ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Behavioural and Health Economics

The predictive power of "education genes" beyond parental education and socioeconomic status in the Health and Retirement Study

Name student: Patrick Zuidhoorn

Student ID number: 432633

Supervisor: S.F.W. Meddens

Second assessor:

Date final version: 07 / 08 / 2020

Abstract

Due to large-scale genome wide association studies (GWAS), the explanatory power of polygenic scores become more and more valuable to researchers. This has led to an increase of researchers investigating polygenic scores, leading to an interesting debate about how and when to use polygenic scores in society. I study the predictive power of polygenic scores for education beyond parental education and socioeconomic status in the HRS and compare findings to the results of Morris et al. (2020a) who use the ALSPAC study. In addition, I looked at different aspects of parental socioeconomic status and how they affect the added predictive power of education genes, and if the extra explanatory power of educational attainment provided by polygenic scores vary across different parental socioeconomic statuses. In line with Morris et al. (2020a), predictions from polygenic scores are inferior to parental environment factors. In contrary to Morris et al (2020a), for people in the HRS, education genes contribute more in predicting educational attainment, beyond parental education and socioeconomic status. Moreover, my results suggest that when parental education is available, other socioeconomic factors, such as financial or social capital, provide little extra value to explain completed years of education. If the extra explanatory power of educational attainment provided by polygenic scores varies across different parental socioeconomic statuses was very hard to interpret, but indicated that there was a bigger increase in R^2 for high parental SES groups, relative to low and middle parental SES groups.

Contents page

Introduction

Educational attainment is partly explained by genes (Rietveld et al., 2013). Effects of genetic variants are accumulated into so-called "polygenic scores" and correlate with educational attainment and other phenotypes (Lee et al., 2018). A phenotype is the complete set of all observable characteristics of an individual as a result of its genotype and environmental influences. Another aspect that explains the variation in completed years of education relatively accurate is parental socioeconomic status, which also partly captures the genetic effect (Sewell & Shah, 1967). Parental socioeconomic status predicts just 7 percent of the variation in educational attainment, compared to 12 percent of polygenic scores for education. Still, parental socioeconomic status is accounted for a lot more within studies, than for education genes, even though education genes have a stronger predictive power (Lee et al., 2018).

Many genetic factors, such as intelligence, personality and emotional stability, drive educational attainment (Okbay et al., 2016). It is likely that polygenic scores (PGS) for education correlate with many aspects of an individuals' environment (Abdellaoui et al., 2019; Harden et al., 2020). In other words, part of genotype from parents that is not inherited by their children can still explain a child's educational attainment by its phenotype. This is also called genetic nurture or dynastic effects (figure 1). Genetic nurture is limiting the results of many researches that look at the predictive power of polygenic scores. Controlling for dynastic effects might indicate the role of environmental factors (Willoughby et al. 2019). To this point, it is not clear if polygenic scores for education are more valuable than phenotypic information solely, such as sex, age and parental education and environment, what causes the dynastic effects. When considering this background information, a lot of attention has been given to education, as parental education seems to be a relatively promising predictor of a child's completed years of education (Chevalier, 2004). Also, Bates et al. (2018) have shown that parental environment strongly correlates with polygenic scores for education.

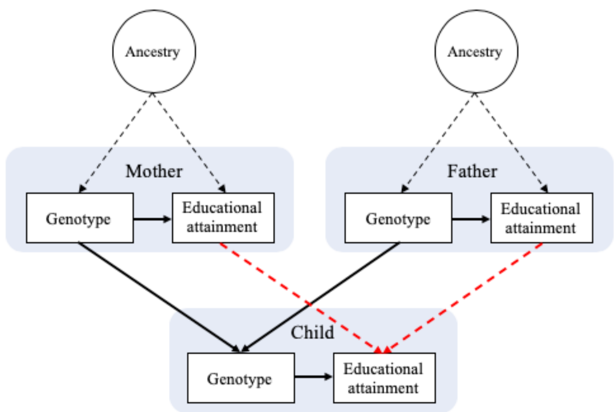

Figure 1: Red lines indicate genetic nurture effect. Adapted from: Morris et al. (2020a)

The role of parental environment in predicting health outcomes and educational attainment is already clearly present in current literature and helps to explain health issues, such as cardiovascular diseases, high blood pressure and diabetes (Lee et al., 2017; Winkleby et al., 1992). Additionally, parental environment also explains completed years of education and completed years of education is a good predictor for income. Therefore, parental environment is a good indicator for income. Further, explanatory power of polygenic scores have been discussed already by many researchers, for example *Morris et al.* in *"Can education be personalised using pupils' genetic data?"*. My paper is built upon the work of Morris et al. (2020a). They have investigated the explanatory power of genomes beyond easily available and obtainable phenotypic data, such as sex, month of birth and prior achievement, for pupil educational achievement and used data from a UK cohort (ALSPAC study). The answer is that genomes do not explain a bigger proportion of the variance of educational achievement on top the proportion of the variance of educational achievement explained by parental educational achievement. Moreover, Morris et al. (2020a) argue that when parental socioeconomic factors are available, polygenic scores offer insignificant added value to explain educational attainment. I focus on the role of the parents' environment; i.e. mechanisms in which parental socioeconomic status affects child education, using the Health and Retirement Study. A century ago, people with a relatively high genetic potential (i.e. better education genes), are less likely to pursue that potential, regardless of nurture effects.

Up to 12 percent of the variance of attainted education can be predicted with polygenic scores for education (Lee et al., 2018). Some researchers, such as Plomin (2019), look at using a baby's DNA to personalize education as an opportunity to improve education and therefore polygenic scores become unquestionably valuable. Others, such as Koellinger & Harden (2018), Kong et al. (2018) and Morris et al. (2020b) disagree and find it unethical to use baby's DNA to better education and suggest that education genes do not have much predictive power anymore, when you take into account the parental environment. From the perspective of researchers, a better understanding of the role of parental environment when predicting educational attainment with polygenic scores is necessary to understand how valuable these polygenic scores might be. Even though the role of polygenic scores might still be undervalued within research, the prediction accuracy, both on the individual and group level, has become more precise, due to large-scale genome wide association studies (GWAS) (Lee et al., 2018). Furthermore, polygenic scores are valuable for policymakers, such as governments and corporations, to better understand the predictive power of education genes on the labor market and its outcome (Conley & Fletcher, 2018; Plomin, 2019).

The novelty of this paper and what is not yet clear in the literature, as far as my knowledge and research goes, is to what extent parental environment contributes to the predictive power of polygenic scores for education, when studying HRS; respondents born between 1905 and 1969. The research question that follows is:

"How well do polygenic scores for education explain educational attainment beyond parental education and socioeconomic status when investigating the Health and Retirement Study?"

To answer this question, the following sub-questions have been formulated:

1: "How much added value do polygenic scores offer in predicting educational attainment beyond parental education?"

2: "How much added value do polygenic scores offer in predicting educational attainment beyond parental socioeconomic status, measured as financial capital, social capital or human capital?"

3: "How much added value do polygenic scores offer in predicting educational attainment beyond parental socioeconomic status, measured as total capital?"

4: "How does the predictive power of polygenic scores on educational attainment change when considering different parental socioeconomic statuses?"

Financial, social, human and total capital will be elaborated on in the "Methodology" section. The aim is to understand how different parental influences are associated with the predictive power of education genes. After answering these sub-questions, the research question is answered and hopefully this further emphasizes the importance and understanding of polygenic scores in general, and more specific in the educational field.

The next part of this paper is methodology, where the data and methods are explained. After that, the results are shown. Finally, in the discussion, conclusions are drawn, validity and limitations of the research are discussed, and a follow-up research is suggested.

Methodology

I use three datasets, all conducted by the Health and Retirement Study (HRS). The Health and Retirement Study includes interviews with Americans in or near retirement. The data consists of thirteen survey years ranging from 1992 till 2016. The first dataset is called *RAND HRS Longitudinal file 2016* and is a simplified, combined and organized version of a Health and Retirement Study (RAND, 2019). The second dataset is called *Cross-Wave: Polygenic Score Data (PGS)*. This dataset consists of different sets of polygenic scores for individuals who participated in the HRS between 2006 and 2012. The individuals were asked to provide salivary DNA to the researchers. The polygenic scores are based on a GWAS. A score is the sum of the effects of over 1 million genetic variants in DNA and is optimized for European ancestry (Mills & Rahal, 2020). The third dataset I use is constructed by Vable et al. (2017) and contains three indexes for parental socioeconomic status (SES): financial capital, social capital and human capital. In total, N = 31,169 participants took part in the HRS. I used N = 10,274 eligible respondents, who were of European ancestry and provided genetic DNA. All datasets can be obtained via the Health and Retirement Study of the university of Michigan and are publicly available.

Parental environment of an individual is most commonly measured by parents' income, education and occupation (Hauser, 1994; Shavers, 2007; Winkleby et al., 1992). Ideally, an instrumental variable would be a promising measure to overcome any biases and find a causal relationship. However, to my knowledge, no suitable instrumental variable has been found. Another promising method would be a fraternal twin study. Twins growing up in the same parental environment, but with different genes, would give a more accurate prediction of educational attainment. Moreover, twin studies have shown that the twin with a higher polygenic score is more likely to attain higher education (Branigan et al., 2013; Daetwyler et al., 2008; Willoughby et al., 2019). This suggests that there are direct genetic effects of education genes that cannot be explained by family environment.

In this study I run a series of OLS regressions. Before running the regressions, the non-linear effects of the variables are removed. Nonlinearity occurs when there is no direct relationship between the response variable and the variables of interest, which is the case with educational attainment and polygenic scores. Furthermore, every OLS regression in this paper has some covariates (*Control_variables*) that always need to be controlled for: *birthyear*, *gender*, *birthyear*×*gender*, *birthyear*×*birthyear* and 10 *principal components* for European population. Those components adjust estimates for systematic differences between subpopulations due to differences in ancestry (Morris et al., 2020a). By convention, while analysing polygenic scores, principal components of the genetic relatedness matrix must always be taken into account (Price et al., 2010).

To answer the sub-questions, multiple OLS regressions are conducted. The OLS regression takes on the form of a multivariate regression and is as following:

$$Educational\_attainment = \beta_0 + \beta_1 Control\_variables + \beta_2 Polygenic\_score\_education + \beta_3 SES + \varepsilon_i$$

SES is determined as, (model 1) *Mother_education*, (2) *Father_education*, (3) *Human_capital*, (4) *Financial_capital*, (5) *Social_capital* or (6) *Total_capital*.

Educational_attainment, *Mother_education* and *Father_education* are measured as completed years of education. The *Human_capital* index of Vable et al. (2017) includes parents' completed years of education. The *Financial_capital* index of Vable et al. (2017) consists of average financial resources in childhood and financial instability in childhood. The following questions were asked to participants to assess *Financial_capital*: "Did your family move for financial reasons before you were 16 years old?", "Did your family receive financial help from relatives before you were 16 years old?", "Did your family declare bankruptcy?", "Did your family lose their business?", "How would you self-rate your childhood SES?", "What is your father's occupation?", "Has your father been unemployed for several months?" and "Does your mother work outside the house?". The *Social_capital* index of Vable et al. (2017) consists of maternal investment (quality of relationship with mother) and family structure before the age of 16 (number of household adults). To assess maternal investment the following questions were asked: "How much effort did your mother put into watching over you and making sure you had a good childhood?", "How much did your mother teach you about life?" and "How much time and attention did your mother give you when you need it?". To determine family structure the following questions were asked: "How many parents do you have?" and "Do you live with your father, mother, both or with grandparents, before you were 16 years old?". *Total_capital* contains the aforementioned three indices of parental socioeconomic status: *Human_capital* + *Financial_capital* + *Social_capital*.

Eventually, the predictive power of polygenic scores is determined as the incremental increase in variance explained (R^2) in educational attainment, when adding *Polygenic_score_education* to the regression. The correlation between the three indices of parental environment are shown in table 1.

Table 1: Correlation coefficients of the three parental socioeconomic status indices from Vable et al. (2017)

Variable		1	2	3
1	Financial capital	1		
2	Social capital	0.11	1	
3	Human capital	0.31	0.00	1

Finally, I stratify the regressions by SES to assess the utility of the polygenic scores on educational attainment in different SES groups: high, middle and low SES. I did this based on the average childhood SES index provided by Vable et. al (2017). The top 33% of this variable is marked as high SES, the middle 33% as middle SES and the bottom 33% as low SES. In this way, it is possible to answer how the predictive power of polygenic scores on educational attainment change when controlling for different childhood socioeconomic statuses, i.e. how much explanatory power is absorbed by these indices of parental socioeconomic status.

$$Educational\_attainment = \beta_0 + \beta_1 Polygenic\_score\_education + \beta_2 Control\_variables + \varepsilon_i$$

The above regression is conditional on SES, with SES (model 7) high, (8) middle or (9) low.

In the end, the research question how well polygenic scores explain educational attainment beyond parental education and socioeconomic status in the Health and Retirement Study, is answered. While adding more phenotypic data or parental information, predictions on educational attainment should improve, while the importance of education genes in the future can be further developed.

Results

The following descriptive statistics about the HRS are derived in table 2.

Table 2: Descriptive statistics of the variables of interest

Variable	Mean	Median	St. dev.	Min	Max
Educational attainment	13.16	12.00	2.56	0	17
Mother education	10.34	11.00	3.02	0	17
Father education	9.90	10.00	3.52	0	17
Birth year	1938	1938	10.48	1905	1969
Female	0.58		0.49		
N = 10,274					

Completed years of education of the respondent is on average approximately 13 years and their birth year is ranging from 1905 till 1969. Mothers completed on average more years of education, compared to fathers. 58 percent of the HRS consists of women.

To examine how much added value polygenic scores offer in predicting educational attainment beyond parental education, I run a series of OLS regressions (model 1-3). As said before, the control variables are always included. Consequently, parental education is added. At last, the incremental increase in R^2 explained by polygenic scores on completed years of education is derived in table 3.

Table 3: R^2 coefficients of educational attainment explained by polygenic scores for model 1-3

Variable	R^2 - model without PGS	R^2 - model with PGS	ΔR^2
1) Mother education	17.25%	21.84%	4.59%
2) Father education	17.44%	21.79%	4.35%
3) Human capital	20.43%	24.35%	3.92%
N = 10,274			

There is almost no difference in R^2 explained in educational attainment, in model 1 and 2 (R^2 = 21.84% and R^2 = 21.79%). The R^2 coefficient of model 3 (R^2 = 24.35%) is higher than the R^2 coefficient in model 1 and 2. Therefore, the ΔR^2 explained by polygenic scores on educational attainment is smaller when adding both parents' education, compared to only include one parent's education (ΔR^2 = 4.59% or ΔR^2 = 4.35% versus ΔR^2 = 3.92%). These results are visualised in figure 1.

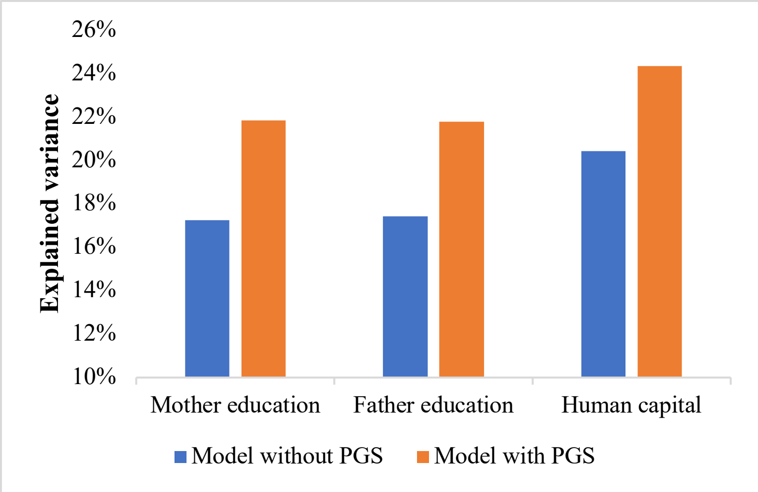

Figure 1: Explained variance of educational attainment, model 1-3

To investigate how much added value polygenic scores provide in predicting educational attainment beyond parental socioeconomic status, measured as financial capital, social capital or human capital, I run another series of OLS regressions (model 3-5). The results are shown in table 4.

Table 4: R^2 coefficients of educational attainment explained by polygenic scores for model 3-5

Variable	R^2 **- model without PGS**	R^2 **- model with PGS**	ΔR^2
3) Human capital	20.43%	24.35%	3.92%
4) Financial capital	8.78%	14.82%	6.04%
5) Social capital	5.05%	11.92%	6.87%
N = 10,272			

Model 3 explains most of R^2 in educational attainment, followed up by model 4 and model 5 has the least explanatory power of educational attainment (figure 2). Model 3 explains a lot more of the variance in educational attainment, compared to model 4 and 5 (R^2 = 24.35% versus R^2 = 14.82% and R^2 = 11.92%). This implies that financial and social capital absorb less of the predictive power of polygenic scores, relative to human capital.

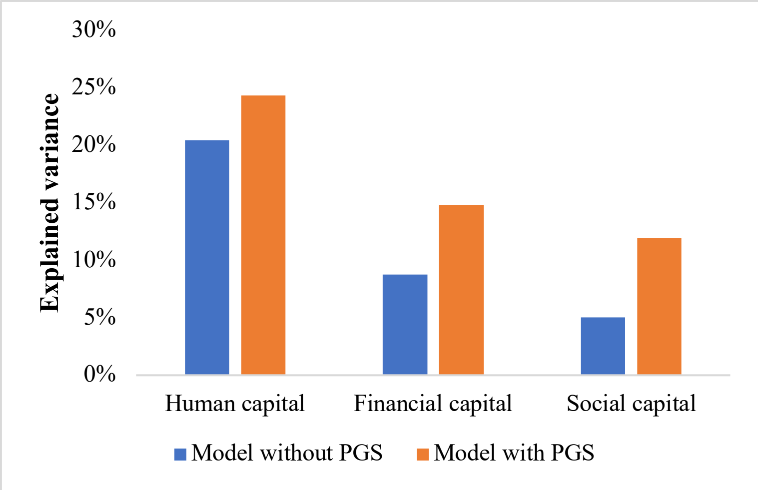

Figure 2: Explained variance of educational attainment, model 3-5

To examine how much added value polygenic scores provide in predicting educational attainment beyond parental socioeconomic status, measured as total capital, I run another OLS regression (model 6). I compare model 6 to model 3, as human capital predicts educational attainment better, relatively to financial and social capital.

Table 5: R^2 coefficients of educational attainment explained by polygenic scores for model 6

Variable	R^2 - model without PGS	R^2 - model with PGS	ΔR^2
3) Human capital	20.43%	24.35%	3.92%
6) Total capital	21.23%	24.99%	3.76%
N = 10,264			

Again, the results in table 5 imply that financial and social capital absorb less of the predictive power of polygenic scores, compared to human capital, as there is almost no difference in explained R^2 of educational attainment between model 3 and 6 (R^2 = 24.35% versus R^2 = 24.99%). Moreover, the ΔR^2 from both models are not very different (R^2 = 24.35% versus R^2 = 14.82%). This suggests correlation which is shown in table 6.

Table 6: Correlation coefficient between human capital and total capital

Variable		1	2
1	Human capital	1	
2	Total capital	0.70	1

To investigate how the predictive power of polygenic scores on educational attainment changes when considering different parental socioeconomic statuses, I run, once again, a series of OLS regressions (7-9). The results are shown in table 7.

Table 7: R^2 coefficients of educational attainment explained by polygenic scores for model 7-9

Variable	**R^2 - model without PGS**	**R^2 - model with PGS**	**Δ R^2**
7) High SES	2.59%	9.72%	7.13%
8) Middle SES	4.23%	9.36%	5.13%
9) Low SES	2.42%	7.40%	4.98%
N = 3,424			

Model 7 (R^2 = 9.72%) and model 8 (R^2 = 9.36%) have more predictive power in educational attainment, compared to model 9 (R^2 = 7.40%). This implies that for respondents in the category's high parental socioeconomic status and middle parental socioeconomic status, the model with polygenic scores included, explain more of the variance in educational attainment, relative to respondents in the low parental socioeconomic status category. The explained variance in model 8 without PGS (R^2 = 4.23%) is higher, compared to model 7 (R^2 = 2.59%) and 9 (R^2 = 2.42%). This indicates that the model without PGS and with people of middle parental SES, has more predictive power of educational attainment, compared to the model with people of high or low SES (Figure 3). Therefore, the added value of polygenic scores in explaining the variance of educational attainment, is the highest for the high parental socioeconomic group (ΔR^2 = 7.13% versus ΔR^2 = 5.13% and ΔR^2 = 4.98%), implying that control variables and respondent' background information, when considering people with high parental socioeconomic status, absorb less of the predictive power of polygenic scores.

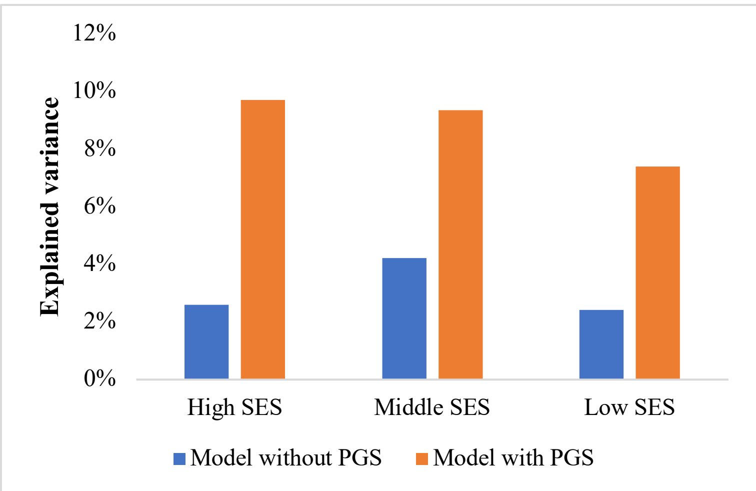

Figure 3: Explained variance of educational attainment, model 7-9

Discussion

The main question in this paper is *"How well do polygenic scores for education explain educational attainment beyond parental education and socioeconomic status when investigating the Health and Retirement Study?".* To answer this question, I examined four sub-questions.

The first sub-question was: "How much added value do polygenic scores offer in predicting educational attainment beyond parental education?". Polygenic scores provide extra predictive power of educational attainment beyond control variables and educational data, namely ΔR^2 = 4.59% on top of maternal education, ΔR^2 = 4.35% beyond paternal education and ΔR^2 = 3.92% above parental education. Morris et al. (2020a) measured an increase in proportion of the variance in educational attainment explained of approximately ΔR^2 = 0.90% and ΔR^2 = 1.90% beyond parental education. Comparing both studies suggests that, with an older generation as respondents (HRS), polygenic scores offer a bigger incremental increase in ΔR^2 explained of educational attainment beyond parental education, than for a younger generation; people born in the early 90's in the UK (ALSPAC study).

The second question was: "How much added value do polygenic scores offer in predicting educational attainment beyond parental socioeconomic status, measured as financial capital, social capital or human capital?". Here, polygenic scores for education provide an increase in explanatory power of educational attainment on top of control variables and parental socioeconomic data, namely ΔR^2 = 3.92% above human capital (i.e. parental education), ΔR^2 = 6.04% beyond financial capital and ΔR^2 = 6.87% on top of social capital. On the one hand, the lower ΔR^2 of human capital implies that human capital is a better measure of educational attainment, compared to financial or social capital. On the other hand, the lower ΔR^2 of human capital indicates that there is less added value of polygenic scores in predicting educational attainment when human capital data is available, relative to financial or social data that is available.

The third question was: "How much added value do polygenic scores offer in predicting educational attainment beyond parental socioeconomic status, measured as total capital?". I measure an increase in ΔR^2 explained of educational attainment above control variables and total capital of ΔR^2 = 3.76%, after adding polygenic scores to the regression. Total capital is a combination of all parental socioeconomic status measures. Beyond parental socioeconomic position, Morris et al. (2020a) calculated an increase of the proportion of ΔR^2 in educational attainment of approximately ΔR^2 = 0.80% and ΔR^2 = 1.80%, after adding polygenic scores. Considering both studies, polygenic scores within the HRS offer more predictive power in educational attainment above parental socioeconomic status, compared to the ALSPAC study.

The fourth question was: "How does the predictive power of polygenic scores on educational attainment change when considering different parental socioeconomic statuses?". The increase in explanatory power of educational attainment, when adding polygenic scores, beyond control variables, changes from $\Delta R^2 = 7.13\%$ for respondents with a high parental socioeconomic status, to $\Delta R^2 = 5.13\%$ for a middle parental SES and to $\Delta R^2 = 4.98\%$ for a low parental SES. This implies that when adding polygenic scores to models 1-6, the added value in explanatory power of educational attainment caused by polygenic scores, is different among various parental socioeconomic status groups.

Now the research question is answered. Education genes provide an increase in the proportion of R^2 of educational attainment explained of $\Delta R^2 = 3.92\%$, above parental education, $\Delta R^2 = 6.04\%$ above financial capital, $\Delta R^2 = 6.87\%$ above social capital and $\Delta R^2 = 3.76\%$ above total capital, which is in all cases more than the findings of Morris et al. (2020a). In line with Morris et al. (2020a), predictions from polygenic scores are inferior to parental environment factors. One reason for this might be the different measurement structure of parental socioeconomic status (Morris et al. (2020a) used the highest parental score on the Cambridge Social Stratification Score scale as measure for parental SES). Morris et al. (2020a) argue that when parental socioeconomic factors are available, polygenic scores offer insignificant added value to explain educational attainment. My results suggest that when parental education is available, other socioeconomic factors, such as financial or social capital, provide little extra value to explain completed years of education. In contrary to Morris et al (2020a), for people in the HRS, education genes contribute more in predicting educational attainment, beyond parental education and socioeconomic status, compared to the ALSPAC Study. A reason for this could be that parents in the ALSPAC study have more opportunities to apply their genetic potential in the current society, compared to parents in the HRS. Finally, when considering different various parental SES groups, there seem to be a variation in increase in explanatory power in educational attainment, caused by polygenic scores.

The Internal validity of this thesis can be disputed, because of measurement errors and I did not do a replication of the models on different samples. Therefore, I cannot say with certainty that the results are consistent. The external validity is hard to argue as I looked at a specific dataset: HRS. One limitation is recall bias. Whereas Morris et al. (2020a) could ask parental education directly to the parents, in my paper this was not possible. In the HRS, older participants had to recall their parent's educational attainment and home situation, leading to measurement errors. Also, my results are descriptive and not bootstrapped. In order to examine significant differences between models, all the R^2's in my paper need to be bootstrapped. Another limitation is the variable educational attainment, which is measured as completed years of education. This implies that a person who obtains the same degree as another individual, but took more years to complete it, has a higher educational attainment. Future research could investigate the possibility to improve this variable. For example, instead of using completed years of education, apply degree ranking of education. Furthermore, many variables that influence polygenic scores and educational attainment, such as parenting skills or help with schooling, are not taken into account. These unmeasured variables could be included in future research to improve internal validity of the research. Moreover, Morris et al (2020a) used attainment indicators (free school meals, English as foreign language and special educational needs) in their models, which can partly explain the difference in ΔR^2 measured, as these variables can also absorb part of the explanatory power of polygenic scores for education. Also, future research could look better into the interpretation of the change in predictive power of polygenic scores on educational attainment when considering different parental socioeconomic statuses, as there seem to be a variation. Finally, as I use limited resources to examine the effect of parental education and socioeconomic status, more family data could lead to better analyses.

Reference list

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., ... & Yang, J. (2019). Genetic correlates of social stratification in Great Britain. *Nature human behaviour*, *3*(12), 1332-1342.

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., ... & Gillespie, N. A. (2018). The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics*, *21*(2), 73-83.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces*, *92*(1), 109-140.

Chevalier, A. (2004). Parental education and child's education: A natural experiment.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American journal of sociology*, *100*(5), 1261-1293.

Conley, D., & Fletcher, J. (2018). *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press.

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS one*, *3*(10), e3395.

Harden, K. P., Domingue, B. W., Belsky, D. W., Boardman, J. D., Crosnoe, R., Malanchini, M., ... & Harris, K. M. (2020). Genetic associations with mathematics tracking and persistence in secondary school. *NPJ science of learning*, *5*(1), 1-8.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child development*, *65*(6), 1541-1545.

Janssens, A. C. J., Aulchenko, Y. S., Elefante, S., Borsboom, G. J., Steyerberg, E. W., & van Duijn, C. M. (2006). Predictive testing for complex diseases using multiple genes: fact or fiction?. *Genetics in medicine*, *8*(7), 395-400.

Koellinger, P. D., & Harden, K. P. (2018). Using nature to understand nurture. *Science*, *359*(6374), 386-387.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Gudbjartsson, D. F. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Fontana, M. A. (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.

Lee, M., Khan, M. M., & Wright, B. (2017). Is childhood socioeconomic status related to coronary heart disease? Evidence from the Health and Retirement Study (1992-2012). *Gerontology and Geriatric Medicine*, *3*, 2333721417696673.

Mills, M.C and Rahal, C., (2020). 'The GWAS Diversity Monitor Tracks diversity by disease in real time'. *Nature Genetics, 52,* 242-243. https://doi.org/10.1038/s41588-020-0580-y

Morris, T. T., Davies, N. M., & Smith, G. D. (2020a). Can education be personalised using pupils' genetic data?. *Elife*, *9*, e49962.

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020b). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*(16), eaay0328.

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

Papageorge, N. W., & Thom, K. (2018). Genes, education, and labor market outcomes: evidence from the health and retirement study. *Journal of the European Economic Association*.

Plomin, R. (2019). *Blueprint: How DNA makes us who we are*. Mit Press.

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 459-463.

RAND Social and Economic Well-being. (2019). *RAND HRS Longitudinal File 2016 (V1) Documentation*. Via: https://www.rand.org/content/dam/rand/www/external/labor/aging/dataprod/randhrs1992_2016v1.pdf

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.

Sewell, W. H., & Shah, V. P. (1967). Socioeconomic status, intelligence, and the attainment of higher education. *Sociology of education*, 1-23.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the national medical association*, *99*(9), 1013.

Vable, A. M., Gilsanz, P., Nguyen, T. T., Kawachi, I., & Glymour, M. M. (2017). Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the Health and Retirement Study. *PloS one*, *12*(10), e0185898.

Willoughby, E. A., McGue, M., Iacono, W. G., Rustichini, A., & Lee, J. J. (2019). The role of parental genotype in predicting offspring years of education: evidence for genetic nurture. *Molecular psychiatry*, 1-9.

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *American journal of public health*, *82*(6), 816-820.

Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet*, *6*(2), e1000864.