



**MAX PLANCK INSTITUTE**  
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · [www.demogr.mpg.de](http://www.demogr.mpg.de)

MPIDR Working Paper WP 2023-015 | April 2023  
<https://doi.org/10.4054/MPIDR-WP-2023-015>

## **Educational tracking and the polygenic prediction of education**

**Hannu Lahtinen**  
**Pekka Martikainen**  
**Kaarina Korhonen**  
**Tim Morris**  
**Mikko Myrskylä** | [office-myrskylä@demogr.mpg.de](mailto:office-myrskylä@demogr.mpg.de)

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

# Educational tracking and the polygenic prediction of education

## **Acknowledgements**

Special thanks for Aysu Okbay for providing education GWAS summary results excluding overlapping samples. We also thank the Finnish National Agency for Education for providing municipal-specific school-reform implementation years. The genetic samples used for the research were obtained from the THL Biobank (study number: THLBB2020\_8), and we thank all study participants for their generous participation in the THL Biobank.

## **Author names and affiliations**

Hannu Lahtinen, Population Research Unit, University of Helsinki

Pekka Martikainen, Population Research Unit, University of Helsinki, Max Planck Institute for Demographic Research, Rostock, Germany; Max Planck – University of Helsinki Center for Social Inequalities in Population Health, Helsinki, Finland

Kaarina Korhonen, Population Research Unit, University of Helsinki

Tim Morris, MRC Integrative Epidemiology Unit at the University of Bristol

Mikko Myrskylä, University of Helsinki, Helsinki, Finland; Max Planck Institute for Demographic Research, Rostock, Germany; and Max Planck – University of Helsinki Center for Social Inequalities in Population Health, Rostock, Germany and Helsinki, Finland

## **Funding**

HL was supported by the Academy of Finland (grant no #345219). MM was supported by the Strategic Research Council (SRC), FLUX consortium, ( #345130 and #345131); by grants to the Max Planck – University of Helsinki Center from the Jane and Aatos Erkko Foundation, the Max Planck Society, Faculty of Social Sciences at the University of Helsinki, and Cities of Helsinki, Vantaa and Espoo; and the European Union (ERC Synergy, BIOSFER, #101071773). PM was supported by the European Research Council under the European Union’s Horizon 2020 research and innovation program (#101019329), the Strategic Research Council (SRC) within the Academy of Finland grants for ACElife (#352543-352572) and LIFECON (#308247), and grants to the Max Planck – University of Helsinki Center from the Jane and Aatos Erkko Foundation, the Max Planck Society, University of Helsinki, and Cities of Helsinki, Vantaa and Espoo. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

## **Key words**

Educational tracking, educational attainment, polygenic score, gene–environment interaction, natural experiment

## **Corresponding author**

Hannu Lahtinen, University of Helsinki, Faculty of Social Sciences, Address: P.O. Box 42 (Unioninkatu 33) 00014 UNIVERSITY OF HELSINKI, Finland, phone: +358294124911, email: hannu.lahtinen@helsinki.fi

## Abstract

Although it is well known that individuals' genetics relate to their educational attainment, our understanding of how this may vary across differing educational institutional contexts is limited. In an educational system that does not separate students into different tracks early on, individuals' unique skills and interests may have more time to manifest, which could potentially strengthen the genetic prediction of education. We test such a hypothesis exploiting the natural experiment of the Finnish comprehensive school reform employed gradually and regionally across the country between 1972 and 1977, using genetically informed population-representative surveys linked to data from administrative registers. We observed that the genetic prediction of education was stronger after the reform by one-third among men and those coming from low-educated families. We observed no evidence for reform effects among women or those from high-educated families. The increase in genetic prediction was particularly pronounced among the first cohort experiencing the new system. From the perspective of genetic prediction, the reform to a more universalist curriculum was successful in promoting equality of opportunity. The results also highlight the potential of various turbulent circumstances – such as puberty or ongoing restructuring of institutional practices – in magnifying genetic effects.

## Introduction

The level of educational attainment is one of the most persistent social characteristics between generations. The correlation of years of education between parents and children is typically around 0.4 in Western Europe (Björklund and Jäntti 2020; Hertz et al. 2008). Twin decompositions indicate that the strong family-related effect stems from both biologically heritable and shared environmental components (Branigan, McCallum, and Freese 2013; Freese and Jao 2017; Silventoinen et al. 2020), and more recently developed molecular genetic methods have confirmed this (Lee et al. 2018; Okbay et al. 2022). However, for complex societal outcomes such as education, genes do not manifest in a vacuum but must necessarily operate within societal and institutional contexts, involving various social forces. Many behavioral geneticists have argued that the high heritability of education may act as an indicator of equality of educational opportunity (Ayorech et al. 2017; Conley 2016; Harden 2021; Plomin 2019; Selita and Kovas 2019; Silventoinen et al. 2020), since in such conditions, individuals have better opportunities to follow their personal talents and interests regardless of societal obstacles or privileges. Thus, it is crucial to understand in what kinds of educational contexts the genetic prediction of educational attainment is strong.

Previous research has shown that the amount of tracking in school curriculums, i.e., the extent to which students are separated in differing streams, particularly at younger ages, is robustly related to the importance of family socioeconomic background in educational attainment (Pekkarinen, Uusitalo, and Kerr 2009; Pfeffer 2008; van de Werfhorst 2018; van de Werfhorst and Mijs 2010). In early-tracked systems, family socioeconomic position tends to predict offspring's educational outcomes more strongly than later-tracked systems with more universalist curriculums. This is possibly since with later tracking, children and adolescents have more time to express their interests and abilities before being sorted into more specialized tracks. In early-tracked systems, in turn, family resources independent of the child's

characteristics play a relatively more pronounced role. In this study, we test whether an analogous pattern is also seen in stronger genetic prediction of education when the tracking is postponed. More specifically, we pose three research questions: 1) how did the universalist curriculum reform in Finland 1972–1977 modify the association between the known genetic propensity for educational attainment; 2) are the effects heterogeneous across individuals from different socioeconomic backgrounds or between men and women; and 3) does the possible effect vary according to the proximity of the reform.

We address these questions using data from genetically informed population-representative epidemiological surveys combined with data derived from administrative population registers. We utilize the Finnish Comprehensive School Reform, carried out between 1972 and 1977, as a natural experiment. The school reform abolished early tracking into academically and vocationally oriented schools at the age of 11 and replaced it with a system where students attend the same school until the age of 16. Due to exogenous sorting of individuals between the old and new systems, this design allows us to circumvent a common limitation of many previous analyses of gene–environment interactions, namely the non-random distribution of genomes across environments (Schmitz and Conley 2017). Moreover, the gradual enrollment of the reform across the country between 1972 and 1977 allows us to control for secular trends and regional differences potentially confounding the estimates on educational attainment.

## The interplay of genetics and the environment in education

Almost all human traits are to some degree heritable with a measurable genetic component (Polderman et al. 2015; Turkheimer 2000), and education is no exception to this (Branigan et al. 2013; Silventoinen et al. 2020). In molecular genetics, associations are tested between single nucleotide polymorphisms (SNPs), the most fundamental unit of variation in DNA, and a trait of interest. Complex behavioral and social traits are influenced by a huge number of SNPs (they

are “polygenic”), with thousands identified that relate to educational attainment (Okbay et al. 2022). Where multiple SNPs have been identified to associate with a trait, their estimated effect sizes can be summarized into an index variable called a polygenic score (PGS, see e.g. Choi, Mak, and O’Reilly 2020; Mills, Barban, and Tropf 2020; Mills and Tropf 2020). PGS consists of a sum SNP allele counts, weighted by their effect sizes in predicting the outcome, obtained from an independent genome-wide association study (GWAS). A PGS therefore provides an estimate of an individual’s genetic propensity for a trait using all SNPs that have been identified to associate with the trait. The relationship between an individual’s genome and their educational attainment is complex and operates through many pathways, such as psychological mechanisms including conscientiousness (Poropat 2009) and cognitive ability (Malanchini et al. 2020). SNPs associated with education have been identified to express predominantly in brain tissue and neurons in particular (Lee et al. 2018).

While there is strong evidence that genetics play a role in educational attainment outcomes, it is important to note that these genetic effects necessarily manifest into achieved education in a context of certain specific educational institutions shaped by social forces. Even though an individual’s as well as population’s DNA is essentially fixed, at least within conventional social-scientific time frames, changes in institutional circumstances may entail rapid changes in the way that genetic propensity for education manifests. Thus, the sociological study of understanding the interplay between such institutional conditions and individual genotypes is warranted, and presents an exciting opportunity to understand the formation and maintenance of educational inequalities. With PGSs, such gene–environment interactions can be modeled straightforwardly in regression frameworks familiar to many sociologists.

However, reliable identification of gene–environment interactions has turned out to be difficult in practice. Analysts must be cautious of endogeneity, as the same reservations that restrict causal inferences from regression main-effect coefficients apply to interactions. A spurious

gene–environment interaction can arise for various reasons. First, individuals tend to be distributed in different environments, such as childhood families, in a way that involves selection on genotypes. Second, individuals actively navigate toward and create environments that suit their genetic predispositions. Third, the environmental variable assessed in a gene–environment interaction analysis may not have a causal influence on the outcome, but merely correlate with the causal factor.<sup>1</sup>

Since the obstacles in drawing causal inferences for interactions are similar as for main effects, solutions can also be sought from the same direction. Exploiting natural experiments that exert exogenous effects on individuals, such as policy changes, constitutes one such strategy (Schmitz and Conley 2017). If individuals cannot self-select into old and new policy regimes, concerns of non-randomly distributed genotypes across categories of environmental exposure are reduced. This provides us a more robust understanding about the societal mechanisms via which PGSs, still largely black boxes, operate. In the current study, we employ one such natural experiment, namely the Finnish comprehensive school reform, which will be discussed in the next section.

### Description of the Finnish comprehensive school reform

Between 1972 and 1977, Finland conducted a comprehensive school reform where the old selective two-track educational system was replaced by a universalist one-track system. The suitability of the reform as a natural experiment has been extensively summarized before by Pekkarinen, Uusitalo and Pekkala Kerr (Pekkala Kerr, Pekkarinen, and Uusitalo 2013; Pekkarinen 2008; Pekkarinen et al. 2009), and this section draws from their work.

The pre-reform system started with a 4-year primary school (“kansakoulu”) for all children at the age of 6–7. At the age of 11, individuals chose either to stay in the primary school or enroll on a general secondary school (“oppikoulu”). Individuals who stayed in the primary school

until the age of 13 continued on to a 3-year civic school (“kansalaiskoulu”), which in turn, usually lasted for 5 years. In the late 1960s, roughly half of students took the general secondary track and half stayed in primary school (Pekkala Kerr et al. 2013; Sysiharju 1969). Follow-up education (if any) for those finishing civic school was usually in vocational schools. The majority of those finishing general secondary school continued to a three-year academic upper secondary school (“lukio” or “gymnasium”), which in turn, opened doors for university education (Sysiharju 1969).

In the post-reform system, by contrast, all students enrolled in a 9-year comprehensive school at the age of 6–7, and everyone followed the same curriculum until the age of 15–16 years. The new curriculum was academically oriented, and among the old-system alternatives, resembled the general secondary track more closely than the practically oriented civic school track. After the comprehensive school, students could continue either to academic upper secondary schools or vocational schools. Admission was based solely on comprehensive school grades (Pekkarinen 2008).

Before the school reform, private schools in secondary education (comprising 55% of the general secondary track students) charged tuition fees, although the majority of the expenses were covered by taxes (Pekkarinen et al. 2009). Tuition fees were abolished with the reform, and free-of-charge tuitions at all levels of education were implemented. In addition, the new system introduced greater special education resources for lower-performing students (Kivirauma and Ruoho 2007). Finally, it should be underlined that, unlike some comprehensive education reforms implemented in other countries during the same period, this reform did not increase mandatory schooling or minimum school leaving age<sup>2</sup>. The possible effects shall thus not be attributable to the amount of mandatory schooling but rather the qualitative differences between the old two-track and the new one-track systems.



The reform was implemented gradually between 1972 and 1977 in 6 different regions, as illustrated in Figure 1. The earliest adoption was in Lapland in Northern Finland and the last in the capital region of Helsinki (for a detailed presentation, see e.g. Pekkarinen et al. 2009). The municipality-specific implementation was planned by the National Board of Education and municipalities could not (at least officially) affect the schedule or the curricula in the new system (Aho, Pitkanen, and Sahlberg 2006). Students in grades 1–5 in the implementation year were transferred to the new system, as well as all subsequent cohorts.

INSERT FIGURE 1 HERE

### Possible consequences of the reform on the genetic prediction of education

Previous evidence implies that more universalist curriculums are associated with a smaller effect of socioeconomic background and greater meritocratic achievement in education (van de Werfhorst 2018; van de Werfhorst and Mijs 2010). Pekkarinen and colleagues (2009) observed that the Finnish comprehensive school reform increased intergenerational income mobility between sons and fathers, and Valkonen et al. (1996, 1998) found suggestive support for a decrease in social class background differences in educational attainment. The mechanism driving the equalizing effect of a universalist curriculum may lie in that students have more time to show their individual talents regardless of the resources of their families. Following such reasoning, we may expect that early tracking mitigates the realization of one's genetic potential in education. Some twin studies have found corresponding associations between delayed tracking and higher heritability (within-population variance that is attributable to genetic differences between individuals) of education (Baier et al. 2022; Knigge et al. 2022; for

a null result, see Mönkediek 2022), as well as higher heritability after some other equality-increasing school reforms (Colodro-Conde et al. 2015; Heath et al. 1985).

The reform-induced change in genetic prediction of education may be heterogeneous by socioeconomic background. Despite that the comprehensive school reform affected every student (in contrast to, for example, school reforms that increase mandatory schooling years, which do not directly affect those who would continue education in any case), the change was arguably larger for those who would have continued in the civic school track in the pre-reform system. These individuals were more likely to come from socioeconomically less advantaged families. In the 1960s, 12% of the general track graduates' principal providers had academic upper secondary ("lukio") or university education, compared to 4% of the overall adult population (Sysiharju 1969). Thus, the reform may have a stronger effect on genetic prediction of education among those with disadvantaged socioeconomic background. Pekkala Kerr et al. (2013) observed that the Finnish comprehensive school reform increased cognitive test scores, especially for individuals with low-educated parents. Such an increase in cognitive performance also provides a further candidate mechanism for the hypothesis of socioeconomically heterogeneous effects. Other potential mechanisms include decreasing economic costs of education, more homogenous peer groups and better prospects for future education, which were likely to affect individuals from less socioeconomically advantaged origins disproportionately.

Heterogeneous effects in opposing directions by socioeconomic family background are also possible. The performance of students who would have enrolled in the general secondary track in the old system may have become hindered by the inclusion of less academically oriented peers in the classroom (Lazear 2001), or other loss of quality in the previously positively selected learning environments. Since the individuals from more advanced socioeconomic backgrounds (who were more likely to continue on the general secondary track in the old

system) were exposed to less positive learning environments in the new system, it is possible that the realization of their genetic propensity of education was weakened.

Furthermore, there has been a discussion of a Scarr–Rowe interaction hypothesis in education. Such a hypothesis, originating from studies of socioeconomic differences in cognitive ability (Rowe, Jacobson, and Van den Oord 1999; Scarr-Salapatek 1971; Tucker-Drob and Bates 2016), states that individuals from more advantaged socioeconomic backgrounds have better possibilities to follow their motivations and abilities, while those with disadvantaged backgrounds suffer more from environmental obstacles. Applied in twin studies of education, this implies the biologic heritability of education should be higher among those with socioeconomically advantaged backgrounds and the environmental component higher among individuals with disadvantaged backgrounds. The evidence on such a pattern in education has been mixed overall, with possible contextual differences (Baier et al. 2022; Baier and Lang 2019; Erola et al. 2021; Lin 2020). Relatedly, Trejo and colleagues (2018) found in the U.S. that PGS was more predictive of educational attainment among higher than lower status high schools among earlier birth cohorts, but this interaction disappeared among later cohorts, possibly due to improving overall opportunities for further education. However, a contrasting “Saunders”<sup>3</sup> hypothesis has been confirmed in some previous studies, where the genetic effects are weaker among those with an advantaged socioeconomic background (Baier et al. 2022; Lin 2020). Such a hypothesis states that genetic effects are more pronounced in challenging or uncertain environments (Baier et al. 2022; Saunders 2010). Relatedly, Harden et al. (2020) found that students with low PGSs from socioeconomically advantaged schools were less likely to drop out of math courses than similar students from disadvantaged schools<sup>4</sup>. This implies that school advantage may offer protection from a failure in class, which is a mechanism that can also lead to lower predictions of PGS among students with socioeconomically more advantaged backgrounds in the old Finnish two-track system. A somewhat analogous pattern

has been a relatively common finding in gene–environment interaction analyses in epidemiological and psychopathological literature, where this is often called a “diathesis-stress” model of gene–environment interactions (Boardman, Daw, and Freese 2013; Manuck and McCaffery 2014). For example, in the epidemiology of alcohol and other substance use, it has been found that environments with less social control or more limited personal resources strengthen the genetic effects of substance consumption and related harm (Neale et al. 2021; Pasma, Verweij, and Vink 2019; Van Der Zwaluw and Engels 2009; Young-Wolff, Enoch, and Prescott 2011).

In addition to socioeconomic background, gender-specific heterogeneous reform effects may be expected as well. A slight majority of the students in the old general secondary track were female, for example, 57% in 1964/1965 (Sysiharju 1969). Thus, the reform possibly brought greater change for curriculums of boys on average. Furthermore, in the new system, the finishing grade point average, on which the qualification for further education was solely based, was determined at the age of 15–16. At this age, boys typically experience the most intense period of puberty, whereas girls have passed peak puberty (Euling et al. 2008). This may have implications to our study. Following reasoning from the Saunders/diathesis-stress hypotheses presented above that genetic effects should be magnified in uncertain and challenging environments, puberty constitutes a candidate for a factor of such a condition, implying that the reform would increase the genetic prediction among men in particular.

On the other hand, there are grounds to expect a stronger reform effect on polygenic prediction of education among women. It is also possible to make a contrary hypothesis that, instead of strengthening genetic effects, tangible but passing phases such as puberty may also add noise which temporarily obscures signals coming from the (permanent) genome of individuals. In addition, (despite already having been the majority in the old general track) there may have been societal obstacles hindering the educational achievement of women which have gradually

diminished over time. Equalizing school reforms may have played a part in such an ease of obstacles. According to previous evidence, the Finnish comprehensive school reform increased the achieved educational level of women but not men (Pekkarinen 2008). Corresponding observations have been made with regards to genetic prediction across birth cohorts, which has strengthened among women (Herd et al. 2019).

Finally, it is also possible that the effect of the reform was dependent on the time-related proximity of its implementation. The oldest cohort transferred to the new system attended comprehensive school for five years, the second cohort for six years and so on. Thus, it may be argued that the reform intensity was lower among the oldest birth cohorts transferred to the new system. In line with such reasoning, Ollikainen (2021) observed that stronger reform intensity (i.e., later birth cohorts) was negatively associated with early career employment, and mixed results have been obtained with regards to intensity and final education (Ollikainen 2021; Pekkarinen 2008). However, opposite effects may also be expected. It is important to note that even the first cohort exposed in the reform already had the two-track system fully abolished in their secondary educational paths. Thus, if the mechanism producing change in genetic prediction is precisely in the tracking system and later sorting of students, the “reform intensity” should not matter much. Moreover, the earliest cohorts had also a unique position in the sense that they were the first to experience new practices, which had to be established and possibly were still seeking their final form during the pioneering school cohorts. In line with the Saunders and diathesis-stress hypotheses discussed in this section, such instability may lead the genetic effects to manifest more strongly precisely at the dawn of the new educational regime.

### Educational stratification in Finland

The association between parental and offspring education (Hertz et al. 2008; Pfeffer 2008), as well as that between siblings (Grätz et al. 2021), in Finland is among the smallest in

international comparisons. This reflects the objectives of the universalist welfare state well, where the reduction of educational differentials by social background has been explicitly pursued (Kalalahti and Varjo 2020). The cohorts born in the early 1960s experienced a period of strong educational expansion; the average level of education rose strongly, particularly among women, whose educational level exceeded that of men among the studied cohorts (Härkönen and Sirniö 2020; Pekkarinen 2012). Furthermore, the birth cohorts around the reform years have possibly even historically low intergenerational educational reproduction, with a Spearman's rank-order parent-child correlation of education around 0.25 (Härkönen and Sirniö 2020; Karhunen and Uusitalo 2017). Part of the weak intergenerational association may possibly be attributed specifically to the comprehensive school reform addressed in this study. The effect of the reform was previously shown in increased income mobility between generations and increased cognitive test scores among those with low socioeconomic backgrounds, as reviewed above (Pekkala Kerr et al. 2013; Pekkarinen et al. 2009). However, for the cohorts born from the 1960s onward, the association between parental and offspring education has strengthened again (Härkönen and Sirniö 2020; Karhunen and Uusitalo 2017; Lahtinen, Martikainen, and Tarkiainen 2022), although a similar increase has not been observed in sibling similarity in education (Lahtinen et al. 2022). Earlier studies mostly covering cohorts born before the 1960s, in turn, have indicated a weakening or stabilizing intergenerational association over time (Hertz et al. 2008; Kivinen, Ahola, and Hedman 2001; Pfeffer 2008).

## Data & methods

### *Data and variables*

The sample used in this study consists of population-based FINRISK survey rounds 1992 (n: 6,024), 1997 (n: 8,387), 2002 (n: 8,775), 2007 (n: 6,216) and 2012 (n: 5,748), collected every 5 years in North Karelia, Northern Savonia, Lapland, Northern Ostrobothnia, Kainuu, Turku and Loimaa, as well as Helsinki and Vantaa. These data are pooled together with data from the

Health2000 survey (n: 6,693, with replacement n: 762 in 2011), which covered the whole of mainland Finland. These data have been evaluated to represent the population of their target age well in the areas studied (Aromaa and Koskinen 2004; Borodulin et al. 2018). These studies included clinical examinations, during which DNA samples were collected. The response rates vary between 65% and 93%, with an overall average of 73%. Genotype information was available for 88% of respondents. The genetic data was put through quality control and imputation procedures according to SISU v3 reference panel protocols (Pärn et al. 2018, 2019). The genotyped data were linked to population registers including annual information on municipality of residence in 1972–1977 when the reform took place as well as the highest level of education in the household in 1970 and 1975, and the personal educational degrees obtained by the end of 2019.

We restricted the analysis to genotyped participants in cohorts  $\pm 4$  years around the reform in their municipality of residence, including individuals from cohorts born between 1957 and 1969 and whose municipality of residence in 1972 could be determined in mainland Finland. This gave us an initial sample of 6,159 individuals. We excluded those who moved between municipalities with different school reform implementation years between 1972 and 1977 (n: 78). Finally, to mitigate possible bias between PGS and outcome arising from shared environment between related individuals, sometimes titled “cryptic relatedness” bias (Choi et al. 2020; Mills et al. 2020), we excluded one individual from pairs sharing more than 0.177 of the variation of their genome, which is the expected lower bound for second-degree relatives (n: 135 excluded). These exclusions gave us the final sample size of 5,946 individuals.

We measured individuals’ educational achievement as the years of education based on the expected number of years of schooling after basic level to obtain the highest degree that an individual has. Following the guidelines of the International Standard Classification of Education 2011 (UNESCO 2012), these were the following: 1) No qualification beyond basic

education (ISCED 0–2): 0 years; 2) Upper-secondary education (ISCED 3): 3 years; 3) Specialized vocational qualification (ISCED 4): 4 years; 4) Lowest tertiary degree (ISCED 5): 5 years; 5) Bachelor’s, or equivalent, including basic polytechnic degree (ISCED 6): 6 years; 6) Master’s level or equivalent (ISCED 7): 8 years; and 7) Doctoral or licentiate degree (ISCED 8): 12 years.

Figure 2 illustrates included birth cohorts with their respective number of observations in each reform region. The exposure to the school reform was an indicator variable defined on the basis of the implementation year of the municipality of residence and year of birth. Students were transferred to the new system if they were in the fifth grade (i.e., they turned 11 in the year) during the year of reform implementation. This means that an individual was assigned as exposed to the reform if the difference between the reform implementation in the municipality of residence and the year of birth was at least 11 (e.g., those born in 1961 or later were assigned “exposed” if the reform took place in 1972 in their municipality of residence). We also performed analyses with the exposure variable in three categories: 1) pre-reform cohorts (4 cohorts in each region); 2) the first cohort in the new system (1 cohort in each region); and 3) subsequent reform cohorts (3 cohorts in each region).

INSERT FIGURE 2 HERE

The PGS of education was defined according to the GWAS by Okbay et al. (2022). To avoid overfitting in the genetic data, all individuals in our study that overlapped in the GWAS analysis were removed from the summary statistics. In addition, those individuals in the GWAS obtained from 23andme were excluded due to privacy policies of the company. A PGS provides a summary measure of the known genetic propensity for a trait, multiplying the effect size (obtained from GWAS summary statistics) of a given copy of a SNP by the number of copies



that an individual has (0, 1 or 2). We employed the SBayesR method (Lloyd-Jones et al. 2019) to adjust the GWAS scores for linkage disequilibrium (LD, that is, correlation between occurrences of SNPs in different loci), using the base GWAS summary results by Okbay et al. (2022), as well as a banded LD matrix provided by the authors of the GCTB software which was used in implementing the SBayesR Monte Carlo Markov Chain simulations (Zeng et al. 2020). These re-weighted scores were then assigned to the individuals using autosomal HapMap3 variants with a minor allele frequency of at least 0.01 and not strongly deviating from Hardy–Weinberg equilibrium ( $p > 10^{-8}$ ) in our data. The PGS is standardized to have mean 0 and standard deviation (SD) 1.

The Finnish population is generally of European ancestry and genetically homogenous, save a difference between the Eastern and Western regions, following closely the border set in the historical treaty of Pähkinäsaari in 1323 (Kerminen et al. 2017). To account for population stratification by genetic ancestry, we adjusted our models for the first ten principal components of the genome (Price et al. 2006). The reform regions roughly correspond to this division, with early reform regions mainly consisting of areas in eastern and later reform regions of the western genetic sphere. Our region-specific analyses thus further control for such population stratification.

To account for the subtle bias that may arise from, for example, differing genotyping equipment or differences in participant recruitment between data collections, our models were adjusted for genotyping batch and survey round. We also measured family education on the basis of the highest education in the household in 1970, and if missing, in 1975, in two categories: all members of the household had basic education / at least one member had education beyond basic education.

*Analysis strategy*

Our main parameter of interest was the difference-in-difference estimate on whether the effect of the PGS on achieved education was different before and after the reform. For these purposes, we performed estimation in two stages.

First, we fitted reform region-specific ordinary least squares regressions to estimate the interaction between PGS and reform status through the following model:

$$edu_i = b_0 + b_1 PGS_i + b_2 Reform_i + b_3 (PGS * Reform)_i + \overline{b_4}' \overline{Controls}_i + \epsilon_i \quad (1),$$

where  $edu_i$  denotes years of education for individual  $i$ ;  $b_0$  is the intercept,  $PGS$  is the polygenic score of education;  $Reform$  is an indicator of being exposed to the school reform;  $\overline{Controls}$  is a vector of control variables. They include continuous year of birth, a female indicator, the first ten principal components of the genome, data collection round indicators and genotyping batch indicators.  $\epsilon$  is the residual term. We used Huber–White standard errors to account for potential heteroscedasticity of residuals (Mansournia et al. 2021).

Second, we combined these separate reform region-specific estimates utilizing the fixed-effect inverse variance-weighted meta-analysis method (Borenstein et al. 2009:11). We conducted meta-analysis on the pre-reform effect parameter ( $b_1$  of equation 1), post-reform effect parameter ( $b_1+b_3$  of equation 1)<sup>5</sup>, as well as the difference-in-difference estimate, which is the interaction parameter  $b_3$ . Calculating a meta-analytic difference-in-difference by subtracting the meta-analytic post-reform estimate from pre-reform estimate creates slightly different results in some cases. These differences are attributable to the heteroscedasticity correction of the standard errors, as it affects the given weight of the parameters.

The rationale behind resorting to this two-step modeling strategy stems from the fact that the exposure to the reform was defined on the basis of the combination of year of birth and reform region. Thus, including both in the model would create very high (or even full if they are measured categorically and interacted) multi-collinearity with the reform variable. This would

also create highly complex models with hard-to-interpret parameters sensitive to all lower-order interactions. Estimating within-region regression also controls for regional differences in education in a flexible manner, including potential interactive patterns in addition to main effects.

After estimating the effect for the whole population through meta-analyses, we assessed heterogeneous effects by conducting identical estimation stratified by educational family background (basic education/at least one household member with a further degree during adolescence) and gender (men/women, excluding the gender indicator from regression). Finally, we also performed these analyses with an alternate formulation of the reform status including three categories (unexposed/first cohort of exposure/later cohorts of exposure).

All statistical tests reported in this study are two-tailed.

INSERT TABLE 1 HERE

Table 1 presents descriptive statistics of the variables used. The PGS of education and the achieved years of education had a correlation of 0.29. The average PGS was slightly lower in regions with earlier adoption of the reform (also, the average level of education was slightly lower in these regions). This might indicate genetic population stratification effects across regions, as the reform roughly moved from north to south and from east to west. This gives further warrant to the strategy of estimating the regressions separately by region, as it implicitly adjusts for such differences. The lower number of observations in early and late birth cohorts stems from the fact that they are only included in the analysis of some regions, for example, the 1957 cohort is only included in the analysis of the region where the reform year was 1972 and cohort 1969 where the reform year was 1977 (see also Figure 2).

## Results

INSERT TABLE 2 HERE

Table 2 presents the association between the PGS of education and years of education before and after the reform, obtained from meta-analyses of reform region-specific regressions presented in appendix Tables A1 and A2. Among the whole study sample, a SD higher PGS predicted a 0.59 (95% confidence interval [CI] 0.51; 0.67,  $p=10^{-48}$ ) increase in the achieved years of education before the reform and 0.68 (95% CI 0.61; 0.76,  $p=3*10^{-61}$ ) years after it. Thus, the reform brought a small increase in coefficients which did not reach any conventional level of statistical significance in a two-tailed test (difference-in-difference [DiD] estimate = 0.08, 95% CI -0.03; 0.19,  $p=0.167$ ).

When investigating this change among individuals with differing levels of educational family background or gender, we observed evidence for an increase in genetic prediction of education among those whose family members had only basic education (DiD=0.16, 95% CI 0.02; 0.31,  $p=0.029$ ) and among men (DiD=0.17, 95% CI 0.01; 0.34,  $p=0.041$ ). We did not observe changes in the predictive power of the PGS among those with higher than basic level educational background or among women. Relatedly, we observed stronger genetic prediction of education among individuals with higher than basic family education compared to those with low family education before the reform (difference in pre-reform estimates between educational backgrounds was 0.22, 95% CI 0.06; 0.38,  $p=0.007$ ), but not after the reform ( $b=0.04$ , 95% CI -0.11; 0.20,  $p=0.581$ )<sup>6</sup>. Overall, after the reform PGS predicted education in a more homogenous manner across all the studied subgroups.

INSERT TABLE 3 HERE

Table 3 presents the effects of the comprehensive school reform via an alternate measurement of the reform, namely in three categories: cohorts before the reform (pre-reform cohorts), the first cohort that was exposed to the reform (1st reform cohort), and all subsequent cohorts. These are again based on meta-analysis reform region-specific regressions presented in Tables A3, A4 and A5 in the appendix. The PGS of education was particularly predictive of years of education among the cohort that was the first to experience the school reform, among whom a one SD higher PGS predicted 0.79 (95% CI 0.64; 0.94,  $p=2*10^{-21}$ ) additional years of education. The corresponding figures were lower among cohorts more than one year after the reform (0.65, 95% CI 0.56; 0.74,  $p=10^{-48}$ ) as well as cohorts before the reform (0.59, 95% CI 0.51; 0.67,  $p=3*10^{-42}$ ). The DiD of the effect of PGS between the first reform cohort and pre-reform cohorts was 0.17 (95% CI 0.00; 0.33,  $p=0.051$ ).

Likewise in the previous analyses, the DiD results presented in Table 3 again indicate a substantial reform effect for the first cohort among those with only basic parental education (0.34, 95% CI 0.10; 0.58,  $p=0.006$ ) and among men (0.33, 95% CI 0.05; 0.60,  $p=0.019$ ), while no evidence for reform effect are observed among those whose parents had more than basic education or among women.

To test the robustness of these results, Figure A6 in the appendix presents an additional analysis where the PGS is interacted with a categorical year-to-reform variable (ranging between -4 and +3, 0 denoting reform year) in predicting the years of education. The results of this analysis were consistent with our earlier results. We see higher coefficients among those with only basic parental education and among men after the reform. The largest coefficient for the first reform cohort was clearly visible. Rather peculiarly, two cohorts before the reform were positive

outliers among those with more than basic parental education and among women. However, the data gets rather thinly stretched, and there is substantial uncertainty in these point estimates, as evidenced by their wide 95% CIs. Thus, these results should be interpreted with caution.

## Discussion

In this study, we investigated whether the comprehensive school reform, implemented regionally between 1972 and 1977 in Finland and replacing the old two-track basic educational system with a universal curriculum for all students, affected the polygenic association of educational attainment. We found a modest and statistically non-significant increase in the strength of the association between education PGS and achieved education among the cohorts exposed to the reform. When investigating this effect more closely in population subgroups, we found that this modest overall effect was driven by more substantial effects among two subgroups, namely men and those whose family members had no more than basic education. Thus, our results provide partial support for the results of a recent twin study from the Netherlands (Knigge et al. 2022) which indicated greater heritability of education when there is less tracking. It also resonates with results of Baier et al. (2022) who observed that Germany had comparatively low heritability of education, attributing it to the strong tracking in the German educational system.<sup>7</sup> Our study exploits a design that allows for stronger causal interpretation due to the virtue of a natural experiment of school reform implemented gradually across the country. Furthermore, our strategy of using directly measured genetic variation has allowed for triangulating the previous evidence from alternative (twin) methods with orthogonal limitations.

We observed that the reform increased genetic prediction among adolescents whose family members had no education beyond basic level, whereas among those from higher educated families the effect was close to nil. Correspondingly, we observed a Scarr–Rowe interaction effect, namely stronger genetic prediction among individuals with more advantaged

socioeconomic position, before but not after the reform. This suggests that universalism in educational curricula may be one mechanism that explains the context-specific emergence of the Scarr–Rowe interaction (Baier et al. 2022; Baier and Lang 2019). For example, it is consistent with that perhaps the strongest pattern of a Scarr–Rowe interaction in educational attainment has been found in Germany (Baier and Lang 2019), a context known for its high level of curriculum tracking. It may further provide (a partial) explanation on previous twin results in Finland finding lower heritability of education in earlier than later birth cohorts (Erola et al. 2021; Silventoinen et al. 2004, 2020).

We may interpret both of these family background–related observations – an increase in the genetic prediction of education among those with low-educated families as well as the simultaneous disappearance of the Scarr–Rowe interaction – in terms of equality of opportunity. Many behavioral geneticists have advanced a thesis that higher genetic heritability of education is an indicator of equality of opportunity (Ayorech et al. 2017; Conley 2016; Harden 2021; Plomin 2019; Selita and Kovas 2019; Silventoinen et al. 2020). Correspondingly, the presence of the Scarr–Rowe interaction has been interpreted to show that those with less socioeconomically advantageous origins may have more environmental obstacles in realizing their genetic potential. From both of these perspectives, we may claim that the reform was successful in improving the equality of educational opportunity in Finland, which was also an explicit political goal of the reform (Aho et al. 2006; Kalalahti and Varjo 2020), although with a caveat that the greatest increase in genetic prediction may have been short term. The improvement in equality in educational achievement fits previous evidence on the increase in intergenerational income mobility (Pekkarinen et al. 2009) and cognitive test scores among those with less advantaged socioeconomic background (Pekkala Kerr et al. 2013) as a result of this reform.

The opponents of the reform raised concerns that the universal curriculum would also have an undesirable equalizing effect (Okkonen 2017). The worries were that the most talented individuals could not obtain an optimally challenging education and would be disturbed with ill-behaving peers in the classroom (Lazear 2001). Although we do not have a direct measure of the performance of the students, we may evaluate the argument indirectly through the performance of those students with higher family education who were likely to continue in the general secondary track in the older system (Statistics Finland 2007). In line with a previous study assessing the effect of school reform on cognitive skills (Pekkala Kerr et al. 2013), our results imply steady genetic prediction of those with more advantaged backgrounds. Thus, we did not observe evidence of the manifestation of these worries. Although the point estimate was negative, the size was very small and indistinguishable from stochastic noise stemming from a finite sample size.

Since children inherit their genomes from their parents, genetic pathways produce a straightforward mechanism in the intergenerational transmission of traits. Thus, there is some tension between the increase in socioeconomic mobility after the reduction of tracking observed in previous studies (Pekkarinen et al. 2009; van de Werfhorst 2018; van de Werfhorst and Mijs 2010) and a simultaneous increase in the genetic prediction of education observed here. High genetic heritability and low intergenerational persistence of education have both been considered as indicators of open societies and equality of opportunity. However, Engzell and Tropf (2019) indeed found comparative evidence for an inverse relationship between the genetic heritability of education and the intergenerational educational correlation, suggesting that they are not mutually exclusive views of equality of opportunity. Our results further suggest that educational policies that reduce the intergenerational correlation of educational attainment have the potential to simultaneously strengthen the genetic prediction of education in a population.



We observed that the reform strengthened the genetic prediction of education among men but not women. In a previous study, Pekkarinen (2008) observed that the comprehensive school reform slightly increased the educational success among women but not men, and offered a possible reason for gender divergent findings in puberty. In the new system, the age when grades and decisions affecting future educational paths coincided with the most difficult period of puberty among boys, whereas among adolescent girls, this phase was already starting to pass. Overall, the psychological regression at puberty is found to be stronger on average among adolescent boys than girls (Aalberg and Siimes 2007:74–82), and previous evidence indicates that school performance more specifically varies more strongly with the puberty timing among boys than girls (Koivusilta and Rimpelä 2004). Additionally in our case, puberty, as a passing but tangible phase that may bring turbulence in educational success and decision making, might resemble such unstable or capricious circumstances where the effect of genetic endowments becomes more pronounced, as outlined by the Saunders or diathesis-stress models of gene–environment interactions (Baier et al. 2022; Boardman et al. 2013; Saunders 2010).

When inspecting the reform effects more closely with respect to its timing, we observed that the genetic prediction was particularly strong for the first cohort experiencing the reform. Again, this interaction was driven by the population subgroups of men and those with lower-educated families. It seems that the turbulence around the reform, with lots of practices still seeking their established form, resulted in a stronger differentiation on individuals by their genetic tendencies. Similar to the gender-specific results discussed in this section, this is another pattern roughly analogous to the Saunders or diathesis-stress interaction models.

### Methodological considerations

Strengths of the study include population-representative data with high response rates for genetically informed datasets and register-based measurements which can avoid biases or noise stemming from self-reported data, including social desirability, subjectiveness, or faulty recall.

Employing exogenous variation provided by a natural experiment helped to minimize biases stemming from non-random allocation of genotypes across environments, which is a common limitation in sociogenomic research (Morris, Davies, Hemani, et al. 2020). The gradual implementation of the school reform across the country allowed us to control for both possible secular trends and regional differences.

Despite using exogenous environmental variation, it should be noted that not only the environmental but also the genetic component in gene–environment interactions may be confounded. This is because PGSs may themselves capture environmental or social factors which are strongly but spuriously correlated with genetic variants (Kong et al. 2018; Morris, Davies, Hemani, et al. 2020). In our case, such influences can arise, for example, if those with higher PGSs also had more beneficial family environments independent of their inherited DNA variants (“genetic nurture effect”, Kong et al. 2018; Wang et al. 2021). Previous work has demonstrated that such environmental biases are larger for educational attainment than many other traits (Howe et al. 2022), and therefore it is possible that there remains some bias in the genetic component of our study. This limitation could be addressed using within-sibship analysis designs, but such analyses have very demanding data requirements, and so far, sibling samples of sufficient numbers for these specific cohorts have not been available.

Limitations of the study also include somewhat compromised statistical power arising from the size of the analysis sample. Although all of our main effects were estimated with high statistical certainty (p-values were smaller than  $10^{-10}$  for all main effect parameters presented in Tables 2 and 3), robust discovery of an interaction requires much larger sample than for a corresponding main effect (Gelman, Hill, and Vehtari 2020:16). Additionally, despite ever increasing GWAS sample sizes, the PGSs still contain noise and account for only part of the total heritability of traits. Furthermore, GWAS summary statistics are usually meta-analyses of a large number of cohorts in differing environmental circumstances. Thus, they are likely to capture SNP effects

that manifest relatively uniformly across environments, which can make them conservative when analyzing gene–environment interactions (Conley 2016).

Factors stemming from the implementation of the reform can bias our results to the conservative direction. First, as a compromise to the political struggle involved in the reform, ability groups for students were retained in foreign languages and mathematics. Attending the lowest group closed the doors for further education (Kalalahti and Varjo 2020; Kauko 2019). The ability groups were only abolished in 1985 (Pekkala Kerr et al. 2013). Second, teachers were only gradually educated focusing specifically on the new system (Aho et al. 2006:49–51). Third, there were pilot schools that were exposed to the reform earlier than other schools in some municipalities which we could not identify from our data. Relatedly, it is also possible to delay (or in very rare cases, advance) the school start year of children, and also grade repeating at primary school grade could bring noise to your estimates, but neither was common during the period studied. In the 1960s, yearly 1.3–1.5% of the students in Helsinki had to repeat a grade in primary school (Oinonen 1969:116). Among the study sample of Oinonen (1969:106–9), consisting of (expected) school entry cohorts 1963 and 1964 in a large school district, 13 out of 430 students did not start their school together with their expected cohort.

## Conclusion

Exploiting the natural experiment of the Finnish comprehensive school reform 1972–1977, we have provided evidence that individuals’ genetic predispositions and institutional features such as curricula can be more than a sum of their parts in shaping the educational success of individuals. A decrease in tracking in the schooling system led to stronger genetic prediction of educational achievement, particularly among those with less socioeconomically advantaged family backgrounds. This result suggests that the reform increased equality of educational opportunity through the convergence in the realization of the genetic potential of individuals.

Thus, in line with previous studies (Pekkala Kerr et al. 2013; Pekkarinen 2008; Pekkarinen et al. 2009; Valkonen et al. 1996, 1998), the reform was successful in achieving its central goal of increasing education and socioeconomic equality also from the novel perspective of this study. Since educational attainment and equality in it has well-known downstream effects on a wide variety of factors – relating to for example, later socioeconomic achievement (Hout and DiPrete 2006), crime and deviance (Huttunen et al. 2023), societal integration (Wiborg 2009), political participation (Lindgren, Oskarsson, and Persson 2019), as well as health<sup>8</sup> and mortality (Davies et al. 2018; Lundborg, Lyttkens, and Nystedt 2016) – the reform provided further promise for increasing equality in them and opened questions for future studies into a wider sociogenetic research program (e.g. Mills and Tropf 2020) on whether there were changes in the genetic effects with regards to this wide repertoire of factors. In addition – assuming again that increasing genetic prediction of education implies better realization of the potential of individuals – the results suggest that the reform not only reduced inequality but simultaneously promoted optimal allocation of human capital, with potentially wider macro-economic benefits. These results also shed light on the debate regarding whether individuals’ educational outcomes could be improved by personalized curricula using their genetic data (Asbury and Plomin 2013; Sabatello 2018). Morris, Davies and Davey Smith (2020) demonstrated that PGSs are too coarse to be suitable for tailoring, particularly if predictors such as parental factors or prior achievement, are available. On top of that, our results suggest that more personalized curricula (particularly at early ages) can even be counterproductive in optimizing the realization of capacities. Students’ and their parents’ preferences, test results – or even PGSs – cannot be assumed to be independent from social forces including socioeconomic resources that shape and lay constraints on them.

Finally, the results of this study suggest that conditions marked by instabilities and pending major changes – whether in the form of puberty among adolescents or reform implementation

at the institutional level – may be powerful magnifiers of the genetic effects (see discussion on “dimmers” in Domingue et al. 2020). Future research could explore and provide a more systematic picture on the potential of turbulent conditions interacting with the genome which, intriguingly enough, is one of the most stable factors of an individual with significant consequences for later life success, health, and well-being.

## Notes

---

<sup>1</sup> A fourth type of confounding is also possible. If we examine only a limited set of alleles in the analysis, the assessed genetic variant may not necessarily be the causal variant, but only in linkage disequilibrium with it. This problem is alleviated in our analysis that utilizes PGS (Schmitz and Conley 2017), which uses the information across the whole genome. However, such an advantage of PGS as a measure comes with the cost of increasing ambiguity on the actual biological mechanisms involved.

<sup>2</sup> With an exception that a few municipalities still had only a two-year civic school. However, empirical evidence indicates that the difference provided by the reform in the proportion of those still in school at the age of 15 was negligible (Pekkala Kerr, Pekkarinen, and Uusitalo 2013).

<sup>3</sup> Originating from Saunders's (2010) observation that the largest deviance from meritocratic socioeconomic achievement in the UK was that the advantaged families were able to protect their offspring from downward mobility even among their lower-ability offspring.

<sup>4</sup> Analogous results for a protective effect of socioeconomically advantaged schools have also been obtained using IQ instead of education PGS (Morris, Dorling, and Davey Smith 2016).

<sup>5</sup> Standard error for the post-reform effect  $b_1+b_3$  was estimated with the delta method as  $\sqrt{se_1^2 + se_3^2 + 2 * cov(b_1, b_3)}$ , where “se” is standard error of the corresponding regression coefficient “b”, and “cov” denotes covariance.

<sup>6</sup> As recommended by Clogg et al. (1995:1276) and Brame et al. (1998:8) standard error for difference in coefficients between educational backgrounds ( $b_{highedu}-b_{lowedu}$ ) is estimated as  $(se_{highedu}^2+se_{lowedu}^2)^{0.5}$ . Here, “se” is standard error of the corresponding regression coefficient “b”.

---

<sup>7</sup> However, the results were more mixed in a study investigating the variability of heritability of education within regions in Germany with varying timing and rigidness of tracking (Mönkediek 2022).

<sup>8</sup> Generally, mixed evidence regarding tracking on health and health inequality has been found (Böckerman et al. 2021; Delaruelle, van de Werfhorst, and Bracke 2019; Ravesteijn et al. 2017).

## References

- Aalberg, V., and M. Siimes. 2007. *Lapsesta Aikuiseksi*. 2nd edition. Helsinki: Nemo.
- Aho, Erkki, Kari Pitkanen, and Pasi Sahlberg. 2006. *Policy Development and Reform Principles of Basic and Secondary Education in Finland since 1968*. 2. the World Bank.
- Aromaa, Arpo, and Seppo Koskinen. 2004. *Health and Functional Capacity in Finland: Baseline Results of the Health 2000 Health Examination Survey*. Helsinki: National Public Health Institute.
- Asbury, Kathryn, and Robert Plomin. 2013. *G Is for Genes: The Impact of Genetics on Education and Achievement*. Vol. 24. John Wiley & Sons.
- Ayorech, Ziada, Eva Krapohl, Robert Plomin, and Sophie von Stumm. 2017. “Genetic Influence on Intergenerational Educational Attainment.” *Psychological Science* 28(9):1302–10.
- Baier, Tina, and Volker Lang. 2019. “The Social Stratification of Environmental and Genetic Influences on Education: New Evidence Using a Register-Based Twin Sample.” *Sociological Science* 6:143–71.
- Baier, Tina, Volker Lang, Michael Grätz, Kieron J. Barclay, Dalton C. Conley, Christopher T. Dawes, Thomas Laidley, and Torkild H. Lyngstad. 2022. “Genetic Influences on Educational Achievement in Cross-National Perspective.” *European Sociological Review* 38(6):959–74.
- Björklund, Anders, and Markus Jäntti. 2020. “Intergenerational Mobility, Intergenerational Effects, Sibling Correlations, and Equality of Opportunity: A Comparison of Four Approaches.” *Research in Social Stratification and Mobility* 70:1–11.
- Boardman, Jason D., Jonathan Daw, and Jeremy Freese. 2013. “Defining the Environment in Gene–Environment Research: Lessons from Social Epidemiology.” *American Journal of Public Health* 103(S1):S64–72.
- Böckerman, Petri, Mika Haapanen, Christopher Jepsen, and Alexandra Roulet. 2021. “School Tracking and Mental Health.” *Journal of Human Capital* 15(2):291–345.
- Borenstein, Michael, Larry V. Hedges, Julian PT Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester: Wiley.
- Borodulin, Katja, Hanna Tolonen, Pekka Jousilahti, Antti Jula, Anne Juolevi, Seppo Koskinen, Kari Kuulasmaa, Tiina Laatikainen, Satu Männistö, and Markku Peltonen. 2018. “Cohort Profile: The National FINRISK Study.” *International Journal of Epidemiology* 47(3):696–696i.
- Brame, Robert, Raymond Paternoster, Paul Mazerolle, and Alex Piquero. 1998. “Testing for the Equality of Maximum-Likelihood Regression Coefficients between Two Independent Equations.” *Journal of Quantitative Criminology* 14(3):245–61.
- 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):109–40.
- Choi, Shing Wan, Timothy Shin-Heng Mak, and Paul F. O’Reilly. 2020. “Tutorial: A Guide to Performing Polygenic Risk Score Analyses.” *Nature Protocols* 15(9):2759–72.



- Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. "Statistical Methods for Comparing Regression Coefficients between Models." *American Journal of Sociology* 100(5):1261–93.
- Colodro-Conde, Lucía, Frühling Rijdsdijk, María J. Tornero-Gómez, Juan F. Sánchez-Romera, and Juan R. Ordoñana. 2015. "Equality in Educational Policy and the Heritability of Educational Attainment." *PLoS One* 10(11):e0143796.
- Conley, Dalton. 2016. "Socio-Genomic Research Using Genome-Wide Molecular Data." *Annual Review of Sociology* 42:275–99.
- Davies, Neil M., Matt Dickson, George Davey Smith, Gerard J. van den Berg, and Frank Windmeijer. 2018. "The Causal Effects of Education on Health Outcomes in the UK Biobank." *Nature Human Behaviour* 2(2):117. doi: 10.1038/s41562-017-0279-y.
- Delaruelle, Katrijn, Herman van de Werfhorst, and Piet Bracke. 2019. "Do Comprehensive School Reforms Impact the Health of Early School Leavers? Results of a Comparative Difference-in-Difference Design." *Social Science & Medicine* 239:112542.
- Domingue, Benjamin W., Sam Trejo, Emma Armstrong-Carter, and Elliot M. Tucker-Drob. 2020. "Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges." *Sociological Science* 7:465–86.
- Engzell, Per, and Felix C. Tropf. 2019. "Heritability of Education Rises with Intergenerational Mobility." *Proceedings of the National Academy of Sciences*.
- Erola, Jani, Hannu Lehti, Tina Baier, and Alekski Karhula. 2021. "Socioeconomic Background and Gene-Environment Interplay in Social Stratification across the Early Life Course." *European Sociological Review* jcab026 (online publication ahead of print).
- Euling, Susan Y., Marcia E. Herman-Giddens, Peter A. Lee, Sherry G. Selevan, Anders Juul, Thorkild I. A. Sørensen, Leo Dunkel, John H. Himes, Grete Teilmann, and Shanna H. Swan. 2008. "Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends: Panel Findings." *Pediatrics* 121(Supplement\_3):S172–91. doi: 10.1542/peds.2007-1813D.
- Freese, Jeremy, and Yu-Han Jao. 2017. "Shared Environment Estimates for Educational Attainment: A Puzzle and Possible Solutions." *Journal of Personality* 85(1):79–89.
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press.
- Grätz, Michael, Kieron J. Barclay, Øyvind N. Wiborg, Torkild H. Lyngstad, Alekski Karhula, Jani Erola, Patrick Präg, Thomas Laidley, and Dalton Conley. 2021. "Sibling Similarity in Education across and within Societies." *Demography*.
- Harden, Kathryn Paige. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton: University Press.
- Harden, Kathryn Paige, Benjamin W. Domingue, Daniel W. Belsky, Jason D. Boardman, Robert Crosnoe, Margherita Malanchini, Michel Nivard, Elliot M. Tucker-Drob, and Kathleen Mullan Harris. 2020. "Genetic Associations with Mathematics Tracking and Persistence in Secondary School." *NPJ Science of Learning* 5(1):1–8.

- Härkönen, Juho, and Outi Sirniö. 2020. "Educational Transitions and Educational Inequality: A Multiple Pathways Sequential Logit Model Analysis of Finnish Birth Cohorts 1960–1985." *European Sociological Review*. doi: 10.1093/esr/jcaa019.
- Heath, A. C., K. Berg, L. J. Eaves, M. H. Solaas, L. A. Corey, J. Sundet, P. Magnus, and W. E. Nance. 1985. "Education Policy and the Heritability of Educational Attainment." *Nature* 314(6013):734–36. doi: 10.1038/314734a0.
- Herd, Pamela, Jeremy Freese, Kamil Sicinski, Benjamin W. Domingue, Kathleen Mullan Harris, Caiping Wei, and Robert M. Hauser. 2019. "Genes, Gender Inequality, and Educational Attainment." *American Sociological Review* 84(6):1069–98.
- Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2008. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Analysis & Policy* 7(2):1–46. doi: 10.2202/1935-1682.1775.
- Hout, Michael, and Thomas A. DiPrete. 2006. "What We Have Learned: RC28's Contributions to Knowledge about Social Stratification." *Research in Social Stratification and Mobility* 24(1):1–20. doi: 10.1016/j.rssm.2005.10.001.
- Howe, Laurence J., Michel G. Nivard, Tim T. Morris, Ailin F. Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor, Rafael Ahlskog, Penelope A. Lind, and Teemu Palviainen. 2022. "Within-Sibship Genome-Wide Association Analyses Decrease Bias in Estimates of Direct Genetic Effects." *Nature Genetics* 54(5):581–92.
- Huttunen, Kristiina, Tuomas Pekkarinen, Roope Uusitalo, and Hanna Virtanen. 2023. "Lost Boys? Secondary Education and Crime." *Journal of Public Economics* 218:104804.
- Kalalahti, Mira, and Janne Varjo. 2020. "Revisiting Universalism in the Finnish Education System." *Research on Finnish Society* 12(1):1–16.
- Karhunen, Hannu, and Roope Uusitalo. 2017. "50 vuotta koulutusmahdollisuuksien tasa-arvoa." *Yhteiskuntapolitiikka* 82(3):296–303.
- Kauko, Jaakko. 2019. "The Finnish Comprehensive School: Conflicts, Compromises, and Institutional Robustness." Pp. 122–42 in *Great Policy Successes: Or, A Tale About Why It's Amazing That Governments Get So Little Credit for Their Many Everyday and Extraordinary Achievements as Told by Sympathetic Observers Who Seek to Create Space for a Less Relentlessly Negative View of Our Pivotal Public Institutions*. Oxford: University Press.
- Kerminen, Sini, Aki S. Havulinna, Garrett Hellenthal, Alicia R. Martin, Antti-Pekka Sarin, Markus Perola, Aarno Palotie, Veikko Salomaa, Mark J. Daly, and Samuli Ripatti. 2017. "Fine-Scale Genetic Structure in Finland." *G3: Genes, Genomes, Genetics* 7(10):3459–68.
- Kivinen, Osmo, Sakari Ahola, and Juha Hedman. 2001. "Expanding Education and Improving Odds? Participation in Higher Education in Finland in the 1980s and 1990s." *Acta Sociologica* 44(2):171–81.
- Kivirauma, Joel, and Kari Ruoho. 2007. "Excellence through Special Education? Lessons from the Finnish School Reform." *International Review of Education* 53(3):283–302. doi: 10.1007/s11159-007-9044-1.

- Knigge, Antonie, Ineke Maas, Kim Stienstra, Eveline L. De Zeeuw, and Dorret I. Boomsma. 2022. "Delayed Tracking and Inequality of Opportunity: Gene-Environment Interactions in Educational Attainment." *NPJ Science of Learning* 7(1):1–13.
- Koivusilta, Leena, and A. Rimpelä. 2004. "Pubertal Timing and Educational Careers: A Longitudinal Study." *Annals of Human Biology* 31(4):446–65.
- Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsón, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdóttir, Asmundur Oddsson, Bjarni V. Halldorsson, and Gisli Masson. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359(6374):424–28.
- Lahtinen, Hannu, Pekka Martikainen, and Lasse Tarkiainen. 2022. "Changes in Sibling Similarity in Education among Finnish Cohorts Born in 1950–89: The Contribution of Paternal and Maternal Education." *Longitudinal and Life Course Studies* 13(4):496–526.
- Lazear, Edward P. 2001. "Educational Production." *The Quarterly Journal of Economics* 116(3):777–803.
- Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzián, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, Mark Alan Fontana, Tushar Kundu, Chanwook Lee, Hui Li, Ruoxi Li, and David Cesarini. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50(8):1112–21.
- Lin, Meng-Jung. 2020. "The Social and Genetic Inheritance of Educational Attainment: Genes, Parental Education, and Educational Expansion." *Social Science Research* 86:102387.
- Lindgren, Karl-Oskar, Sven Oskarsson, and Mikael Persson. 2019. "Enhancing Electoral Equality: Can Education Compensate for Family Background Differences in Voting Participation?" *American Political Science Review* 1–15.
- Lloyd-Jones, Luke R., Jian Zeng, Julia Sidorenko, Loïc Yengo, Gerhard Moser, Kathryn E. Kemper, Huanwei Wang, Zhili Zheng, Reedik Magi, and Tonu Esko. 2019. "Improved Polygenic Prediction by Bayesian Multiple Regression on Summary Statistics." *Nature Communications* 10(1):1–11.
- Lundborg, Petter, Carl Hampus Lyttkens, and Paul Nystedt. 2016. "The Effect of Schooling on Mortality: New Evidence from 50,000 Swedish Twins." *Demography* 53(4):1135–68.
- Malanchini, Margherita, Kaili Rimfeld, Andrea G. Allegrini, Stuart J. Ritchie, and Robert Plomin. 2020. "Cognitive Ability and Education: How Behavioural Genetic Research Has Advanced Our Knowledge and Understanding of Their Association." *Neuroscience & Biobehavioral Reviews* 111:229–45.
- Mansournia, Mohammad Ali, Maryam Nazemipour, Ashley I. Naimi, Gary S. Collins, and Michael J. Campbell. 2021. "Reflection on Modern Methods: Demystifying Robust Standard Errors for Epidemiologists." *International Journal of Epidemiology* 50(1):346–51. doi: 10.1093/ije/dyaa260.
- Manuck, Stephen B., and Jeanne M. McCaffery. 2014. "Gene-Environment Interaction." *Annual Review of Psychology* 65:41–70.

- Mills, Melinda C., Nicola Barban, and Felix C. Tropf. 2020. *An Introduction to Statistical Genetic Data Analysis*. Cambridge: MIT Press.
- Mills, Melinda C., and Felix C. Tropf. 2020. “Sociology, Genetics, and the Coming of Age of Sociogenomics.” *Annual Review of Sociology* 46(1):553–81. doi: 10.1146/annurev-soc-121919-054756.
- Mönkediek, Bastian. 2022. “How Variants of Tracking Affect the Role of Genes and Environment in Explaining Child Attendance at Upper Secondary School.” *Research in Social Stratification and Mobility* 81:100714.
- Morris, Tim T., Neil M. Davies, and George Davey Smith. 2020. “Can Education Be Personalised Using Pupils’ Genetic Data?” *Elife* 9:e49962.
- Morris, Tim T., Neil M. Davies, Gibran Hemani, and George Davey Smith. 2020. “Population Phenomena Inflate Genetic Associations of Complex Social Traits.” *Science Advances* 6(16):eaay0328.
- Morris, Tim T., Danny Dorling, and George Davey Smith. 2016. “How Well Can We Predict Educational Outcomes? Examining the Roles of Cognitive Ability and Social Position in Educational Attainment.” *Contemporary Social Science* 11(2–3):154–68. doi: 10.1080/21582041.2016.1138502.
- Neale, Zoe E., I. Sally, Chun Kuo, and Danielle M. Dick. 2021. “A Systematic Review of Gene-by-Intervention Studies of Alcohol and Other Substance Use.” *Development and Psychopathology* 33(4):1410–27.
- Oinonen, Päivi. 1969. *Kouluvalmiuden Ongelma: Milloin Lapsi on Kypsä Kouluun*. W. Söderström.
- Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko, Hyeokmoon Kweon, Grant Goldman, and Tamara Gjorgjieva. 2022. “Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals.” *Nature Genetics* 54(4):437–49.
- Okkonen, Ville. 2017. “Peruskoulua Vastaan. Yksityisoppikoulut Ja Yhteiskuntajärjestys 1966–1975.” *Turun Yliopiston Julkaisuja, Turun Yliopisto*.
- Ollikainen, Jani-Petteri. 2021. “Comprehensive School Reform and Labor Market Outcomes over the Lifecycle: Evidence from Finland.” *Labour Economics* 68:101952. doi: 10.1016/j.labeco.2020.101952.
- Pärn, Kalle, Marita A. Isokallio, Javier Nunez-Fontarnau, Aarno Palotie, Samuli Ripatti, and Priit Palta. 2018. *Genotype Imputation Workflow v3.0*.
- Pärn, Kalle, Javier Nunez-Fontarnau, Marita A. Isokallio, Timo Sipilä, Elina Kilpeläinen, Aarno Palotie, Samuli Ripatti, and Priit Palta. 2019. *Genotyping Chip Data Lift-over to Reference Genome Build GRCh38/Hg38. Protocols Io*.
- Pasman, Joëlle A., Karin JH Verweij, and Jacqueline M. Vink. 2019. “Systematic Review of Polygenic Gene–Environment Interaction in Tobacco, Alcohol, and Cannabis Use.” *Behavior Genetics* 49(4):349–65.

- Pekkala Kerr, Sari, Tuomas Pekkarinen, and Roope Uusitalo. 2013. "School Tracking and Development of Cognitive Skills." *Journal of Labor Economics* 31(3):577–602.
- Pekkarinen, Tuomas. 2008. "Gender Differences in Educational Attainment: Evidence on the Role of Tracking from a Finnish Quasi-Experiment." *Scandinavian Journal of Economics* 110(4):807–25.
- Pekkarinen, Tuomas. 2012. "Gender Differences in Education." *Nordic Economic Policy Review* 1(1):165–94.
- Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr. 2009. "School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform." *Journal of Public Economics* 93(7–8):965–73.
- Pfeffer, Fabian T. 2008. "Persistent Inequality in Educational Attainment and Its Institutional Context." *European Sociological Review* 24(5):543–65. doi: 10.1093/esr/jcn026.
- Plomin, Robert. 2019. *Blueprint: How DNA Makes Us Who We Are*. Cambridge: MIT Press.
- Polderman, Tinca JC, Beben Benyamin, Christiaan A. De Leeuw, Patrick F. Sullivan, Arjen Van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies." *Nature Genetics* 47(7):702–9.
- Poropat, Arthur E. 2009. "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance." *Psychological Bulletin* 135(2):322.
- Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38(8):904–9.
- Ravesteijn, Bastian, Hans van Kippersluis, Mauricio Avendano, Pekka Martikainen, Hannu Vessari, and Eddy Van Doorslaer. 2017. "The Impact of Later Tracking on Mortality by Parental Income in Finland."
- Rowe, David C., Kristen C. Jacobson, and Edwin JCG Van den Oord. 1999. "Genetic and Environmental Influences on Vocabulary IQ: Parental Education Level as Moderator." *Child Development* 70(5):1151–62.
- Sabatello, Maya. 2018. "A Genomically Informed Education System? Challenges for Behavioral Genetics." *Journal of Law, Medicine & Ethics* 46(1):130–44.
- Saunders, Peter. 2010. *Social Mobility Myths*. Civitas.
- Scarr-Salapatek, Sandra. 1971. "Race, Social Class, and IQ: Population Differences in Heritability of IQ Scores Were Found for Racial and Social Class Groups." *Science* 174(4016):1285–95.
- Schmitz, Lauren, and Dalton Conley. 2017. "Modeling Gene-Environment Interactions with Quasi-Natural Experiments." *Journal of Personality* 85(1):10–21.
- Selita, Fatos, and Yulia Kovas. 2019. "Genes and Gini: What Inequality Means for Heritability." *Journal of Biosocial Science* 51(1):18–47.

- Silventoinen, Karri, Aline Jelenkovic, Reijo Sund, Antti Latvala, Chika Honda, Fujio Inui, Rie Tomizawa, Mikio Watanabe, Norio Sakai, and Esther Rebato. 2020. “Genetic and Environmental Variation in Educational Attainment: An Individual-Based Analysis of 28 Twin Cohorts.” *Scientific Reports* 10(1):1–11.
- Silventoinen, Karri, Robert F. Krueger, Thomas J. Bouchard, Jaakko Kaprio, and Matt McGue. 2004. “Heritability of Body Height and Educational Attainment in an International Context: Comparison of Adult Twins in Minnesota and Finland.” *American Journal of Human Biology* 16(5):544–55. doi: 10.1002/ajhb.20060.
- Statistics Finland. 2007. “Finland 1917-2007. Education in Finland: More Education for More People.” Retrieved November 30, 2022 ([https://www.stat.fi/tup/suomi90/marraskuu\\_en.html](https://www.stat.fi/tup/suomi90/marraskuu_en.html)).
- Sysiharju, Anna-Liisa. 1969. “Social Factors and Recruitment to the Gymnasium in Finland.” *Pedagogisk Forskning* 13(1):79–102.
- Trejo, Sam, Daniel W. Belsky, Jason D. Boardman, Jeremy Freese, Kathleen Mullan Harris, Pam Herd, Kamil Sicinski, and Benjamin W. Domingue. 2018. “Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health.” *Sociological Science* 5:513–40. doi: 10.15195/v5.a22.
- Tucker-Drob, Elliot M., and Timothy C. Bates. 2016. “Large Cross-National Differences in Gene  $\times$  Socioeconomic Status Interaction on Intelligence.” *Psychological Science* 27(2):138–49.
- Turkheimer, Eric. 2000. “Three Laws of Behavior Genetics and What They Mean.” *Current Directions in Psychological Science* 9(5):160–64.
- UNESCO. 2012. *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics.
- Valkonen, Tapani, Tiina Pensola, and Marika Jalovaara. 1996. “Peruskoulu-Uudistuksen Vaikutus Sukupuolten, Alueiden Ja Sosiaaliryhmien Väliin Eroihin Perusasteen Jälkeisessä Koulutuksessa.” Pp. 316–35 in *Toteuttaako peruskoulu tasa-arvoa?*, Arvioni. Helsinki: Opetushallitus.
- Valkonen, Tapani, Tiina Pensola, and Marika Jalovaara. 1998. “Peruskoulu-Uudistuksen Vaikutus Sosiaaliryhmien Väliin Eroihin Ylioppilastutkinnon Suorittamisessa.” Pp. 229–40 in *Rekisteriaineistot yhteiskunta- ja sosiaalitutkimuksessa*. Helsinki: Gaudeamus.
- Van Der Zwaluw, Carmen S., and Rutger C. M. E. Engels. 2009. “Gene–Environment Interactions and Alcohol Use and Dependence: Current Status and Future Challenges.” *Addiction* 104(6):907–14. doi: 10.1111/j.1360-0443.2009.02563.x.
- Wang, Biyao, Jessie R. Baldwin, Tabea Schoeler, Rosa Cheesman, Wikus Barkhuizen, Frank Dudbridge, David Bann, Tim T. Morris, and Jean-Baptiste Pingault. 2021. “Robust Genetic Nurture Effects on Education: A Systematic Review and Meta-Analysis Based on 38,654 Families across 8 Cohorts.” *The American Journal of Human Genetics* 108(9):1780–91.

- van de Werfhorst, Herman G. 2018. "Early Tracking and Socioeconomic Inequality in Academic Achievement: Studying Reforms in Nine Countries." *Research in Social Stratification and Mobility* 58:22–32. doi: 10.1016/j.rssm.2018.09.002.
- van de Werfhorst, Herman G., and Jonathan JB Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28.
- Wiborg, Susanne. 2009. *Education and Social Integration: Comprehensive Schooling in Europe*. Springer.
- Young-Wolff, Kelly C., Mary-Anne Enoch, and Carol A. Prescott. 2011. "The Influence of Gene–Environment Interactions on Alcohol Consumption and Alcohol Use Disorders: A Comprehensive Review." *Clinical Psychology Review* 31(5):800–816. doi: 10.1016/j.cpr.2011.03.005.
- Zeng, J., J. Yang, F. Zhang, Z. Zheng, L. Lloyd-Jones, and M. Goddard. 2020. "GCTB: A Tool for Genome-Wide Complex Trait Bayesian Analysis."

## Tables

Table 1. Descriptive statistics

	<b>Mean</b>	<b>SD</b>			
PGS of education	0.00	1.0			
Years of education beyond basic	4.29	2.4			
Correlation: PGS*education	0.29				
	<b>%</b>	<b>Mean PGS</b>	<b>SD PGS</b>	<b>Mean education</b>	<b>SD education</b>
<b>Reform exposure</b>					
No	51	0.01	1.0	4.2	2.3
Yes	49	-0.01	1.0	4.4	2.4
<b>Reform exposure (alternative measurement)</b>					
No	51	0.01	1.0	4.2	2.3
1 <sup>st</sup> reform cohort	13	0.01	1.0	4.3	2.4
Subsequent cohorts	36	-0.02	1.0	4.5	2.4
<b>Gender</b>					
Men	45	0.01	1.0	3.9	2.3
Women	55	0.00	1.0	4.6	2.3
<b>Year of birth</b>					
1957	1	-0.12	0.9	3.7	1.8
1958	4	0.03	0.9	3.9	2.2
1959	7	0.01	1.0	4.1	2.2
1960	9	-0.06	1.0	4.0	2.1
1961	11	-0.05	1.0	4.2	2.4
1962	13	0.11	1.0	4.3	2.4
1963	12	-0.05	1.0	4.2	2.4
1964	12	-0.02	1.0	4.4	2.4
1965	10	0.00	1.0	4.5	2.3
1966	9	0.05	1.0	4.5	2.5
1967	6	0.00	0.9	4.5	2.7
1968	4	-0.05	1.0	4.3	2.4
1969	1	0.16	1.0	4.9	2.6
<b>Reform region</b>					
1972	11	-0.09	1.0	4.1	2.1
1973	17	-0.06	1.0	4.1	2.2
1974	27	0.00	1.0	4.3	2.3
1975	17	0.07	1.0	4.6	2.5
1976	18	0.01	1.0	4.2	2.4
1977	10	0.08	1.0	4.3	2.8
<b>Highest family education</b>					
Only basic	54	-0.13	1.0	3.8	2.1
More than basic	46	0.15	1.0	4.9	2.5
Total (N)	5,946				



Table 2. The association between one standard deviation change in education PGS and achieved years of education before and after the comprehensive school reform

	Pre-reform		Post-reform		Difference-in-difference		N
	b	95% CI	b	95% CI	b	95% CI	
<b>Whole sample</b>	0.59***	(0.51; 0.67)	0.68***	(0.61; 0.76)	0.08	(-0.03; 0.19)	5,946
<b>By subgroup</b>							
Men	0.49***	(0.37; 0.61)	0.69***	(0.57; 0.80)	0.17*	(0.01; 0.34)	2,680
Women	0.66***	(0.55; 0.76)	0.67***	(0.57; 0.78)	0.01	(-0.14; 0.15)	3,266
Basic family education	0.44***	(0.35; 0.54)	0.60***	(0.49; 0.71)	0.16*	(0.02; 0.31)	3,236
More than basic family education	0.67***	(0.54; 0.79)	0.64***	(0.53; 0.76)	-0.04	(-0.21; 0.13)	2,710

Notes: Inverse variance-weighted meta-analytic estimates from reform region -specific regressions including reform indicator, education PGS, and their interactions, adjusted for gender, year of birth, first ten principal components of the genome, study collection round and genotyping batch. These regression reform region-specific regression results are presented in Tables A1 and A2 in the appendix. Possible inequivalence between difference-in-difference estimates (which is the meta-analyzed interaction coefficient from reform region-specific regressions) and post-reform minus pre-reform estimates stems from the heteroscedasticity correction of the standard errors of reform-specific regressions. \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$  (two-tailed)

Table 3. The association between one standard deviation change in education PGS and years of education before, during and after the comprehensive school reform

	Pre-reform cohorts		1 <sup>st</sup> Reform cohort		Subsequent cohorts		N
	b	95% CI	b	95% CI	b	95% CI	
<b>Whole sample</b>	0.59***	(0.51; 0.67)	0.79***	(0.64; 0.94)	0.65***	(0.56; 0.74)	5,946
<b>By subgroup</b>							
Men	0.49***	(0.37; 0.61)	0.84***	(0.60; 1.09)	0.63***	(0.50; 0.76)	2,680
Women	0.66***	(0.55; 0.76)	0.72***	(0.54; 0.90)	0.65***	(0.52; 0.77)	3,266
Basic family education	0.45***	(0.35; 0.54)	0.79***	(0.57; 1.01)	0.54***	(0.42; 0.66)	3,236
More than basic family education	0.67***	(0.54; 0.79)	0.70***	(0.50; 0.89)	0.63***	(0.50; 0.77)	2,710
<b>Difference-in-difference</b>							
		Reform cohort vs. pre-reform		Subsequent vs. pre-reform			
	b	95% CI	b	95% CI			
<b>Whole sample</b>	0.17	(0.00; 0.33)	0.05	(-0.07; 0.17)			
<b>By subgroup</b>							
Men	0.33*	(0.05; 0.60)	0.11	(-0.06; 0.29)			
Women	0.04	(-0.17; 0.25)	-0.01	(-0.17; 0.15)			
Basic family education	0.34**	(0.10; 0.58)	0.10	(-0.06; 0.26)			
More than basic family education	-0.03	(-0.26; 0.21)	-0.04	(-0.22; 0.14)			

Notes: Inverse variance-weighted meta-analytic estimates from reform region -specific regressions including status in three categories, education PGS, and their interactions, adjusted for gender, year of birth, first ten principal components of the genome, study collection round and genotyping batch. These regression reform region-specific regression results are presented in Tables A3, A4 and A5 in the appendix. Possible inequivalence between difference-in-difference estimates (which is the meta-analyzed interaction coefficient from reform region-specific regressions) and post-reform minus pre-reform estimates stems from the heteroscedasticity correction of the standard errors of reform-specific regressions. \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$  (two-tailed)

Figures

Figure 1. School reform implementation years across municipalities in Finland

[Figure 1 removed from current version due to copyright reasons. It is a reproduction of the Figure 2 in Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment. *The Scandinavian Journal of Economics*, 110(4), 807–825.]

Figure 2. Number of observations by birth cohort and reform region (total N: 5,946)

1977						94	58	78	66	86	80	79	68
1976					130	135	132	139	122	135	139	133	
1975				110	138	135	125	128	116	122	132		
1974			215	195	179	212	200	199	193	194			
1973		152	141	136	127	123	111	111	124				
1972	88	80	80	87	86	74	73	86					
<b>Reform year region</b>	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
	<b>Birth cohort</b>												

Note: Lighter-shaded cells are pre-reform cohorts (“non-exposed”), darker-shaded cells are post reform (“exposed”) cohorts

## Appendix

Table A1. The association between one standard deviation change in education PGS and achieved years of education before and after the comprehensive school reform by reform region

Region	Pre-reform		Post-reform		Difference-in-difference		N (pre; post-reform)
	b	95% CI	b	95% CI	b	95% CI	
1972	0.66	(0.39; 0.92)	0.65	(0.41; 0.89)	-0.01	(-0.36; 0.35)	654 (335; 319)
1973	0.40	(0.22; 0.58)	0.64	(0.45; 0.82)	0.23	(-0.02; 0.49)	1,025 (556; 469)
1974	0.55	(0.41; 0.69)	0.58	(0.43; 0.72)	0.02	(-0.18; 0.22)	1,587 (801; 786)
1975	0.53	(0.34; 0.72)	0.67	(0.46; 0.87)	0.13	(-0.14; 0.41)	1,006 (508; 498)
1976	0.81	(0.61; 1.01)	0.66	(0.47; 0.84)	-0.15	(-0.42; 0.11)	1,065 (536; 529)
1977	0.82	(0.55; 1.10)	1.20	(0.95; 1.45)	0.38	(0.00; 0.75)	609 (296; 313)
Meta-analytic	0.59	(0.51; 0.67)	0.68	(0.61; 0.76)	0.08	(-0.03; 0.19)	5946 (3032; 2914)

Notes: Each point estimate is based on reform region (1972–1977) -specific linear regression models predicting years of education. Independent variables include reform indicator, education PGS, and their interaction, adjusted for gender, cohort, the first ten principal components of the genome, study collection round and genotyping batch. Meta-analytic estimate is an inverse variance–weighted fixed-effect estimate on these region-specific estimates.

Table A2. The association between one standard deviation change in education PGS and achieved years of education before and after the comprehensive school reform by reform region, heterogeneous effects by gender and family education

<b>Men</b>							
Region	Pre-reform		Post-reform		Difference-in-difference		N (pre; post-reform)
	b	95% CI	b	95% CI	b	95% CI	
1972	0.82	(0.39; 1.26)	0.23	(-0.13; 0.59)	-0.60	(-1.14; -0.05)	269 (148; 121)
1973	0.26	(0.02; 0.50)	0.77	(0.47; 1.06)	0.50	(0.13; 0.88)	469 (247; 222)
1974	0.41	(0.21; 0.62)	0.52	(0.32; 0.73)	0.11	(-0.17; 0.40)	720 (363; 357)
1975	0.47	(0.17; 0.76)	0.81	(0.49; 1.14)	0.35	(-0.08; 0.77)	456 (237; 219)
1976	0.76	(0.44; 1.08)	0.70	(0.43; 0.97)	-0.06	(-0.48; 0.35)	477 (238; 239)
1977	0.82	(0.39; 1.26)	1.44	(1.06; 1.82)	0.62	(0.03; 1.20)	289 (150; 139)
Meta-analytic	0.49	(0.37; 0.61)	0.69	(0.57; 0.80)	0.17	(0.01; 0.34)	2680 (1383; 1297)
<b>Women</b>							
Region	Pre-reform		Post-reform		Difference-in-difference		N (pre; post-reform)
	b	95% CI	b	95% CI	b	95% CI	
1972	0.52	(0.17; 0.87)	0.91	(0.63; 1.19)	0.39	(-0.07; 0.84)	385 (187; 198)
1973	0.52	(0.27; 0.78)	0.54	(0.30; 0.78)	0.01	(-0.33; 0.36)	556 (309; 247)
1974	0.68	(0.49; 0.87)	0.62	(0.41; 0.82)	-0.06	(-0.34; 0.22)	867 (438; 429)
1975	0.59	(0.34; 0.84)	0.54	(0.27; 0.81)	-0.05	(-0.40; 0.31)	550 (271; 279)
1976	0.82	(0.57; 1.07)	0.65	(0.40; 0.90)	-0.17	(-0.52; 0.18)	588 (298; 290)
1977	0.78	(0.41; 1.14)	1.01	(0.68; 1.35)	0.24	(-0.27; 0.75)	320 (146; 174)
Meta-analytic	0.66	(0.55; 0.76)	0.67	(0.57; 0.78)	0.01	(-0.14; 0.15)	3266 (1649; 1617)
<b>Basic family education</b>							
Region	Pre-reform		Post-reform		Difference-in-difference		N (pre; post-reform)
	b	95% CI	b	95% CI	b	95% CI	
1972	0.49	(0.15; 0.83)	0.42	(0.11; 0.74)	-0.07	(-0.54; 0.40)	401 (208; 193)
1973	0.37	(0.15; 0.58)	0.68	(0.44; 0.91)	0.31	(-0.01; 0.63)	670 (370; 300)
1974	0.41	(0.24; 0.58)	0.52	(0.33; 0.70)	0.11	(-0.14; 0.35)	928 (492; 436)
1975	0.43	(0.17; 0.69)	0.63	(0.32; 0.95)	0.20	(-0.20; 0.60)	480 (255; 225)
1976	0.61	(0.36; 0.86)	0.77	(0.47; 1.06)	0.16	(-0.23; 0.55)	494 (275; 219)
1977	0.45	(0.05; 0.86)	0.73	(0.30; 1.16)	0.28	(-0.29; 0.84)	263 (144; 119)
Meta-analytic	0.44	(0.35; 0.54)	0.60	(0.49; 0.71)	0.16	(0.02; 0.31)	3236 (1744; 1492)
<b>More than basic family education</b>							
Region	Pre-reform		Post-reform		Difference-in-difference		N (pre; post-reform)
	b	95% CI	b	95% CI	b	95% CI	
1972	0.97	(0.50; 1.43)	0.81	(0.42; 1.20)	-0.16	(-0.76; 0.45)	253 (127; 126)
1973	0.48	(0.19; 0.78)	0.48	(0.15; 0.81)	-0.01	(-0.44; 0.43)	355 (186; 169)
1974	0.62	(0.38; 0.87)	0.57	(0.35; 0.79)	-0.06	(-0.38; 0.27)	659 (309; 350)
1975	0.43	(0.13; 0.72)	0.58	(0.32; 0.83)	0.15	(-0.23; 0.53)	526 (253; 273)
1976	0.85	(0.55; 1.16)	0.50	(0.27; 0.73)	-0.35	(-0.73; 0.02)	571 (261; 310)
1977	0.97	(0.60; 1.34)	1.21	(0.90; 1.52)	0.24	(-0.26; 0.74)	346 (152; 194)
Meta-analytic	0.67	(0.54; 0.79)	0.64	(0.53; 0.76)	-0.04	(-0.21; 0.13)	2710 (1288; 1422)

Notes: Each point estimate is based on reform region (1972–1977) -specific linear regression models predicting years of education. Independent variables include reform indicator (unless stratified), education PGS, and their interaction, adjusted for gender, birth cohort, first ten principal components of the genome, study collection round and genotyping batch. Meta-analytic estimate is an inverse variance–weighted fixed-effect estimate on these region-specific estimates.

Table A3 The association between one standard deviation change in education PGS and achieved years of education before, during and after the comprehensive school reform by reform region

Region	Pre-reform cohorts		1 <sup>st</sup> Reform cohort		Subsequent cohorts	
	b	95% CI	b	95% CI	b	95% CI
1972	0.66	(0.39; 0.92)	0.64	(0.16; 1.13)	0.66	(0.38; 0.94)
1973	0.40	(0.22; 0.58)	0.81	(0.44; 1.18)	0.59	(0.37; 0.81)
1974	0.55	(0.41; 0.69)	0.74	(0.46; 1.01)	0.53	(0.36; 0.69)
1975	0.53	(0.34; 0.72)	0.85	(0.46; 1.24)	0.60	(0.37; 0.84)
1976	0.81	(0.61; 1.01)	0.47	(0.11; 0.82)	0.71	(0.50; 0.93)
1977	0.83	(0.55; 1.10)	1.27	(0.89; 1.64)	1.17	(0.85; 1.49)
Meta-analytic	0.59	(0.51; 0.67)	0.79	(0.64; 0.94)	0.65	(0.56; 0.74)
Difference-in-difference						
	1 <sup>st</sup> Reform cohort vs. pre-reform		Subsequent cohorts vs. pre-reform			
Region	b	95% CI	b	95% CI		
1972	-0.01	(-0.56; 0.53)	0.00	(-0.39; 0.39)		
1973	0.41	(0.00; 0.82)	0.19	(-0.10; 0.47)		
1974	0.19	(-0.12; 0.49)	-0.03	(-0.25; 0.19)		
1975	0.32	(-0.12; 0.75)	0.07	(-0.22; 0.37)		
1976	-0.34	(-0.74; 0.06)	-0.09	(-0.38; 0.19)		
1977	0.44	(-0.03; 0.91)	0.35	(-0.08; 0.77)		
Meta-analytic	0.17	(0.00; 0.33)	0.05	(-0.07; 0.17)		

Notes: Each row (1972–1977) is based on separate reform region -specific linear regression models including reform status in three categories, education PGS, and their interaction, adjusted for gender, cohort, first ten principal components of the genome, study collection round and genotyping batch. Meta-analytic estimate is an inverse variance–weighted fixed-effect estimate on these region-specific estimates.

Table A4. The association between one standard deviation change in education PGS and years of education before, during and after the comprehensive school reform by reform region, heterogeneous effect by gender

<b>Men</b>						
Region	Pre-reform cohorts		1 <sup>st</sup> Reform cohort		Subsequent cohorts	
	b	95% CI	b	95% CI	b	95% CI
1972	0.82	(0.38; 1.26)	0.23	(-0.46; 0.92)	0.23	(-0.24; 0.69)
1973	0.26	(0.02; 0.50)	0.97	(0.46; 1.48)	0.70	(0.33; 1.07)
1974	0.41	(0.20; 0.62)	0.83	(0.34; 1.31)	0.44	(0.22; 0.66)
1975	0.46	(0.17; 0.76)	1.07	(0.36; 1.77)	0.73	(0.39; 1.07)
1976	0.76	(0.44; 1.08)	0.92	(0.22; 1.62)	0.64	(0.35; 0.93)
1977	0.83	(0.39; 1.27)	0.95	(0.30; 1.60)	1.67	(1.19; 2.16)
Meta-analytic	0.49	(0.37; 0.61)	0.84	(0.60; 1.09)	0.63	(0.50; 0.76)
Difference-in-difference						
Region	Reform cohort vs. pre-reform		Subsequent cohorts vs. pre-reform			
	b	95% CI	b	95% CI		
1972	-0.59	(-1.38; 0.20)	-0.60	(-1.22; 0.03)		
1973	0.71	(0.16; 1.27)	0.44	(-0.01; 0.89)		
1974	0.42	(-0.12; 0.95)	0.03	(-0.27; 0.32)		
1975	0.60	(-0.14; 1.35)	0.26	(-0.18; 0.70)		
1976	0.16	(-0.61; 0.93)	-0.12	(-0.54; 0.31)		
1977	0.12	(-0.66; 0.90)	0.84	(0.18; 1.51)		
Meta-analytic	0.33	(0.05; 0.60)	0.11	(-0.06; 0.29)		
<b>Women</b>						
Region	Pre-reform		1 <sup>st</sup> Reform cohort		Subsequent cohort	
	b	95% CI	b	95% CI	b	95% CI
1972	0.52	(0.17; 0.87)	0.88	(0.26; 1.49)	0.90	(0.59; 1.22)
1973	0.52	(0.27; 0.78)	0.68	(0.16; 1.20)	0.50	(0.23; 0.78)
1974	0.68	(0.49; 0.87)	0.70	(0.38; 1.01)	0.58	(0.33; 0.83)
1975	0.59	(0.34; 0.84)	0.70	(0.20; 1.20)	0.49	(0.18; 0.80)
1976	0.81	(0.56; 1.06)	0.26	(-0.13; 0.64)	0.79	(0.49; 1.09)
1977	0.78	(0.41; 1.14)	1.54	(1.02; 2.05)	0.76	(0.31; 1.21)
Meta-analytic	0.66	(0.55; 0.76)	0.72	(0.54; 0.90)	0.65	(0.52; 0.77)
Difference-in-difference						
Region	Reform cohort vs. pre-reform		Subsequent cohorts vs. pre-reform			
	b	95% CI	b	95% CI		
1972	0.36	(-0.35; 1.07)	0.39	(-0.09; 0.87)		
1973	0.15	(-0.42; 0.72)	-0.02	(-0.39; 0.35)		
1974	0.02	(-0.35; 0.38)	-0.10	(-0.42; 0.22)		
1975	0.11	(-0.45; 0.67)	-0.10	(-0.48; 0.29)		
1976	-0.55	(-1.01; -0.10)	-0.02	(-0.41; 0.37)		
1977	0.76	(0.11; 1.40)	-0.02	(-0.60; 0.56)		
Meta-analytic	0.04	(-0.17; 0.25)	-0.01	(-0.17; 0.15)		

Notes: Each row (1972–1977) is based on separate reform region -specific linear regression models including reform status in three categories, education PGS, and their interaction, adjusted for cohort, first ten principal components of the genome, study collection round and genotyping batch. Meta-analytic estimate is an inverse variance-weighted fixed-effect estimate on these region-specific estimates.

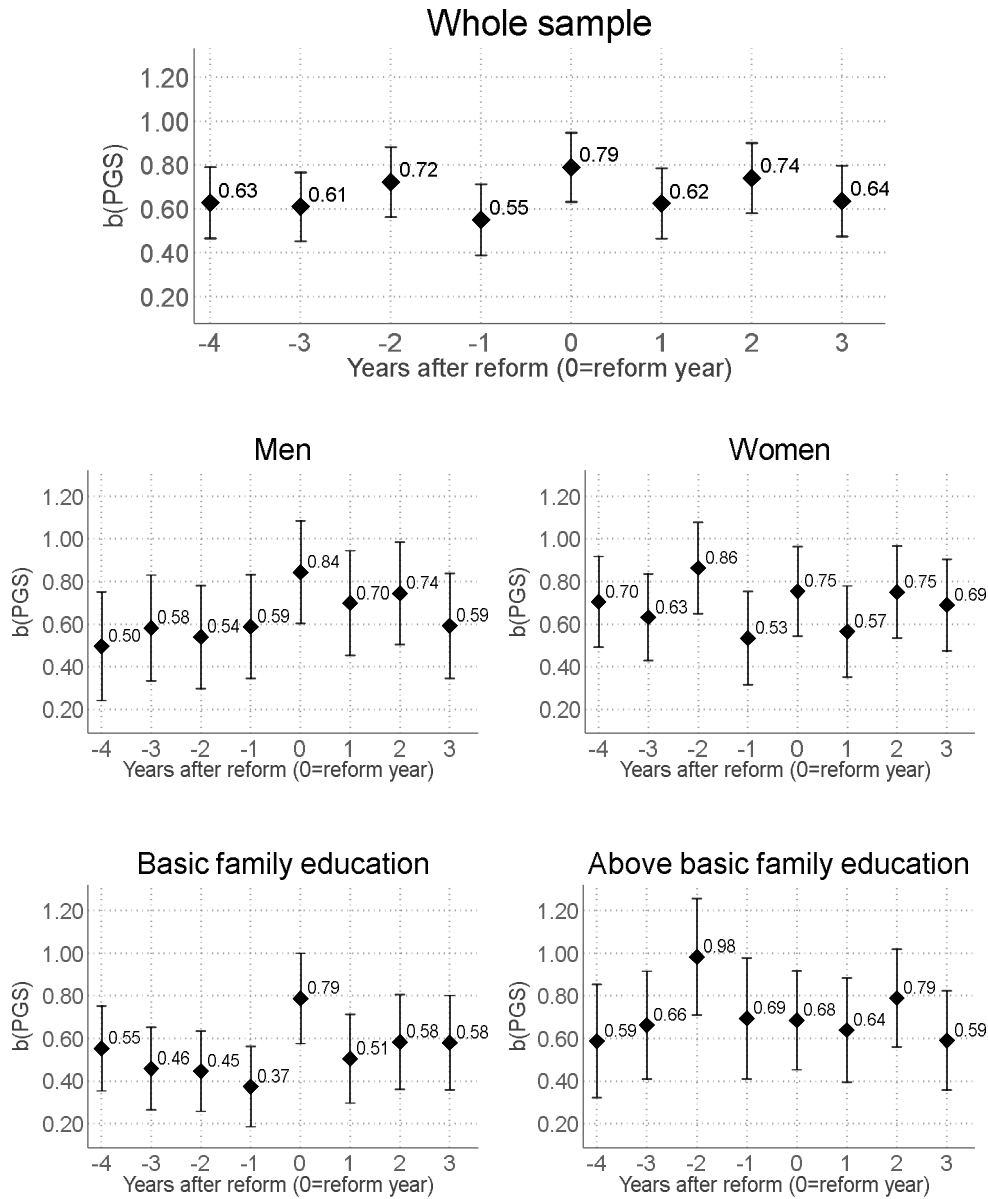
Table A5 The association between one standard deviation change in education PGS and years of education before, during and after the comprehensive school reform by reform region, heterogeneous effect by parental education

<b>Those with basic family education</b>						
Region	Pre-reform		1 <sup>st</sup> Reform cohort		Subsequent cohort	
	b	95% CI	b	95% CI	b	95% CI
1972	0.50	(0.16; 0.84)	0.61	(-0.03; 1.25)	0.35	(0.00; 0.7)
1973	0.37	(0.15; 0.59)	1.06	(0.56; 1.56)	0.57	(0.3; 0.84)
1974	0.41	(0.24; 0.58)	0.62	(0.23; 1.00)	0.48	(0.27; 0.69)
1975	0.43	(0.17; 0.69)	0.66	(0.00; 1.31)	0.61	(0.26; 0.96)
1976	0.60	(0.35; 0.85)	0.97	(0.41; 1.54)	0.74	(0.40; 1.07)
1977	0.47	(0.06; 0.88)	0.89	(0.16; 1.62)	0.57	(0.02; 1.11)
Meta-analytic	0.45	(0.35; 0.54)	0.79	(0.57; 1.01)	0.54	(0.42; 0.66)
Difference-in-difference						
Region	1 <sup>st</sup> Reform cohort vs. pre-reform		Subsequent vs. pre-reform			
	b	95% CI	b	95% CI		
1972	0.12	(-0.60; 0.83)	-0.15	(-0.65; 0.35)		
1973	0.69	(0.15; 1.23)	0.20	(-0.15; 0.55)		
1974	0.20	(-0.22; 0.62)	0.07	(-0.20; 0.34)		
1975	0.23	(-0.47; 0.92)	0.18	(-0.24; 0.6)		
1976	0.37	(-0.25; 0.99)	0.13	(-0.29; 0.55)		
1977	0.42	(-0.40; 1.24)	0.10	(-0.57; 0.76)		
Meta-analytic	0.34	(0.10; 0.58)	0.10	(-0.06; 0.26)		
<b>Those with more than basic family education</b>						
Region	Pre-reform		Reform cohort		Subsequent cohort	
	b	95% CI	b	95% CI	b	95% CI
1972	0.96	(0.50; 1.42)	0.57	(-0.2; 1.34)	0.95	(0.52; 1.39)
1973	0.48	(0.18; 0.78)	0.29	(-0.35; 0.93)	0.54	(0.14; 0.94)
1974	0.62	(0.38; 0.87)	0.80	(0.42; 1.19)	0.50	(0.24; 0.76)
1975	0.43	(0.13; 0.73)	0.65	(0.17; 1.12)	0.54	(0.26; 0.83)
1976	0.85	(0.55; 1.16)	0.22	(-0.18; 0.62)	0.60	(0.33; 0.87)
1977	0.96	(0.59; 1.34)	1.49	(1.02; 1.96)	1.07	(0.66; 1.49)
Meta-analytic	0.67	(0.54; 0.79)	0.70	(0.50; 0.89)	0.63	(0.50; 0.77)
Difference-in-difference						
Region	1 <sup>st</sup> Reform cohort vs. pre-reform		Subsequent vs. pre-reform			
	b	95% CI	b	95% CI		
1972	-0.39	(-1.29; 0.51)	-0.01	(-0.65; 0.64)		
1973	-0.20	(-0.89; 0.50)	0.06	(-0.43; 0.55)		
1974	0.18	(-0.27; 0.63)	-0.12	(-0.47; 0.23)		
1975	0.22	(-0.33; 0.77)	0.12	(-0.29; 0.52)		
1976	-0.63	(-1.14; -0.13)	-0.26	(-0.66; 0.15)		
1977	0.53	(-0.10; 1.16)	0.11	(-0.45; 0.67)		
Meta-analytic	-0.03	(-0.26; 0.21)	-0.04	(-0.22; 0.14)		

Notes: Each row (1972–1977) is based on separate reform region -specific linear regression models including reform status in three categories, education PGS, and their interaction, adjusted for gender, cohort, first ten principal components of the genome, study collection round and genotyping batch. Meta-analytic estimate is an inverse variance–weighted fixed-effect estimate on these region-specific estimates.



Figure A6. The association between one standard deviation change in education PGS and years of education by the years to reform



Notes: Sub-figures are based on linear regression including categorical years-after-reform variable, education PGS and the interaction between them. Capped bars are 95% confidence intervals. Models adjusted by gender (if not stratified), cohort, first ten principal components of the genome, study collection round and genotyping batch.