Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Major: Health and Behavioral Economics

Different environmental factors in gene-environment interactions on educational attainment

---

Anne de Vries (481457)

Supervisor *Second assessor*

Rita Dias Pereira

Date final version:

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Contents

1. Introduction 2

1.1 Genetic and environmental influence 2

1.2 Research question and hypotheses 4

1.3 Scientific and societal relevance 4

2. Data and Methodology 6

2.1 Dataset 6

2.2 Description of variables 7

2.3 Methodology 10

3. Results 11

3.1 Results of the regression with different variables for environment 11

3.2 Results of the regression with the socioeconomic variable 13

4. Conclusions and Discussion 15

4.1 Discussion of the results 15

4.2 Concluding remarks 17

1. Introduction

Success in the labor market and income depend on different factors that interact with each other. These factors include environments, education and abilities. The interaction between genetics and environment has long been a topic of discussion and has come with different methodological challenges. Economists have tried to find and map these factors and the interactions between them (Griliches & Mason, 1972). Early research has found that education has an impact on the distribution of earnings and that education was impacted by a substantial amount of "institutional" factors (Becker & Chiswick, 1966). Rowe, Vesterdal and Rodgers (1998) investigated the genetic and environmental effects separately. They found that the genetic influence was stronger than often assumed in the social sciences. In this study, full and half siblings were observed while the same environment was assumed for both. This idea, also used in classical twin studies, did however result in biased outcomes because of the tendency to choose a partner with similar characteristics also known as assortative mating (Vandenberg, 1972). Further research by Loehlin, Harden and Turkheimer (2009) indicated that the genetic effect increased when assuming greater assortative mating. This is in line with the fact that higher parental income seems to increase heritability of IQ (Turkheimer et al., 2003).

1.1 Genetic and environmental influence

In recent years, a considerable amount of research relating to genes and their impact on different outcomes was conducted. When the Human Genome Project came to an end, this opened a new way of looking at genetic factors. Instead of comparing siblings and twins with each other under a substantial amount of assumptions, it now became possible to identify specific genes. This research was mainly conducted in the form of candidate-gene studies. This type of studies focusses on the association between phenotypes and a set of pre-specified SNPs (single-nucleotide polymorphisms). SNPs are tiny variations in the DNA sequence. When studying the complex behavioral traits, the statistical power in these candidate-gene studies was not sufficient. This low statistical power was one of the reasons that most findings could not be replicated. This led to a great deal of false findings in the field of behavioral genetics (Benjamin et al., 2012) and eventually to an editorial policy from the journal *Behavioral genetics* (Hewitt, 2012). As a result, the focus in identifying genes responsible for different complex phenotypes shifted towards Genome-wide association studies (GWAS)(Pearson & Manolio, 2008). The GWAS on the effects of genes on educational attainment conducted by Rietveld et al. (2014), found that three different SNPs significantly impacted educational attainment. This paper proved the robustness of GWAS on complex behavioral traits. In the paper the results of Rietveld et al. (2013) were replicated in three different studies. In later

research by Okbay et al. (2016) 74 SNPs were identified from a sample size of 293,723 people. In the study by Lee et al. (2018) substantially more markers were found from a sample size of 1,131,881 people. A strong relationship between genetics and educational attainment was hereby established in research.

Even though genes have a considerable effect, educational level is dependent on environmental factors as well. These factors consist of the financial, educational and social background of the childhood home. Mani et al. (2013) found that a poor environment will have a negative effect on cognitive performance. This could mean that social mobility is limited in society and that therefore a big part of human potential is not fulfilled. This also means that scores, for example from IQ-tests, cannot be trusted in this type of research. Two individuals with the same score but from different backgrounds are unlikely to have the same genetic potential. Wilson, Haveman and Smeeding (2008) showed that there is a big income-related gap in higher education. The income and socioeconomic background of the parents has a substantial influence on the children's chances of graduating a top tier college. This indicates that the educational success is reliant on the financial situation of the household. In addition to financial background, the educational background of the parents has been found to be a predictor for potential downward mobility in research in Sweden (Alm, 2011). Governments are unable to change genetics, but if the environmental factors correlating with genetic potential can be identified this could open up a variety of options for policy. The socioeconomic variables created by Vable et al. (2017) will be the measures for environmental influences. In this paper the influences of human capital, financial capital and social capital will be compared. The human capital index is the sum of parental educational years. The financial capital scale consists of the sum of the average financial resources in childhood and a variable for financial instability. The social capital variable is a sum of the maternal investment and the family structure. These are measured by the self-reported quality of the relationship with the mother and the number of household adults. This way we can identify the different types of socioeconomic influences. The educational attainment will be represented by two different measures: the years of education and the highest degree obtained.

Traditionally nature and nurture were thought of as two independent factors but in recent years it has been established that most phenotypic outcomes are a result of an interaction between genes and environment (Heckman, 2007). Therefore, it is important to explore the interplay between the environmental and genetic factors (Turkheimer, 2000). This field of research tries to estimate a part of the interplay between genetic inputs and life outcomes. The small effect of genetics found on phenotype led to the idea of 'missing' heritability (Zuk et al., 2012). A possible explanation for this is the fact that the influence of genetics is partly conditional on the environment (Manolio et al., 2009, Zuk et al., 2012). As a

result there has been a substantial amount of research exploring the gene-environment interactions on different phenotypes. These phenotypes include obesity (Qi & Cho, 2008), ADHD (Grizenko et al., 2012) and smoking (Nilsson et al., 2009). Educational attainment has been a topic for gene-environment interaction studies as well (Thompson, 2014).

1.2 Research question and hypotheses

The focus of this paper will be on the interaction terms between the different socioeconomic measures and the polygenic scores for educational attainment with the following research question in mind: How do different environmental factors interact with the genetic predictors for educational attainment? Rowe et al. (1999) displayed that heritability of verbal IQ was considerably higher among children with higher educated parents, suggesting a positive interaction. Wertz et al. (2020) found that the parental investment has a positive influence on attainment. The parental investment was also correlated with the SNPs with positive influence on education. Therefore the relationship between parenting and educational attainment was slightly confounded by genetics. Thompson (2014) found that the economic background measured through household income also has a positive association with educational attainment. There is extensive literature suggesting relations between the educational level and financial status of the parents and the resulting educational attainment. However not much information is directly stating this for the self-reported relationship with the mother. Krein and Beller (1988) did however find that children from single-parent families tend to have worse educational attainment. A part of this effect is due to the lack of parental investment in single-parent homes. To summarize, from the literature we can assume that both financial background and the educational background of the childhood improve the chances of higher educational attainment. The relationship between social capital and educational attainment is less clear. To test whether the socioeconomic background interacts with the polygenic scores the same hypothesis will be tested for each measurement of environment. The following hypotheses will be tested:

H_0: The socioeconomic background does not interact positively with the polygenic scores for educational attainment

H_1: The socioeconomic background interacts positively with the polygenic scores for educational attainment

1.3 Scientific and societal relevance

This research is important for different reasons. First of all, results of this research might imply the loss of potential human capital. If genes interact with the environment a substantial amount of people with bad

circumstances will never reach their potential, resulting in a loss of potential human capital. It also means that talented people from a poor background will in all likelihood not reach their educational potential. This has a negative influence on the overall social mobility in society. From a governmental perspective, it is important to know the reasons behind this. This research helps identifying the problems and the reasons for it. A good example could be the allocation of federal aid for students. Research has shown that this is an effective measure for nontraditional students (Seftor & Turner, 2002).

This research will specifically investigate educational attainment. Educational attainment seems to have a considerable influence on different factors and is therefore an important topic for governments. The importance for educational attainment is illustrated by the economic progress that comes with it (Mankiw et al., 1992), but educational attainment is also an important indicator for different social outcomes. Research shows that educational attainment can have a positive effect on life expectancy (Barro and Lee, 1994) and reduces child mortality (Breierova and Duflo, 2004). Health improvements related to education do not seem to alter by race or gender (Cutler et al., 2006). Zhou et al. (2019) found that investing in the low-income youth should result in a strong increase in social mobility. A high educational attainment is therefore substantially more important than just the diploma, because it might have a positive impact on different parts of people's lives. This research will contribute to the existing literature because of the new variables used for environment. It will widen our view on the impact of different socioeconomic factors on educational attainment. For the polygenic scores the results from Lee et al. (2018) will be used. To lower the chances of population stratification we will also use the different principal components found in that study. This study will cover different social variables as opposed to most earlier research and will therefore extend the available knowledge on possible gene-environment interactions. Also, the use of the new extended information on predicting SNPs and the new polygenic scores add to the value of this research. With the discovery of specific SNPs we will be able to approximate the causal effect of nature more precisely. These PGS-scores that follow from huge datasets are less likely to be impacted by confounders as opposed to measurements like IQ-scores. This research therefore differs from earlier research using measures like IQ-scores which are highly correlated with environment. A substantial amount of this type of research has been conducted using twin studies. These studies assume that twins grow up in the same environment while this is often not the case. Polygenic scores are therefore expected to be better predictors for genetics, especially when the scores are attained trough GWAS of huge databases.

2. Data and methodology

2.1 Dataset

For this study I will use three different datasets. The 2018 version of the longitudinal dataset from the Health and Retirement Study (HRS) will be used as the master dataset. Next to this dataset I will use two complementary datasets created by Vable et al. (2017) and Lee et al. (2018). The dataset from Vable et al. consists of different variables for socioeconomic background in childhood. The information is gathered through publicly available data and matched with the HRS. All the variables were tested and found acceptable on their construct validity and predictive validity. The dataset from Lee et al. (2018) consists of the outcomes from genome-wide association analyses of roughly 1.1 million individuals of European-descent, from 71 cohorts, with a range in birth year of approximately 1901 to 1989. One of the outcomes of this research are the polygenic scores for educational attainment obtained from the GWAS. These polygenic scores will be used together with the principal components recognized by the research. The interaction terms were created by multiplying the polygenic scores with the different measures for educational attainment. The dataset of Lee et al. (2018) only contains polygenic scores of people with European ancestry. After merging the three datasets together the merged dataset contains 8652 observations. Since all incomplete observations were dropped, the merged dataset should only contain data of individuals with a European ancestry. However, after checking this with the data on the race of participants from the HRS there were still a few participants with an African American background. All observations that were not identified as white or Caucasian have been dropped.

The HRS contains an ordered categorical variable for the highest degree obtained. The categories of this variable are listed below.

Highest degree

0. No degree
1. GED (General Educational Development)
2. HS (High School Diploma)
3. HS/GED (HS or GED)
4. AA/ Lt BA (Two-year college degree)
5. BA (Four-year college degree)
6. MA/MBA (Master's degree)
7. Law/MD/PhD (Professional degree, Ph.D., M.D., J.D.)
8. Other

To use this variable as an ordered categorical variable it must be clear that the categories are in a logical order. There are however two downsides in the order of this variable. Firstly, the overlap at categories 1, 2 and 3. *GED* represents a test on General Educational Development which has roughly the same value as a high school diploma. Passing the GED will be rewarded with a certificate of high school equivalency and will allow students to enroll in college. With *HS/GED* the person has a high school diploma or a GED-test. The distinction between these categories and the first two is therefore unclear. The difference in academic skills between a GED and a High school diploma is minimal. Thus, the categories 1, 2 and 3 will be taken together and recoded to create: *HS/GED*. The category 'other' only contains one observation. There is no clear reason why this observation should be higher ordered in this variable. Because of this the observation has been dropped. After this there are six categories left. The final dataset contains 8610 observations.

2.2 Description of variables

The formula used to test gene-environment interaction looks as follows:

$$Educational\ Attainment = \beta_0 + \beta_1 PGS + \beta_2 Environment + \beta_3 PGSxEnvironment + \beta_i X_i + \varepsilon$$

Where β_0 is the constant, *PGS* is the genetic effect, *Environment* the environmental effect and *PGSxEnvironment* the interaction term. ε stands for the error term and $\beta_i X_i$ represents possible control variables. This formula is the base for this paper. The variable of interest here is the interaction term between the genetic effect (PGS) and environment. In this section the variables for each part of the formula will be introduced. After this we can complete the methodology. Summary statistics of the important variables are provided in table 1 and table 2. The mean and standard deviation will be provided for the continuous variables and the percentages for the categorical variables.

In this paper the dependent variables will be the number of educational years and the highest degree obtained. As shown in the table, *educational years* is a simple numerical variable with values from 0 to 17 representing the number of years and individual spent in education. The average number of years in education in this set is 13.155 years and the proportion of individuals with 10 years of education or less is only 11.17%, as shown in table 5 in the appendix. These observations can be explained by the fact that attending primary school and part of high school is often obligatory. However, it does mean that the majority of individuals who did not obtain a high school degree still had at least 10 years of education. Therefore, *educational years* and the *highest degree* will both be used as dependent variables as the possible outcomes for *highest degree* provide a more meaningful distinction for academic success.

The genetic effect is measured with the results from the GWAS by Lee et al. (2018). The polygenic

scores are the weighted sum of genetic traits found for educational attainment obtained from the GWAS and only represent the predictors found in their study. These scores contain a small portion of the genetic information predicting educational attainment. A potential problem with the results of a GWAS is that the association between the polygenic scores and the educational attainment is due to population stratification. This would mean that the association is caused by genetic differences between ancestral subpopulations as a result of non-random mating. Lee et al. (2018) therefore recommend to control for principal components since that would minimize the risk. The ten recommended ancestry-specific principal components will be added as control variables.

As mentioned in the introduction three different variables for socioeconomic environment and one variable combining all three will be used. The variables used are *human capital, social capital* and *financial capital* and are coded in a way that higher numbers reflect a higher score for capital and therefore a preferable socioeconomic status. The variable for human capital is conceptualized as an index. In the case of missing or dichotomized data Vable et al. (2017) imputed continuous information using expectation maximization. *Social capital* and *financial capital* are scale variables. The variables for human, social and financial capital have been combined to create the childhood socioeconomic score index or *CSES*, which is the average of the three variables. The interaction terms were created by multiplying the PGS with the variables for environment.

In addition, we will control for factors that are expected to influence the educational attainment. In the dataset the discrete variable birth-year ranges from 1905 to 1974. This cohort differs from the cohort mentioned above because of the incomplete observations that were dropped. Since environment is expected to be different for people born this far apart, we will control for birthyear in the analysis. Furthermore, we are looking at educational attainment for men and women in the first half of the 20th century when women were not always expected to get education. Therefore, in addition we will control for gender. *Gender* is a dichotomous variable where 1 represents males and 2 represents females. 58,41% of the individuals in the sample are women as shown in table 2.

Table 1: Summary statistics of numerical variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Birthyear	8,610	1937.806	10.420	1905	1974
Educational years	8,597	13.155	2.542	0	17
Human Capital	8,610	.412	.852	-2.574	2.420
Social Capital	8,610	-.011	1.130	-5.638	1.476
Financial Capital	8,606	.033	1.024	-3.101	3.044
Socioeconomic index	8,610	.274	.868	-3.322	2.809
Polygenic Scores	8,610	-.242	.146	-.774	.323
pc1	8,610	.0000152	.0108	-.0189	.0553
pc2	8,610	-3.83e-06	.0108	-.0393	.0733
pc3	8,610	6.02e-06	.0108	-.0395	.0481
pc4	8,610	.0000118	.0108	-.0441	.0711
pc5	8,610	-.0000201	.0108	-.0432	.0372
pc6	8,610	3.38e-06	.0108	-.0542	.0435
pc7	8,610	-.0000127	.0108	-.0537	.0374
pc8	8,610	-3.76e-06	.0108	-.0380	.0521
pc9	8,610	-.0000151	.0107	-.0427	.0361
pc10	8,610	-4.82e-06	.0108	-.0585	.0462

Table 2: Summary statistics of categorical variables

Variable	Category	Frequency	Percentage	Cumulative
Gender	Male	3,581	41.59	41.59
	Female	5,029	58.41	100.00
Highest Degree	No degree	1,116	12.96	12.96
	HS/GED	4,910	57.03	69.99
	AA/ Lt BA	464	5.39	75.38
	BA	1,268	14.73	90.10
	MA/MBA	653	7.58	97.69
	Law/MD/PhD	199	2.31	100.00

2.3 Methodology

In this section the base formula will be combined with the mentioned variables to complete the regression. The regression will contain all environmental factors and interaction terms at the same time. This means that each regression will have three variables of interest. The benefit of using all interaction terms and environmental factors in the regression is that the environmental influences also act as control variables. The first regression will be performed with *educational years* as measure for educational attainment, in the second regression educational attainment will be defined by highest obtained degree. The first two regressions will then look like this:

(1) $Educational\ years = \beta_0 + \beta_1 Polygenic\ scores(PGS) + \beta_2\ Financial\ Capital + \beta_3\ Social\ Capital + \beta_4 Human\ Capital + \beta_5\ (Financial\ Capital * Polygenic\ Scores) + \beta_6\ (Social\ Capital * Polygenic\ Scores) + \beta_7\ (Human\ Capital * Polygenic\ Scores) + \beta_8 Gender + \beta_9 Birthyear + \beta_{10} PC1 + \beta_{11} PC2 + \beta_{12} PC3 + \beta_{13} PC4 + \beta_{14} PC5 + \beta_{15} PC6 + \beta_{16} PC7 + \beta_{17} PC8 + \beta_{18} PC9 + \beta_{19} PC10 + \varepsilon$

(2) $Highest\ degree = \beta_0 + \beta_1 Polygenic\ scores(PGS) + \beta_2\ Financial\ Capital + \beta_3\ Social\ Capital + \beta_4 Human\ Capital + \beta_5\ (Financial\ Capital * Polygenic\ Scores) + \beta_6\ (Social\ Capital * Polygenic\ Scores) + \beta_7\ (Human\ Capital * Polygenic\ Scores) + \beta_8 Gender + \beta_9 Birthyear + \beta_{10} PC1 + \beta_{11} PC2 + \beta_{12} PC3 + \beta_{13} PC4 + \beta_{14} PC5 + \beta_{15} PC6 + \beta_{16} PC7 + \beta_{17} PC8 + \beta_{18} PC9 + \beta_{19} PC10 + \varepsilon$

To measure the combined effect of the different environmental factors an additional regression will be performed where the factors financial, social and human capital are combined into the socioeconomic score index variable. In this regression there will not be any other environmental factors or interaction terms except for the socioeconomic score. The rest of the regression will remain the same and looks as follows:

(3) $Educational\ Attainment = \beta_0 + \beta_1 Polygenic\ scores(PGS) + \beta_2\ CSES + \beta_3\ (CSES * Polygenic\ Scores) + \beta_4 Gender + \beta_5 Birthyear + \beta_6 PC1 + \beta_7 PC2 + \beta_8 PC3 + \beta_9 PC4 + \beta_{10} PC5 + \beta_{11} PC6 + \beta_{12} PC7 + \beta_{13} PC8 + \beta_{14} PC9 + \beta_{15} PC10 + \varepsilon$

3. Results

3.1 Results of the regression with three variables for environment

In the introduction four hypotheses were formulated relating to the environmental factors. A positive and significant interaction term is expected for financial capital, social capital, human capital and the socioeconomic index (cses). After running the first two regressions introduced above, we get the results shown in table 3.

Table 3: Results of the linear regressions of the different environmental factors on Highest degree and Educational years where pc stands for principal component

	Educational years	Highest degree
Polygenic score (PGS)	4.278***	1.873***
	(0.184)	(0.0913)
Financial capital	0.155***	0.0966***
	(0.0464)	(0.0231)
Social capital	0.0192	0.0317
	(0.0423)	(0.0210)
Human capital	1.080***	0.601***
	(0.0552)	(0.0274)
Financial capitalxPGS	-0.0553	0.123
	(0.163)	(0.0811)
Social capitalxPGS	-0.122	0.0426
	(0.147)	(0.0729)
Human capitalxPGS	0.0227	0.655***
	(0.192)	(0.0952)
Gender	-0.283***	-0.264***
	(0.0478)	(0.0237)
Birthyear	0.0148***	0.00703***
	(0.00240)	(0.00119)
pc1	12.37***	3.961***
	(-2.192)	-1.090)
pc2	-2.927	(-0.299
	(-2.182)	(-1.085)
pc3	5.483*	2.213*
	(-2.181)	(-1.084)
pc4	0.857	-0.704
	(-2.182)	(-1.085)
pc5	2.341	1.737
	(-2.183)	(-1.085)

pc6	3.804	1.550
	(-2.181)	(-1.084)
pc7	-0.823	-0.854
	(-2.182)	(-1.084)
pc8	3.058	1.337
	(-2.182)	(-1.085)
pc9	-1.186	-0.0632
	(-2.184)	(-1.086)
pc10	-3.135	-1.959
	(-2.180)	(-1.084)
Constant	-14.52**	-11.42***
	(-4.639)	(-2.303)
N	8593	8606

Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

There are significant results for the influence of the identified genes on the number of educational years and on the highest degree obtained. The effects the polygenic scores do however differ considerably across the regressions. As presented in table 3, a point higher in the *polygenic score* significantly results in 4.278 extra years of education. The coefficient of the *polygenic score* in the regression on highest degree is 1.873 and significant as well. This means that individuals with a polygenic score of a point higher are expected to have a degree almost 2 categories higher. The difference between these coefficients is explained by the difference between the variables. Most steps in the categorical variable represent a degree of 2-3 years.

Out of the three environmental variables *financial capital* and *human capital* have a significant effect on both dependent variables. An increase of one point in financial capital results in 0.155 more educational years and a 0.0966 rise in the categories for degree. The control variables gender and birthyear both have a significant effect as well. The effect of gender is negative suggesting that women in this sample are less educated than man in similar circumstances. The effect of birthyear suggests that people got slightly more educated over the years.

The interaction term of financial capital and the polygenic score is negative for educational years and positive for highest degree. The coefficients are small and do not have a significant effect ($p > 0.05$). This means that for this sample the hypothesis that financial capital has a positive and significant interaction term should be rejected. The same applies to the interaction term of social capital and therefore for this sample we must also reject the hypothesis that social capital has a positive and significant effect. The coefficient for the interaction term with human capital differs between the two regressions.

There is a small and insignificant interaction effect on the educational years. This suggests that human capital and the polygenic score do not affect each other when predicting the educational years. In the regression on highest degree there is a positive and significant effect of the interaction term ($p<0.001$). This suggests that the influence of a higher polygenic score increases when the parents have a higher education. We should still reject the hypothesis because of the insignificant effect for the regression on educational years.

3.2 Results of the regression with the socioeconomic variable

The fourth hypothesis, stating that the index variable for socioeconomic status has a positive and significant interaction term, will be tested based on the second regression introduced earlier. Table 4 shows the results of the linear regressions. The results for the variable of interest look roughly the same as with the interaction term of human capital. The interaction is positive and significant for the highest degree obtained but not for the number of educational years. This means that the last hypothesis should also be rejected. The rest of the coefficient differ marginally from the coefficients in the earlier regression. The polygenic score still has a significant effect. The socioeconomic index variable also influences both dependent variables significantly.

Table 4: *Results of the linear regressions of the socioeconomic index on Highest degree and Educational years*

	Educational years	Highest degree
Polygenic score	4.685***	2.127***
	(0.178)	(0.0875)
cses_index	0.929***	0.544***
	(0.0528)	(0.0260)
csesxPGS	0.0378	0.680***
	(0.183)	(0.0904)
Gender (female)	-0.310***	-0.275***
	(0.0488)	(0.0241)
Birthyear	0.0277***	0.0124***
	(0.00238)	(0.00117)
pc1	8.732***	2.369*
	(-2.238)	-1.105
pc2	-3.791	-0.699
	(-2.235)	(-1.103)
pc3	6.338**	2.605*
	(-2.235)	(-1.103)
pc4	1.809	-0.369
	(-2.236)	(-1.104)
pc5	1.987	1.671
	(-2.237)	(-1.104)
pc6	3.778	1.484
	(-2.236)	(-1.103)
pc7	-0.558	-0.770
	(-2.237)	(-1.104)
pc8	3.368	1.393
	(-2.236)	(-1.104)
pc9	-2.235	-0.537
	(-2.238)	(-1.104)
pc10	-3.096	-1.984
	(-2.235)	(-1.103)
Constant	-39.13***	-21.60***
	(-4.603)	(-2.269)
N	8597	8610

Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

4. Conclusions and discussion

The interplay between nature and nurture has been the subject of discussion in different scientific fields. They have historically been treated separately as well as in conjunction. The view of nature and nurture as separate factors dominated parts of the 20th century (Logan & Johnson, 2007). This view has been surpassed in science for a while now. With the upcoming developments in genotyping individuals there is substantially more achievable in the field of geno-economics. Gene-environment interactions play a substantial role in this. Since the effect of environment on outcomes could be enhanced by genetics and vice versa, this field could lead to highly effective policy.

4.1 Discussion of the results

The purpose of this study was to research the influence of different variables for environment and their interaction with the genes. For social capital and financial capital there was no significant interaction term found. The results suggested that social capital and financial capital were not even significant predictors for educational attainment. Social capital did not have a significant effect on educational attainment in this sample. Financial capital did have a significant effect, but the effect was minor. The size of the effect for financial capital was unexpected given that there is literature showing that the financial background has an influence on educational attainment. A possible explanation for this is that the variable for financial capital was not extensive enough. The number of children per household or the spending patterns could be needed to increase the explanatory value. Another possible explanation is that the predictors correlate with each other. Education often influences income, meaning that a part of the financial effect could be explained by the educational background of the parents. In this way, incorporating educational background in this regression might have reduced the effect of *financial capital*. The hypotheses, expecting a significant interaction term, do not hold for both social and financial capital. Human capital and the socioeconomic index (CSES) were better predictors of educational success. The coefficients for both were higher in the regressions on highest degree than on educational years. This resulted in a positive and significant interaction term in the regressions on highest degree but not on educational years. A possible explanation could be that the degree is a clearer measurement unit for educational attainment due to the fact that it provides result-based differences and not time-based differences. The fact that there is some form of obligation in educational years could also be an explanation. It could on the other hand also suggest that a better socioeconomic background increases educational success in less years. This could be an interesting topic for future research. However, this research focused on both educational years and the highest degree and therefore the hypotheses should be rejected. If the hypotheses only represented the

highest degree as dependent variable, the hypothesis would be accepted for *human capital* and *CSES*. The interpretation of the positive and significant interaction between human capital and the polygenic score in the regression on highest degree comes with a few problems. The first problem in interpreting the results is the fact that the genome of the parents influences both the human capital as the polygenic scores. This would mean that both the effects of *human capital* and the *polygenic score* would partly be the mediated effect of the parental genes on educational attainment. The second problem is the fact that educational attainment is in addition related to other positive phenotypes. As seen in the introduction these phenotypes also include phenotypes like health. This could mean that the positive interaction term is actually caused by different phenotypical outcomes which are represented by educational attainment in this regression.

This type of study is subject to a number of limitations. The first limitation is the use of polygenic scores found in a GWAS. Even though the polygenic scores resulted from an extensive GWAS, they only account for a small part of the genetic influences on educational attainment. This means that they do not give a clear view on the entire genetic disposition. Another problem with GWAS is that it is unclear to what extent the SNPs influence the phenotypic outcome. Therefore, we cannot know for certain what the causal effect is between the polygenic scores and the educational attainment. The SNPs found for educational attainment are directly linked to the educational attainment people got, meaning the genetic predisposition and the resulting outcome are dependent on each other.

Another limitation is the use of weak predictors as independent variables. The first problem was that not every measure for environment had a significant and positive effect on the educational attainment. The variables that did are not all specific enough or are hard to translate into policy implications. An obstacle with finding interactions is the fact that we cannot always know the underlying correlations. To provide a clear view of the specific factors that interact with each other we need to have a clear view of the existing correlations. It is however challenging to come up with an exogenous measure for environment. Part of this problem is the earlier explained idea of assortative mating, resulting in men and women marrying more frequently with partners with the same educational level. This would increase the estimated heritability. In addition, there might be genes involved in parental education that were not accounted for in the polygenic scores. Part of the interaction found there could therefore be the result of accounted genes interacting with unaccounted genes.

Another potential problem is the presence of unobserved variables that influence both environment and the outcome. These variables will then confound the results and therefore create a bias. An example of this could be personality traits. These traits might influence both the financial and

educational success of the parents and also have a positive impact on an individual's education. The selection of dependent variables plays an important role as well in these types of interactions. The first question when selecting a variable is, whether or not it is a useful variable. Educational attainment is a suitable variable because it has an influence on a variety of other outcomes. As mentioned in the introduction, there is evidence that higher education also influences of social outcomes such as health and social life (Hout, 2012). In this paper we saw that the different measures for educational attainment can result in very different outcomes. Therefore, there must be a clear view of which form of attainment influences individuals' lives the most. The specific choice of variables is therefore important to enable useful policy recommendations.

4.2 Concluding remarks

From this paper we must conclude that there is no decisive evidence in this sample for a gene-environment interaction on both educational years and the highest degree obtained. From the regressions we can conclude that social capital did not have significant predictive power. Human capital did seem to be a good predictor but in the interaction there are possible biases. The variables for educational attainment are not exogenous enough to draw conclusions from it.

While this paper does not have results for policy implications there are some implications for future research. The exogeneity of the environmental variables is very important. This does not mean that the variable should be completely exogeneous, but research should focus on finding the part of the effect that is. This should be done using methods where unobservable differences do not bias the results. Such methods could be instrumental variables, regression discontinuity and differences-in-differences. Furthermore, it is important to gather a broader range of environmental variables. This will make sure that environmental variables like financial background will not also carry the effect of education. Because of the fact that the degree gives a clearer distinction of educational attainment there is reason to assume that the parental education interacts positively. This sample can however not provide a decisive answer. The results do suggest that having higher educated parents leads to a higher degree without a big increase in educational years. This could be interesting for future research because this would mean that individuals with a better socioeconomic background are more efficient with the same number of educational years.

At last, we should be aware of the information provided by the GWAS. The polygenic scores are not completely exogenous either. GWAS does not give us the exact causal effect and only explains a small part of the genotype. It is therefore important is this field of research that we map the different ways

environment and genes influence each other and correlate with each other. In the future more genetic studies will be conducted. This will eventually lead to a better prediction of the genetics. If we map specific environmental factors and identify relations between the environmental factors, better prediction will be drawn from gene-environment interactions.

Reference list

Alm, S. (2011). Downward social mobility across generations: the role of parental mobility and education. *Sociological Research Online*, *16*(3), 1-14.

Barro, R. J., & Lee, J. W. (1994). Sources of economic growth. In *Carnegie-Rochester conference series on public policy* (Vol. 40, pp. 1-46). North-Holland.

Becker, G. S., & Chiswick, B. R. (1966). Education and the Distribution of Earnings. *The American Economic Review*, *56*(1/2), 358-369.

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, Gudnason V, Lichtenstein P. The promises and pitfalls of genoeconomics. Annual Review of Economics. 2012; 4(1):627–662.

Breierova, L., & Duflo, E. (2004). *The impact of education on fertility and child mortality: Do fathers really matter less than mothers?* (No. w10513). National bureau of economic research.

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: evaluating theories and evidence* (No. w12352). National bureau of economic research.

Griliches, Z., & Mason, W. M. (1972). Education, income, and ability. *Journal of political Economy*, *80*(3, Part 2), S74-S103.

Grizenko, N., Fortier, M. E., Zadorozny, C., Thakur, G., Schmitz, N., Duval, R., & Joober, R. (2012). Maternal stress during pregnancy, ADHD symptomatology in children and genotype: gene-environment interaction. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *21*(1), 9.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. Proceedings of the National Academy of Sciences, 104(33):13250–5.

Hewitt, J. K. (2012). Editorial policy on candidate gene association and candi-date gene-by-environment interaction studies of complex traits.Behaviorgenetics,42(1), 1.

Hout, M. (2012). Social and economic returns to college education in the United States. Annual review of sociology, 38, 379-400.

Krein, S. F., & Beller, A. H. (1988). Educational attainment of children from single-parent families: Differences by exposure, gender, and race. *Demography*, *25*(2), 221-234.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, *50*(8), 1112-1121.

Loehlin, J. C., Harden, K. P., & Turkheimer, E. (2009). The effect of assumptions about parental assortative mating and genotype–income correlation on estimates of genotype–environment interaction in the national merit twin study. Behavior Genetics, 39(2), 165-169.

Logan, C. A., & Johnston, T. D. (2007). Synthesis and separation in the history of "nature" and "nurture". *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *49*(8), 758-769.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. science, 341(6149), 976-980.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, *107*(2), 407-437.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.

Nilsson, K. W., Oreland, L., Kronstrand, R., & Leppert, J. (2009). Smoking as a product of gene–environment interaction. *Upsala journal of medical sciences*, *114*(2), 100-107.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Salomaa, V. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539-542.

Pearson, T. A., & Manolio, T. A. (2008). How to interpret a genome-wide association study. *Jama*, *299*(11), 1335-1344.

Qi, L., & Cho, Y. A. (2008). Gene-environment interaction and obesity. *Nutrition reviews*, *66*(12), 684-694.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. science, 340(6139), 1467-1471.

Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., & Vinkhuyzen, A. A. E... Social Science

Genetics Association Consortium.(2014). Replicability and robustness of genome-wide-association studies for behavioral traits. Psychological Science, 25(11), 1975-1986.

Rowe, D. C., Vesterdal, W. J., & Rodgers, J. L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. Intelligence, 26(4), 405-423.

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child development*, *70*(5), 1151-1162.

Seftor, N. S., & Turner, S. E. (2002). Back to school: Federal student aid policy and adult college enrollment. *Journal of Human resources*, 336-352.

Thompson, O. (2014). Economic background and educational attainment the role of gene-environment interactions. *Journal of Human Resources*, *49*(2), 263-294.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. Current Directions in Psychological Science, 9(5):160–164.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological science*, *14*(6), 623-628.

Vable AM, Gilsanz P, Nguyen TT, Kawachi I, Glymour MM (2017) Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the Health and Retirement Study. PLoS ONE 12(10): e0185898. https://doi.org/10.1371/journal.pone.0185898

Vandenberg, S. G. (1972). Assortative mating, or who marries whom?. *Behavior genetics*, *2*(2), 127-157.

Wertz, J., Moffitt, T. E., Agnew-Blais, J., Arseneault, L., Belsky, D. W., Corcoran, D. L., ... & Caspi, A. (2020). Using DNA from mothers and children to study parental investment in children's educational attainment. *Child development*, *91*(5), 1745-1761.

Wilson, K., Smeeding, T., & Haveman, R. (2008). The Role of Education and Occupation in US Social Mobility: A Glimpse Inside the Black Box.

Zhou, X. (2019). Equalization or selection? Reassessing the "meritocratic power" of a college degree in intergenerational income mobility. *American Sociological Review*, *84*(3), 459-485.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193-1198.

Appendix

Table 5: Proportions of educational years

Educational years	Freq.	Percent	Cum.
0.none	12	0.14	0.14
2	4	0.05	0.19
3	7	0.08	0.27
4	9	0.10	0.37
5	14	0.16	0.54
6	35	0.41	0.94
7	62	0.72	1.66
8	236	2.75	4.41
9	221	2.57	6.98
10	361	4.20	11.18
11	325	3.78	14.96
12	3,117	36.26	51.22
13	707	8.22	59.44
14	947	11.02	70.45
15	360	4.19	74.64
16	1,059	12.32	86.96
17.17+ yrs	1,121	13.04	100.00
Total	8,597	100.00	