple Imputation

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					1905~30 AHEAD & CODA		1931~41 HRS		1948~59 WB, EBB, & MBB	
Standardized Whole-genome Polygenic Score (PGS)	.301***	-1.769*	-1.455*	-1.571*	.229***	.235***	.340***	.345***	.304***	.306***
	(.018)	(.702)	(.703)	(.705)	(.036)	(.037)	(.029)	(.030)	(.030)	(.030)
Standardized Parental Education	.497***	.498***	-2.106*	-1.846*	.384***	.400***	.598***	.599***	.494***	.494***
	(.020)	(.019)	(.813)	(.780)	(.041)	(.043)	(.032)	(.032)	(.034)	(.034)
Parental Education*PGS	-.037*	-.036*	-.026	.230+		-.075+		-.040		-.014
	(.018)	(.018)	(.019)	(.134)		(.040)		(.029)		(.030)
Female	-.341***	-.341***	-.341***	-.343***	-.367***	-.371***	-.415***	-.416***	-.248***	-.247***
	(.035)	(.035)	(.035)	(.035)	(.071)	(.071)	(.057)	(.057)	(.058)	(.058)
Age/10	-.267***	-.198	-.200	-.182	-.208**	-.211**	-.269**	-.264**	-.073	-.074
	(.017)	(.209)	(.208)	(.209)	(.079)	(.079)	(.090)	(.090)	(.070)	(.070)
(Age/10)2		-.005	-.005	-.006						
		(.015)	(.015)	(.015)						
PGS*(Age/10)		.598**	.512*	.545**						
		(.205)	(.205)	(.206)						
PGS*(Age/10)2		-.042**	-.036*	-.039**						
		(.015)	(.015)	(.015)						
Parental Education*(Age/10)			.744**	.662**						
			(.239)	(.230)						
Parental Education*(Age/10)2			-.053**	-.047**						
			(.017)	(.017)						
Parental Education*PGS*(Age/10)				-.038+						
				(.020)						
Region (0=Northeast)										
Midwest	-.075	-.077	-.086+	-.085+	-.190+	-.193+	.025	.021	-.086	-.085
	(.048)	(.048)	(.048)	(.048)	(.099)	(.099)	(.079)	(.079)	(.079)	(.079)
South	-.113*	-.109*	-.125*	-.122*	.088	.089	-.094	-.094	-.235**	-.233**
	(.051)	(.051)	(.051)	(.051)	(.105)	(.105)	(.083)	(.083)	(.087)	(.087)
West	.081	.091	.090	.092	.166	.166	.127	.127	-.087	-.086
	(.063)	(.063)	(.063)	(.063)	(.119)	(.119)	(.110)	(.110)	(.104)	(.104)
Other	.003	-.008	-.015	-.016	.320	.287	-.037	-.038	.129	.133
	(.121)	(.121)	(.122)	(.121)	(.443)	(.442)	(.146)	(.146)	(.299)	(.299)
Missing	-.119	-.117	-.129	-.126	-.377*	-.383*	-.487	-.492	-.104	-.104
	(.093)	(.093)	(.093)	(.093)	(.178)	(.178)	(.471)	(.473)	(.121)	(.121)
Rural (0=Urban)										
Rural	-.430***	-.432***	-.444***	-.442***	-.587***	-.584***	-.469***	-.466***	-.286***	-.287***
	(.036)	(.036)	(.036)	(.036)	(.075)	(.076)	(.059)	(.059)	(.059)	(.059)
Missing	-1.567***	-1.575***	-1.580***	-1.576***	-3.270***	-3.273***	-.365*	-.358*	-.852**	-.847**
	(.112)	(.113)	(.113)	(.113)	(.218)	(.218)	(.176)	(.176)	(.297)	(.297)
Stage (0=GED/HS vs. No Degree)										
2-yr/Some College	-2.956***	-2.960***	-2.949***	-2.947***	-2.815***	-2.824***	-3.040***	-3.047***	-3.273***	-3.278***
	(.044)	(.044)	(.045)	(.044)	(.084)	(.084)	(.071)	(.072)	(.085)	(.086)

4-yr College	-.872***	-.877***	-.863***	-.860***	-.181	-.190	-.678***	-.688***	-1.682***	-1.686***
	(.064)	(.063)	(.064)	(.063)	(.141)	(.141)	(.109)	(.110)	(.105)	(.106)
MA/PhD	-3.029***	-3.033***	-3.006***	-3.003***	-2.854***	-2.859***	-2.958***	-2.961***	-3.535***	-3.538***
	(.061)	(.061)	(.061)	(.061)	(.124)	(.124)	(.101)	(.101)	(.107)	(.108)
Population Stratification										
PC1	1.593	1.503	-.328	-.266	-94.067*	-89.777+	-4.440	-3.107	38.138*	38.249*
	(8.430)	(8.407)	(8.457)	(8.363)	(47.194)	(47.292)	(14.808)	(15.001)	(15.563)	(15.603)
PC2	13.101	13.771	14.456	14.177	55.437+	52.440+	-.388	.476	68.380+	68.362+
	(12.519)	(12.521)	(12.687)	(12.602)	(29.191)	(29.214)	(23.092)	(23.254)	(35.303)	(35.318)
PC3	12.445	12.847	14.123	14.301	23.902	22.236	-5.638	-4.519	59.797+	59.587+
	(12.241)	(12.148)	(12.333)	(12.293)	(2.031)	(2.071)	(21.258)	(21.424)	(32.787)	(32.807)
PC4	-9.830	-11.465	-8.162	-6.220	-87.722	-89.457	55.328	53.588	-15.197	-15.928
	(36.783)	(36.645)	(36.767)	(36.645)	(8.350)	(8.419)	(6.354)	(6.425)	(59.252)	(59.293)
PC5	22.748	23.285	22.743	21.188	2.878	1.620	6.510	61.334	-3.561	-3.158
	(36.117)	(36.089)	(35.937)	(36.067)	(73.060)	(73.129)	(6.101)	(6.114)	(59.920)	(59.983)
PC6	-16.795	-18.206	-19.036	-17.188	-109.016	-106.017	17.932	19.038	52.678	52.628
	(34.274)	(34.306)	(34.239)	(34.271)	(75.461)	(75.536)	(54.698)	(54.756)	(57.294)	(57.313)
PC7	2.083	2.552	21.860	22.896	7.858	9.375	-29.045	-29.336	78.579+	78.567+
	(28.001)	(28.093)	(28.095)	(28.067)	(59.924)	(59.963)	(46.030)	(46.058)	(46.123)	(46.134)
PC8	14.218	12.167	12.712	12.210	-97.882+	-97.439+	-31.233	-28.638	101.495*	101.692*
	(25.999)	(26.043)	(25.901)	(25.901)	(58.465)	(58.486)	(45.373)	(45.453)	(41.122)	(41.180)
PC9	-6.172	-7.781	-6.446	-6.173	-61.211*	-58.886*	-8.227	-7.595	31.458	31.250
	(13.656)	(13.664)	(13.726)	(13.719)	(28.875)	(28.874)	(26.164)	(26.161)	(26.205)	(26.205)
PC10	-42.743+	-43.067+	-44.951+	-43.184+	-31.688	-29.650	-37.487	-36.857	-26.111	-26.357
	(25.703)	(25.503)	(25.964)	(25.647)	(54.834)	(54.835)	(43.511)	(43.600)	(42.360)	(42.398)
Constant	4.475***	4.246***	4.254***	4.175***	3.843***	3.879***	4.399***	4.377***	3.801***	3.813***
	(.141)	(.719)	(.716)	(.719)	(.658)	(.659)	(.638)	(.638)	(.427)	(.427)
-2 Log-Likelihood	21136	21125	21191	21196	5211	5207	7870	7868	7689	7689
Observations	22,264	22,264	22,264	22,264	5,426	5,426	8,409	8,409	8,429	8,429
N	9,215	9,215	9,215	9,215	2,414	2,414	3,578	3,578	3,223	3,223

Standard errors in parentheses.
*** p<.001, ** p<.01, * p<.05, + p<.1

APPENDIX D: GREML RESULTS

Appendix Table 8. Estimation of the Heritability of Years of Schooling Using Genomic-Relatedness-Matrix Restricted Maximum Likelihood (GREML), by Cohort

	Full			Older Cohort (1905~41 AHEAD, CODA, and HRS)			Younger Cohort (1948~59 WB, EBB, & MBB)		
	Unconditional	No SES	Final	Unconditional	No SES	Final	Unconditional	No SES	Final
V(G): Variance in genotype	2.583	1.733	0.904	3.673	2.291	1.437	1.141	0.748	0.096
S.E.	0.502	0.450	0.382	0.803	0.726	0.626	1.091	1.016	0.832
V(e): Residual error	4.284	4.223	4.047	3.834	4.144	3.948	4.324	4.306	4.050
S.E.	0.348	0.315	0.272	0.552	0.505	0.439	0.766	0.715	0.591
V(P): Variance in phenotype	6.867	5.956	4.951	7.507	6.435	5.385	5.466	5.054	4.146
S.E.	0.195	0.172	0.143	0.296	0.262	0.223	0.366	0.340	0.277
V(G)/V(P)	0.376	0.291	0.183	0.489	0.356	0.267	0.209	0.148	0.023
S.E.	0.064	0.069	0.073	0.090	0.100	0.107	0.187	0.192	0.199
logL	-10487.491	-10043.394	-9493.575	-6884.039	-6598.884	-6247.707	-3499.141	-3387.064	-3180.298
logL0	-10503.525	-10051.664	-9496.652	-6896.871	-6604.203	-6250.429	-3499.717	-3387.362	-3180.305
Likelihood Ratio Test	32.068	16.539	6.154	25.663	10.639	5.444	1.153	0.597	0.015
p-value (df=1)	0.000	0.000	0.007	0.000	0.001	0.010	0.142	0.220	0.452
N	7482	7482	7482	4826	4826	4826	2656	2656	2656

REFERENCES

Agresti, Alan. 2006. *An Introduction to Categorical Data Analysis, Second Edition*.

Allison, Paul D. 2012. *Logistic Regression Using SAS: Theory and Application*. SAS Institute, 2012.

Behrman, Jere R., and Paul Taubman. 1989. "Is Schooling "Mostly in the Genes"? Nature-Nurture Decomposition Using Data on Relatives." *The Journal of Political Economy* 97(6): 1425-1446.

Belsky, Daniel W., and Salomon Israel. 2014. "Integrating Genetics and Social Science: Genetic Risk Scores." *Biodemography and Social Biology* 60(2): 137-155.

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

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* 92(1): 109-140.

Bronfenbrenner, Urie, and Stephen J. Ceci. 1994. "Nature-Nurture Reconceptualized in Developmental Perspective: A Bioecological Model." *Psychological Review* 101(4): 568.

Campbell, Catarina D., Elizabeth L Ogburn, Kathryn L Lunetta, Helen N Lyon, Matthew L Freedman, Leif C Groop, David Altshuler, Kristin G Ardlie, and Joel N Hirschhorn. 2005. "Demonstrating Stratification in a European American Population." *Nature Genetics* 37(8): 868-872.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *The American Economic Review* 104(5): 141-147.

Conley, Dalton, and Benjamin Domingue. 2016. "The Bell Curve Revisited: Testing Controversial Hypotheses with Molecular Genetic Data." *Sociological Science* 3: 520-539.

Conley, Dalton, Benjamin W. Domingue, David Cesarini, Christopher Dawes, Cornelius A. Rietveld, and Jason D. Boardman. 2015. "Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?" *Sociological Science* 2: 82-105.

Davies, Neil M.,Gibran Hemani, Nic J. Timpson, Frank Windmeijer, and George Davey Smith. 2015. "The Role of Common Genetic Variation in Educational Attainment and Income: Evidence from the National Child Development Study." *Scientific Reports* 5:16509.

Domingue, Benjamin W., Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason 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.

Duncan, Otis Dudley. 1968. “Ability and Achievement” *Eugenics Quarterly* 15(1):1-11.

Eckland, Bruce K. 1967. “Genetics and Sociology: A Reconsideration.” *American Sociological Review* 32(2): 173-194.

Freese, Jeremy. 2008. “Genetics and the Social Science Explanation of Individual Outcomes.” *American Journal of Sociology* 114(S1): S1-S35.

Ganzeboom, Harry BG, Donald J. Treiman, and Wout C. Ultee. 1991. “Comparative Intergenerational Stratification Research: Three Generations and Beyond.” *Annual Review of Sociology*: 277-302.

Guo, Guang, and Elizabeth Stearns. 2002. “The Social Influences on the Realization of Genetic Potential for Intellectual Development.” *Social Forces* 80(3): 881-910.

Heath, Andrew 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(6013): 734-736.

Hosmer, D. W., Jr., S. Lemeshow, and R. X. Sturdivant. 2013. “Applied Logistic Regression.” 3rd ed. Hoboken, NJ: Wiley.

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

Jonsson, Jan O., David B. Grusky, Matthew Di Carlo, and Reinhard Pollak. 2011. "It's A Decent Bet that Our Children Will be Professor Too." *The Inequality Reader*: 499-516.

Krapohl, E., and R. Plomin. 2015. “Genetic Link between Family Socioeconomic Status and Children’s Educational Achievement Estimated from Genome-Wide SNPs.” *Molecular Psychiatry* 21:437-443.

Mare, Robert D. 1980. “Social Background and School Continuation Decisions.” *Journal of the American Statistical Association* 75(370): 295-305.

Mumper, Michael, Lawrence E. Gladieux, Jacqueline E. King, and Melanie E. Corrigan. 2011. “The Federal Government and Higher Education” Pp. 113-138 in *American higher education in the twenty-first century: Social, political, and economic challenges,* edited by Altbach, Philip G., Patricia J. Gumport, and Robert O. Berdahl. Baltimore, Maryland:

The Johns Hopkins University Press, 2011.

Nielsen, François, and J. Micah Roos. 2015. “Genetics of Educational Attainment and the Persistence of Privilege at the Turn of the 21st Century.” *Social Forces* 94(2): 535-561.

Nielsen, François. 2006. “Achievement and Ascription in Educational Attainment: Genetic and Environmental Influences on Adolescent Schooling.” *Social Forces* 85(1): 193-216.

Nielsen, François. 2016. “The Status-Achievement Process: Insights from Genetics.” *Frontiers in Sociology* 1: 9.

Okbay, Aysu, Jonathan P. Beauchamp, Mark Alan Fontana, James J. Lee, Tune H. Pers, et al. 2016. “Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment.” *Nature* 533(7604): 539-542.

Pareto, Vilfredo. 1909/1971. *Manual of Political Economy*. New York: A. M. Kelley.

Perry, Bruce D. 2002. “Childhood Experience and the Expression of Genetic Potential: What Childhood Neglect Tells Us about Nature and Nurture?” *Brain and mind* 3(1): 79-100.

Purcell, Shaun M., Naomi R. Wray, Jennifer L. Stone, Peter M. Visscher, Michael C. O’Donovan, Patrick F. Sullivan et al. 2009. “Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder.” *Nature* 460(7256): 748-752.

Raftery, Adrian E., and Michael Hout. 1993. “Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921-75.” *Sociology of Education* 66(1): 41-62.

Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, Jian Yang, Tõnu Esko, et al. 2013. “GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment.” *Science* 340(6139): 1467-1471.

Rietveld, Cornelius A., Tõnu Eskoc, Gail Daviesg, Tune H. Persc, Patrick Turleyi, et al. 2014. “Common Genetic Variants Associated with Cognitive Performance Identified Using the Proxy-Phenotype Method." *Proceedings of the National Academy of Sciences* 111(38): 13790-13794.

Roksa, Josipa, Eric Grodsky, Richard Arum, and Adam Gamoran. 2007. “United States: Changes in Higher Education and Social Stratification” Pp. 165-193 in *Stratification in Higher Education: A Comparative Study*, edited by Yossi Shavit, Richard Arum, and Adam Gamoran, pp.140-164. Stanford, CA: Stanford University Press.

Saunders, Peter. 2010. *Social Mobility Myths*. London: Civitas.

Scarr, Sandra, and Richard A. Weinberg. 1978. “The Influence of “Family Background” on Intellectual Attainment.” *American Sociological Review* 43(5): 674-692.

Shanahan, Michael J., and Scott M. Hofer. 2005. “Social Context in Gene–Environment Interactions: Retrospect and Prospect.” *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 60. Special Issue 1: 65-76.

Trow, Martin. 1972. “The Expansion and Transformation of Higher Education.” *International Review of Education* 18(1): 61-84.

Trow, Martin. 2007. “Reflections on the Transition from Elite to Mass to Universal Access: Forms and Phases of Higher Education in Modern Societies since WWII.” Pp. 243-280 in *International Handbook of Higher Education*. Netherlands: Springer.

Tucker-Drob, Elliot M., and Timothy C. Bates. 2016. “Large Cross-national Differences in Gene x Socioeconomic Status Interaction on Intelligence. *Psychological Science.* 27, 138–149. doi:10.1177/0956797615612727

Tucker-Drob, Elliot M., Mijke Rhemtulla, K. Paige Harden, Eric Turkheimer, and David Fask. 2011. “Emergence of a Gene× Socioeconomic Status Interaction on Infant Mental Ability between 10 Months and 2 Years.” *Psychological Science* 22(1): 125-133.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D’Onofrio, and Irving I. Gottesman. 2003. “Socioeconomic Status Modifies Heritability of IQ in Young Children.” *Psychological Science* 14(6): 623-628.

U.S. Department of Education, National Center for Education Statistics. 2016. *Digest of Education Statistics, 2014* (NCES 2016-006), Chapter 3.

Wray, Naomi R., Sang Hong Lee Divya Mehta, Anna A.E. Vinkhuyzen, Frank Dudbridge, and Christel M. Middeldorp. 2014. “Research Review: Polygenic Methods and their Application to Psychiatric Traits." *Journal of Child Psychology and Psychiatry* 55(10): 1068-1087.