

Polygenic Prediction of Education and Its Role in the Intergenerational Transmission of Education: Cohort Changes Among Finnish Men and Women Born in 1925–1989

Hannu Lahtinen, Kaarina Korhonen, Pekka Martikainen, and Tim Morris

ABSTRACT Major changes in the educational distribution of the population and in institutions over the past century have affected the societal barriers to educational attainment. These changes can possibly result in stronger genetic associations. Using genetically informed, population-representative Finnish surveys linked to administrative registers, we investigated the polygenic associations and intergenerational transmission of education for those born between 1925 and 1989. First, we found that a polygenic index (PGI) designed to capture genetic predisposition to education strongly increased the predictiveness of educational attainment in pre-1950s cohorts, particularly among women. When decomposing the total contribution of PGI across different educational transitions, the transition between the basic and academic secondary tracks was the most important. This transition accounted for 60–80% of the total PGI–education association among most cohorts. The transition between academic secondary and higher tertiary levels increased its contribution across cohorts. Second, for cohorts born between 1955 and 1984, we observed that one eighth of the association between parental and one’s own education is explained by the PGI. There was also an increase in the intergenerational correlation of education among these cohorts, which was partly explained by an increasing association between family education of origin and the PGI.

KEYWORDS Polygenic index • Education • Intergenerational • Multiple pathways sequential logit model • Partial rank-order correlation

Introduction

Over the last century, the educational distribution of the population and institutions have undergone dramatic changes. Once a pursuit primarily of young men from privileged backgrounds, higher education has become increasingly accessible for broader segments of the population (Breen 2010). Having only the minimum educational qualifications, in turn, has increasingly become a disadvantaged position (Gesthuizen et al. 2011). Education is the main pathway to many occupations, providing an opportunity for social mobility on the one hand and to reinforce the intergenerational persistence of social position on the other (Breen and Müller 2020; Hout and DiPrete 2006).

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (<https://doi.org/10.1215/00703370-10963788>) contains supplementary material.

Education also influences social networks and relationships (Huang et al. 2009; Mäenpää 2014; Savage 2015), political participation (Gallego 2007), income (Uusitalo 1999), and labor force participation (Gesthuizen et al. 2011; Lahtinen et al. 2020) and is among the most established social risk factors for early labor market exit (Järnefelt 2010), poor health, and mortality (Davies et al. 2018; Mackenbach et al. 2016).

Virtually every human trait is influenced in some part by genetics (Polderman et al. 2015; Turkheimer 2000), and educational attainment is no exception (Branigan et al. 2013; Ge et al. 2017; Lee et al. 2018; Silventoinen et al. 2020). The genetic composition of populations typically changes substantially only over many generations (Kong et al. 2017), thus changes in genetic distributions are an unlikely explanation for the faster-paced trends of societal changes, including the sharp rise in average educational level observed across cohorts. In addition, factors such as improvements in the overall standard of living and welfare-state provisions—including longer compulsory schooling, better school support and cost subvention, and decreases in the use of curriculum tracks—may have reduced the significance of the role of social background in educational attainment (e.g., Breen et al. 2009, 2010). Hence, social changes may have reduced the variation in education that can be attributed to environmental factors, resulting in a greater amount of variation being explained by genetics.

Many behavioral geneticists have advanced the thesis that a higher genetic heritability of education (i.e., the proportion of variation in education within a population that is a result of genetic differences between individuals) is indicative of equality of opportunity (Ayorech et al. 2017; Conley 2016; Harden 2021; Heath et al. 1985; Plomin 2019; Selita and Kovas 2019; Silventoinen et al. 2020; Trzaskowski et al. 2014). According to this view, higher heritability indicates smaller environmental influences and is expected in contexts where individuals have better social or environmental opportunities to fulfill their natural potential in the absence of societal obstacles or privileges (i.e., there is less environmental variation). Several empirical results support such a view. Genetic factors were more predictive of educational outcomes in Estonia after the collapse of the Soviet Union (Rimfeld et al. 2018), and the heritability of education was found to be higher after educational reforms aiming to increase equality of opportunity (Colodro-Conde et al. 2015; Heath et al. 1985).

The persistence of education between generations has received considerable attention in sociology, economics, and related social sciences (e.g., Björklund and Jäntti 2020; Björklund and Salvanes 2011; Breen and Jonsson 2005; Breen and Müller 2020; Pfeffer 2008). An important motivation has been to assess equality of opportunity. A weak association between parental social position and offspring education is generally considered to reflect more equal life chances for individuals with differing social origins, and thus differing resources and other structural constraints. Because the genome constitutes a direct mechanism through which differences are transmitted from parents to children, claims that both weak intergenerational association and strong heritability of education imply equality of opportunity provide an interesting tension.¹ Despite the possible friction between these two views for equality

¹ However, the view that greater heritability in socioeconomic achievement implies a fairer society can be problematized as well. Another commonly made distinction in this context is between ascribed and achieved status characteristics of individuals (Blau and Duncan 1967; Linton 1936: chapter VIII). A society relying more on achieved (earned) characteristics is typically considered as giving fairer life chances

of opportunity, Engzell and Tropf (2019) observed a negative association between country- and cohort-level intergenerational correlations of education and its heritability, suggesting that they may indeed be compatible.

Recent methodological developments allow researchers to predict individuals' outcomes directly from their molecular genetic information. The associations of individual genetic variants on many traits, such as educational attainment, are typically each very small, but can be combined into a polygenic index (PGI; e.g., Becker et al. 2021; Mills et al. 2020) that has greater explanatory power. The PGI provides a summary measure of the known genetic propensity for a trait. The index is generated by multiplying the association strength of a given copy of a single-nucleotide polymorphism (SNP, the most fundamental unit of genetic variation), obtained from an independent genome-wide association study (GWAS), by the number of copies that an individual has (zero, one, or two). That is: $PGI_i = \sum w_{snp} a_{snp,i}$ where w is the GWAS weight for the SNP in question and a is the allele count for this SNP for an individual i . The strongest PGI of educational attainment to date was based on a study of three million individuals and has been shown to predict 12–16% of the variation in the years of education in independent samples (Okbay et al. 2022). With PGIs, we can model genetic associations flexibly in population-representative samples and triangulate results obtained from other designs, such as intergenerational, sibling, twin, or adoption studies.

Using genetically informed, population-representative Finnish surveys linked to register-based data, we will approach the issues of genetic prediction of education and its role in the intergenerational transmission of education with a twofold investigation. First, we assess how polygenic prediction of education has changed among Finnish men and women born between 1925 and 1989. Second, we quantify the extent to which the overall intergenerational association of education can be attributed to a genetic propensity for educational attainment, and how this has changed across cohorts born between 1955 and 1984.

Models of Genetic Prediction of Education in the Changing Finnish Context

Genetics of Educational Transitions

The first aim of this study is to analyze the changes in polygenic prediction of education across the 1925–1989 cohorts in Finland. A recurrent challenge in analyzing changes in the correlates of education over time is that these changes are muddled by changing distributions of education. Educational distributions are determined by factors such as the structure of educational institutions and the demand for education, which are often not considered interesting to investigate, but rather something to be controlled for in models (Breen and Jonsson 2005; Härkönen and Sirniö 2020). A solution proposed by Mare (1980, 1981) specified sequential logistic models of

under the dominant liberal ethos of modern societies. In contrast, genetic material arguably belongs in the ascribed (unearned) characteristics of an individual (although this argument can be contested as well; see Nielsen 2006).

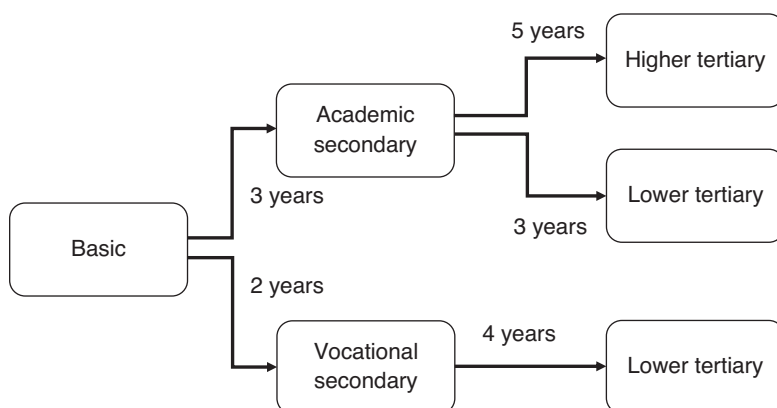


Fig. 1 A simplified model of educational transitions and the associated gains in expected years of education in the Finnish educational system. Model and associated years are from Härkönen and Sirniö (2020).

educational transitions. This approach was more parsimoniously reformulated by Buis (2017). Härkönen and Sirniö (2020) extended the model to allow multiple transitions at any given point of educational decision-making and specified a model for the Finnish educational system that we apply here. Following Härkönen and Sirniö (2020), we refer to it as a multiple pathways sequential logit model.

The starting point of this approach is to consider education as a series of successive transitions between educational levels. The association of each transition is first estimated through a logistic model and then weighted. A transition receives more weight where (1) more people are at risk of the transition in question; (2) the transition differentiates between more individuals, that is, the probability of the transition is close to .5; and (3) people gain more years of education by passing this transition.

Beyond merely solving technical issues and being able to present results of multiple models parsimoniously, such an approach is appealing since it mimics the structurally constrained decision-making process that individual agents face in navigating their educational careers. Focusing on specific transitions—for example, between academically and vocationally oriented tracks as well as between lower and higher level transitions—also provides an opportunity for a more nuanced understanding of the processes via which genetic variation manifests in educational attainment.

Figure 1 presents the model of educational transitions and associated gains in the expected years of education as applied in the Finnish educational system, following the specification of Härkönen and Sirniö (2020). At each level (except the rightmost ones), individuals can either finish their education or continue on to a further degree. It should be noted that the model is a simplified model of the Finnish educational system. For example, a joint completion of vocational and academic tracks is possible. In these cases, an individual is classified in the academic track in our model. Unsurprisingly, the educational system has also changed over the years covered by this study. In particular, the time to attain a lower tertiary degree was often 2–3 years (typically vocational college degrees), but from 1990 onward it typically took four years (polytechnic degrees), while three years is also possible (bachelor's degrees from universities). For comparability and parsimony, each of these is considered to

take three years in our model. It is also possible to continue to university after obtaining a vocational degree, but this transition is rare (1–2% of the study population) and was thus combined with a lower tertiary degree, in line with the original formulation of the model (Härkönen and Sirniö 2020).

Increases in educational attainment across cohorts have generally been larger among women than among men (DiPrete and Buchmann 2013; Pekkarinen 2012), which suggests that it may also be worthwhile to examine gender-related differences in the polygenic prediction of education over time. Historically, societal obstacles have hindered women's education, which may suggest stronger changes in the relationship between PGI and education among women compared with men. In the United States, Herd and colleagues (2019) observed an increase in the genetic associations with education across cohorts among women but not among men. Many women with a high genetic propensity to education also returned to school after educational opportunities improved in the 1970s and 1980s.

The Role of Genetics in the Intergenerational Transmission of Education

The second aim of this study is to situate the polygenic prediction of education in the wider context of the intergenerational reproduction of education. We analyze how the PGI contributes to the intergenerational transmission of education and whether this contribution has changed over time. Because individuals obtain their genome completely from the parents (save mutations randomly occurring during the life course), genetic influences act as a natural candidate to explain the intergenerational transmission of traits. Despite this, achieved education necessarily requires manifestation through societal institutions and culture, and thus can never be fully reduced to genetics. Consequently, trends in the intergenerational correlation of education and the correlation between a PGI of education and educational attainment do not need to move in the same direction. Instead of making such an assumption, changes in these three-way interrelationships between education of origin and destination and PGI of education will be empirically assessed in this study.

Figure 2 demonstrates our theoretical path model. We build a path model that we call the origin–genome–destination (*OGD*) triangle.² Origin (*O*) is measured as parental education, genome (*G*) as the PGI of education, and destination (*D*) as an individual's own achieved education. We analyze the strength and trends of each pathway and to what extent the association between origin and destination is mediated by the genome.

In these decompositions, we use partial Spearman's rank-order correlations (Liu et al. 2018). Again, this method is relatively robust to changes in marginal distributions of education in either generation and offers flexibility in assumptions regarding the

² The framework draws inspiration from the decomposition of the origin–education–destination (OED) triangle (Bernardi and Ballarino 2016; Bukodi and Goldthorpe 2016), an influential model in the analyses of social class mobility. In OED decomposition, the association between social class of origin and destination is elaborated by analyzing the extent to which education mediates it. Our analysis can be understood as an extension in this framework. We concentrate on one component of the OED triangle, namely, the OE path, and investigate the contribution of genetics in explaining it.

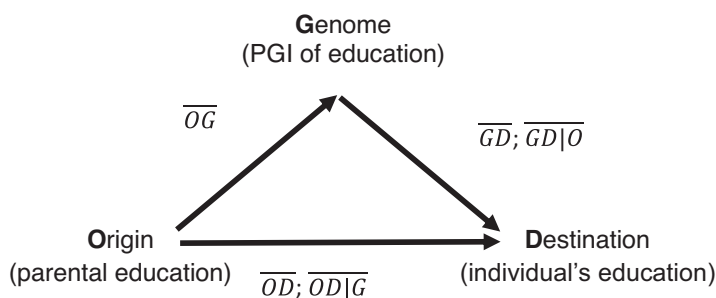


Fig. 2 The origin–genome–destination (OGD) triangle and the estimated pathways

functional form of variables. For example, the method relaxes the need to assume distances between educational categories or that PGI affects education in a linear form.

The Finnish Context

To date, studies exploring educational stratification using molecular genetic data have largely focused on the United States and United Kingdom contexts. These countries have comparatively limited liberal welfare states (see Esping-Andersen 1990), high inequality, and low social mobility compared with Finland, which has a more extensive social-democratic welfare state, with less socioeconomic inequality and a more egalitarian educational system. Finland is characterized by one of the weakest observed associations in educational attainment between parents and children (Pfeffer 2008) and between siblings (Grätz et al. 2021), and also has comparatively small achievement differences between schools (Bernelius and Kauppinen 2012; Debeer et al. 2014). Currently, there is free-of-charge tuition at all levels, as well as widely available student allowances, state-guaranteed student loans, and special education for those with learning difficulties, particularly for the cohorts born after the early 1960s, who experienced comprehensive school reform (Kivirauma and Ruoho 2007). The Finnish educational system employs comparatively little curriculum tracking and has high nationwide standardization (Pfeffer 2008).

There is evidence of a weakening trend in the intergenerational association of education among the earlier cohorts of this study (Hertz et al. 2008; Kivinen et al. 2001; Pfeffer 2008). However, later analyses using register-based data across the entire educational spectrum have shown that for cohorts born from the 1960s onward, intergenerational persistence has again strengthened (Härkönen and Sirniö 2020; Karhunen and Uusitalo 2017; Lahtinen et al. 2022).

Like other high-income countries, Finland experienced a considerable educational expansion during the twentieth century. This expansion stalled and partly reversed for cohorts born in the mid-1970s and thereafter (Härkönen and Sirniö 2020; Pekkarinen 2012). Since the baby boomer cohorts born between 1945 and 1949 onward, women have attained more education than men (Pekkarinen 2012). Currently, among the working-age population, women are 15 percentage points more likely than men to have completed a postsecondary degree (Official Statistics of Finland 2019), and the

gender gap in school performance is comparatively large (OECD 2019). Finland also manifests a high level of gender segregation regarding the field of study (Pekkarinen 2012:173).

Data and Methods

We used rounds 1992, 1997, 2002, 2007, and 2012 of the epidemiological survey FINRISK (Borodulin et al. 2018), Health 2000/2011 (Lundqvist and Mäki-Opas 2016), and FinHealth 2017 (Borodulin and Sääksjärvi 2019), all collected and maintained by the Finnish Institute for Health and Welfare (THL). FINRISK data were collected across six regions: North Karelia, Northern Savonia, Lapland, Northern Ostrobothnia and Kainuu, Turku and Loimaa, and Helsinki and Vantaa. Health 2000/2011 covered 80 healthcare districts and FinHealth 2017 covered 50 healthcare districts across mainland Finland. All surveys aimed to collect representative samples of the Finnish population, although eastern and northern regions were oversampled. These surveys included clinical examinations, during which deoxyribonucleic acid (DNA) samples were collected from participants. The response rates of the surveys varied between 65% and 93%, averaging 73%, showing the usual negative trend over time. The data were further linked to longitudinal administrative population registers using personal identity codes before a pseudonymized dataset was released to the research team. Phenotype variables used in the analyses were based on either these linked registers or—in the case of age, gender, and region in FinHealth—sampling frame information also drawn from population registers.

The genetic data have gone through quality control and imputation procedures according to SISU version 3 reference panel protocols (Pärn, Fontarnau et al. 2018; Pärn, Isokallio et al. 2018). Eighty-eight percent of the participants have genotyped information available, varying between 84% and 95% by data collection round. Excluding nongenotyped individuals gave us an initial pooled sample of 43,325 individuals.

We included cohorts born between 1925 and 1989 and individuals aged less than 80 years at the time of survey. Education-selective mortality (Mackenbach et al. 2016) may introduce survivorship bias in the older population.

One possible source of confounding in analyses of PGI and various outcomes may stem from shared environments among individuals who are related (Freedman et al. 2004; Mills et al. 2020; Sillanpää 2011). To mitigate this possibility, we excluded one individual from pairs of participants who had a proportion of identity by descent of at least .178, corresponding to the expected lower bound of relatedness of second-degree relatives (aunts/uncles to nieces/nephews, grandparents to grandchildren, or between half-siblings). These restrictions resulted in a sample of 37,416 individuals, which is used in the first part of the analysis. In the second part of the analysis, we restricted the sample to cohorts born between 1955 and 1984 and excluded FinHealth 2017 because of unavailable information on childhood family educational level. This resulted in a sample size of 15,143 individuals. Figure A1 in the online appendix presents flowcharts of the stepwise sample construction and the number of observations.

We calculated the PGI for educational attainment using the summary scores of the GWAS by Okbay and colleagues (2022). Participants in 23andMe data collection

were excluded in these summary scores owing to privacy policies, as well as participants overlapping with our analysis sample to avoid overfitting of PGI. The method of defining the PGI was SBayesR (Lloyd-Jones et al. 2019). This method produced linkage disequilibrium–weighted scores using the summary GWAS scores and an external banded linkage disequilibrium matrix from HapMap3 SNP variants with a minor allele frequency of at least .01 and not strongly deviating from Hardy–Weinberg equilibrium ($p > 10^{-8}$) in our data (see Lloyd-Jones et al. 2019).

The population of Finland is genetically homogeneous, except for a difference between eastern and western regions following the border set in the historic treaty of Pähkinäsaari in 1323 (Kerminen et al. 2017). However, to account for possible population stratification bias induced by genetic similarity within regions, all of our analyses were adjusted for the first 10 principal components of the full matrix of approximately independent SNP variants (Price et al. 2006). Additionally, we adjusted for five regions of residence at the time of the study participation, which are hospital-specific catchment areas for highly specialized medical care.³ Furthermore, to adjust for possible biases stemming from methods of genotyping or data collection, we included the combination of data collection round and genotyping batch as dummy variables for each of our models.

First Analysis Strategy: Level and Trends in Polygenic Prediction of Education, 1925–1989

First, we investigated the association between PGI and education, which is estimated with a set of regressions taking the following form:

$$y_{i,j} = b_0 + b_1 PGI + b_2 Female_i + \overline{b_3 Cohort_i} + b_4 PGI \times Female_i + \overline{b_5 PGI_i \times Cohort_i} + \overline{b_6 Female_i \times Cohort_i} + \overline{b_7 PGI_i \times Female_i \times Cohort_i} + \overline{b_8 Controls_i}, \quad (1)$$

where $y_{i,j}$ marks a binary/multinomial logistic link of completing the educational transition j for an individual i . PGI is the standardized (as mean zero, standard deviation one) PGI of education, $Female$ is a dummy variable of female gender, and $Cohort_i$ is a vector of cohort cubic splines (Durrleman and Simon 1989) with four knots.⁴ $Controls_i$ is a vector of covariates, including the first 10 principal components of the genome, dummy variables for region of residence, and dummy variables for the combination of data collection round and genotyping batch. In the beginning of the analysis, we also present results from models with main effects only and models with a $PGI \times Female$ interaction, as well as corresponding linear regression models, where y is the identity link for years of education.

³ See <https://www.stat.fi/en/luokitukset/erva/>, accessed August 16, 2022.

⁴ A four-knot solution was chosen in defining splines because it was preferable relative to three- or five-knot solutions according to both AIC and BIC in multinomial models predicting the basic/vocational/academic secondary transition, as well as linear models predicting years of education. The results with transitions on the tertiary level were more mixed, but the four-knot solution performed relatively well in all cases.

After establishing these associations over the whole study period, we focused on trends across cohorts. We evaluated to what extent changes in the total association could be attributable to different transitions, and to what extent these transitions, in turn, are attributable to distributional changes and the change in the associations of PGI in predicting the transitions. We used the method developed by Buis (2017) and Härkönen and Sirniö (2020), where the association of the PGI on completed education is expressed as a sum of the weighted associations of PGI on each educational transition, that is:

$$\frac{\delta E(L_i)}{\delta PGI_i} = \sum_j effect_{i,j} \times weight_{i,j}, \quad (2)$$

where $\delta E(L_i)$ is the change in expected level of educational attainment for an individual i . The *effect* parameter is the log odds ratio of PGI from the model presented in Eq. (1) for the educational transition j . These are again predicted separately by men and women and for different two-year cohorts (1925–1926, 1927–1928, and so forth). The weight parameter, in turn, is a product of three components:

$$weight_{i,j} = atrisk_{i,j} \times differentiate_{i,j} \times gain_{i,j}. \quad (3)$$

Here, the *atrisk* parameter refers to the proportion of the population at risk of the transition j . For the transition between basic and secondary education, it is assumed to be one. For transitions between academic secondary and lower/higher tertiary degrees, it is the estimated proportion of those who have completed an academic secondary degree. For a transition between vocational and tertiary degrees, it is the proportion that have completed vocational education. These cohort- and gender-specific estimates are obtained from the multinomial logit model presented in Eq. (1), where y_i is the multinomial log link between basic/academic secondary/vocational degrees. The estimates of being at risk are further transformed into a probability scale holding the PGI and covariates at their observed values (Hanmer and Ozan Kalkan 2013).

The *differentiate* parameter is the variance of the estimated probability p of an individual i passing the transition j : $differentiate_{i,j} = p_{i,j} \times (1 - p_{i,j})$. These probabilities are again obtained using the model in Eq. (1), but while the *atrisk* parameter was estimated from the previous transition (or it was assumed to be one for the first transition), these probabilities were specifically for the binary/multinomial logit model of the transition in question.

The *gain* parameter of Eq. (3) is the expected increase in years of education when completing the transition in question. For the transition between vocational and tertiary education, the gain was assumed to be four years. For completing other transitions, these cohort- and gender-specific gains were obtained using the following linear regression model among those at risk on the transition in question:

$$\begin{aligned} eduyears_i = & b_0 + b_1 \overline{Transit_i} + b_2 \overline{Female_i} + b_3 \overline{Cohort_i} + b_4 \overline{Transit_i} \\ & \times \overline{Female_i} + b_5 \overline{Transit_i \times Cohort_i} + b_6 \overline{Female_i \times Cohort_i} \\ & + b_7 \overline{Transit_i \times Female_i \times Cohort_i} + b_8 \overline{PGI_i} + b_9 \overline{Controls_i} + \epsilon_i, \quad (4) \end{aligned}$$

where $eduyears_i$ is the (expected) years of education of an individual after the basic level given their educational path. These years were based on the specification by Härkönen and Sirniö (2020) and specified in Figure 1 (basic = 0 years; basic \rightarrow vocational secondary = 2 years; basic \rightarrow academic secondary = 3 years; basic \rightarrow academic secondary \rightarrow lower tertiary = 6 years; basic \rightarrow academic secondary \rightarrow higher tertiary = 8 years; basic \rightarrow vocational secondary \rightarrow lower tertiary = 6 years). $Transit_i$ refers to a dummy variable that indicates whether an individual i completed the transition in question and ϵ_i is the residual term. Other variables are the same as in the linear specification of Eq. (1). Cohort- and gender-specific confidence intervals for effect, weight, and total parameters were obtained through bootstrap simulations with a non-parametric percentile method (1,501 replications).

Table A1 in the online appendix shows descriptive statistics of the sample included in the first analysis. PGI and years of education are moderately correlated (.27). As expected, the average level of education increased across cohorts, but PGI does not show substantial differences across genders or cohorts. Owing to regionally clustered sampling, the eastern Kuopio region is overrepresented relative to the population size, and the Helsinki and Tampere regions in the south are correspondingly underrepresented.

Second Analysis Strategy: Origin–Genome–Destination Decomposition, Cohorts 1955–1984

In the second phase, we conducted origin–genome–destination (*OGD*) decompositions. This analysis exploited partial Spearman's rank-order correlations with the method proposed by Liu and colleagues (2018). For the total analysis sample and each five-year cohort, we estimated five different partial correlations, which were depicted in Figure 2: (1) between education of origin and education of destination (*OD*), (2) between education of origin and education of destination adjusted by PGI of education (*OD* | *G*), (3) between education of origin and PGI of education (*OG*), (4) between PGI of education and education of destination (*GD*), and (5) between PGI of education and education of destination adjusted by education of origin (*GD* | *O*).

We estimate these correlations, first, for all cohorts born between 1955 and 1984 combined and, second, stratified into five-year cohorts. Because of more limited statistical power and the absence of major gender differences among these cohorts, this analysis was conducted for men and women combined. Theoretically, the mechanisms of intergenerational reproduction of education should work in relatively similar ways between men and women (Breen et al. 2010). Previous evidence also suggests only small differences between genders, albeit a slightly stronger increase in the origin–education association among women than men (Härkönen and Sirniö 2020). The parental education was measured using the dominance approach, that is, using the highest completed degree in the childhood household, measured every five years between 1970 and 1985. Hierarchical ranks for education were (1) basic; (2) secondary degree; (3) lowest tertiary; (4) lower tertiary; and (5) higher tertiary, following the official classification levels of education by Statistics Finland.⁵ Since information

⁵ See https://www.stat.fi/meta/kas/koulutusaste_en, accessed July 4, 2022.

on the type of secondary degree of parental generation was unavailable in cases of completed tertiary degrees, we did not use such information in the second analysis. In this analysis, we ranked the PGI of education into 20 five-percentile quantiles to create an ordinal variable suitable for the analysis method and to relax assumptions regarding its functional form. All correlations were also adjusted by the first 10 principal components of the genome, five-year cohorts (if not stratified by them), a gender dummy variable, region of residence, and the genotyping batch–data collection round combination.

Table A2 in the online appendix shows the descriptive statistics of the sample included in the second phase of the analysis. The latter cohorts have smaller sample sizes in this analysis since younger cohorts were not included in the earlier rounds of FINRISK or Health 2000/2011 data collections because they were underage at the time of data collection. Likely owing to the inclusion of a narrower range of cohorts, PGI and years of education show a slightly larger correlation than in the first analysis sample (.29). Again, as expected, PGI does not show substantial variation across genders or cohorts, and no systematic differences between parental education between genders.

We used Stata 16.1 software in the first analysis and the *PResiduals* package in R 4.0.3 in the second. SBayesR was conducted with *GTCB 2.03 beta*. Resulting PGI scores were attached, and principal components and relatedness were estimated with *Plink 1.9–2.0*.

Results

Trends in Polygenic Prediction of Education

Table 1 presents the regression results of the PGI in predicting years of education and different educational transitions. Model A indicates that a one-standard-deviation-higher PGI is associated with 0.74 (95% confidence interval (CI), 0.71–0.78) additional years of education among men. Women attained 0.36 (95% CI, 0.31–0.41) more years of education than men on average, and the number of years of the education–PGI association was only negligibly lower among women. When inspecting different educational transitions, we observed that the PGI discriminated most strongly in the transition between basic and secondary education, and more strongly for the academic secondary track (odds ratio [OR], 2.57, 95% CI, 2.44–2.71, among men; OR, 2.24, 95% CI, 2.14–2.35, among women)⁶ than the vocational track (OR, 1.40, 95% CI, 1.34–1.46, among men; OR, 1.26, 95% CI, 1.21–1.32, among women). In addition, the PGI was moderately associated with the academic secondary → higher tertiary transition (OR, 1.51, 95% CI, 1.38–1.66, among men; OR, 1.63, 95% CI, 1.51–1.75, among women), as well as the vocational → tertiary transition (OR, 1.48, 95% CI, 1.40–1.56, among men; OR, 1.38, 95% CI, 1.31–1.46, among women).

⁶ Odds ratios among women are calculated by multiplying the PGI and PGI × female parameters. Related confidence intervals are estimated with the delta method.

Table 1 Associations between polygenic index (PGI) of education and years of education (linear coefficients) and between PGI and educational transitions (odds ratios), with gender interactions: Linear (A), multinomial logistic (B and C), and binary logistic (D) regression models

	A: Years of Education	B: Academic Secondary/ Vocational Given Basic Degree		C: Lower/Higher Tertiary Given Academic Secondary Degree		D: Tertiary Given Vocational Degree
		Vocational	Academic Secondary	Lower Tertiary	Higher Tertiary	
PGI of Education	0.74*** (0.71–0.78)	1.40*** (1.34–1.46)	2.57*** (2.44–2.71)	0.97 (0.89–1.06)	1.51*** (1.38–1.66)	1.48*** (1.40–1.56)
Female	0.36*** (0.31–0.41)	0.97 (0.92–1.03)	1.91*** (1.79–2.05)	1.42*** (1.27–1.58)	0.90† (0.80–1.01)	1.01 (0.95–1.09)
PGI × Female	−0.03 (−0.08 to 0.02)	0.90*** (0.85–0.96)	0.87*** (0.81–0.93)	1.14* (1.02–1.28)	1.08 (0.96–1.21)	0.94† (0.87–1.01)
N	37,416	37,416		11,382		16,373

Notes: All models include year of birth cubic splines, 10 first principal components of the genome, region of residence, and combination of data collection round and genotyping batch. Linear Model A reports conventional regression coefficients (with Huber–White standard errors), multinomial logistic Models B and C report multinomial odds ratios, and binary logistic Model D reports odds ratios. 95% confidence intervals are shown in parentheses.

†*p* < .10; **p* < .05; ****p* < .001

Figure 3 shows the multiple pathways sequential logit decomposition of the contribution of different educational transitions to the association between the PGI and years of education attained across cohorts. Among both men and women, we observe a curvilinear pattern, where the association first strengthened, then peaked among men in the 1945–1946 cohort and among women in the 1949–1950 cohort, and declined until cohorts born roughly in the mid-1960s. After these cohorts, the association between PGI and education was rather stable. The increase among early cohorts was stronger among women (from 0.36, 95% CI, 0.23–0.51, in 1925–1926 to 0.91, 95% CI, 0.84–0.98, in 1949–1950) than among men (from 0.66, 95% CI, 0.49–0.88, in 1925–1926 to 0.92, 95% CI, 0.85–1.01, in 1945–1946). For the most recent studied cohorts, the total association was only trivially higher among women (e.g., in the 1981–1982 cohort, 0.78, 95% CI, 0.66–0.93, among women, and 0.74, 95% CI, 0.60–0.90, among men).

When examining the contribution of specific transitions on the overall PGI–education association in Figure 3 (see Figure A2 in the online appendix for CIs of specific transitions), the transition from basic to academic secondary school was the most discriminating, accounting for 60–80% of the total net association for most cohorts. Exceptions were women born in the 1980s (50–60%) and men born between the mid-1960s and mid-1970s (just above 80%). The transition between academic secondary and higher tertiary education became more distinctive across cohorts, particularly among women, contributing to about 20% of the total net association among women in the 1970s and 1980s cohorts (*cf.* <10% contribution among the

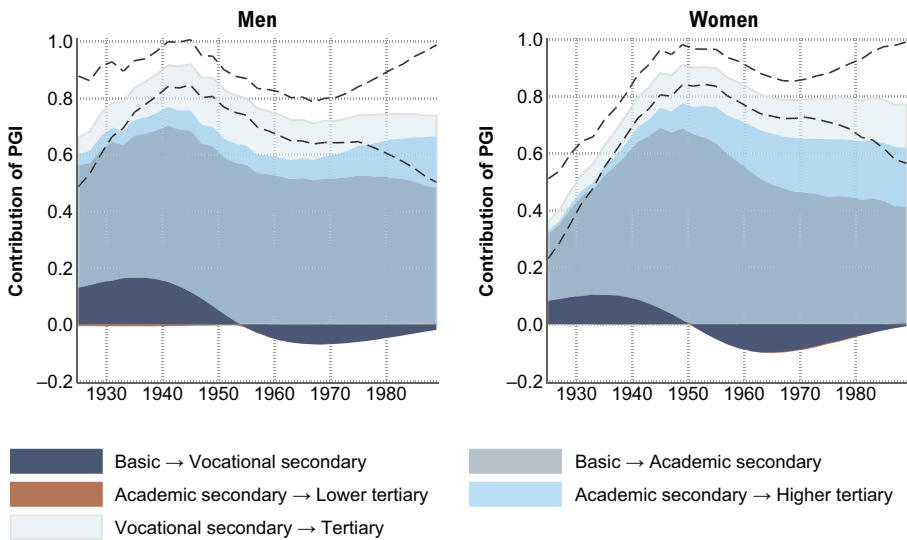


Fig. 3 Contribution of educational transitions to the total PGI–education association by cohort and gender. The graphs show the results of multiple pathways sequential logit decomposition, where the total contribution of the effect of PGI is decomposed with educational transitions. The y-axis indicates weighted log odds ratio of a one-standard-deviation increase in PGI on each transition. Dashed lines represent 95% confidence intervals of the total contribution of all transitions (confidence intervals of the specific transitions are presented in Figure A2 in the online appendix). The contribution of the transition between academic secondary and lower tertiary degrees is negligible and can hardly be distinguished.

pre-1950s cohorts). The transition between vocational and tertiary education had a relatively stable contribution to the total PGI–education association among both men and women (about 10–20% of the total net contribution among most cohorts). The transition from basic to vocational education had a substantial positive contribution to the PGI–education relationship among pre-1950s cohorts, whereas among most cohorts born after 1950, it became negative. The negative contribution peaked among the 1969–1970 cohort (−0.07, 95% CI, −0.11 to −0.03) among men and the 1965–1966 cohort (−0.10, 95% CI, −0.14 to −0.06) among women, after which it started to converge toward zero. Finally, across all studied cohorts, the contribution of the transition between academic secondary and lower tertiary degrees is negligible and virtually indistinguishable in [Figure 3](#).

[Figure 4](#) presents the decomposition of these transitions into their effect and weight parameters (see [Figure A3](#) in the online appendix for further decomposition of the weight parameters). It should be emphasized that the “effect parameter” refers to a technical term in the multiple pathways model—that is, the regression coefficient of the corresponding (multiple) logit model—and is not necessarily causal (see [Buis 2017:652–655](#)). The negative contribution of the basic to vocational secondary path in the cohorts from 1950s onward was a result of the decline of its weight parameter, and further decomposition (see appendix [Figure A3](#)) revealed that this was because the associated gain of years of education became negative. This means that completed vocational education predicted a lower number of final completed years in education relative to those who stayed in basic or completed academic secondary education.

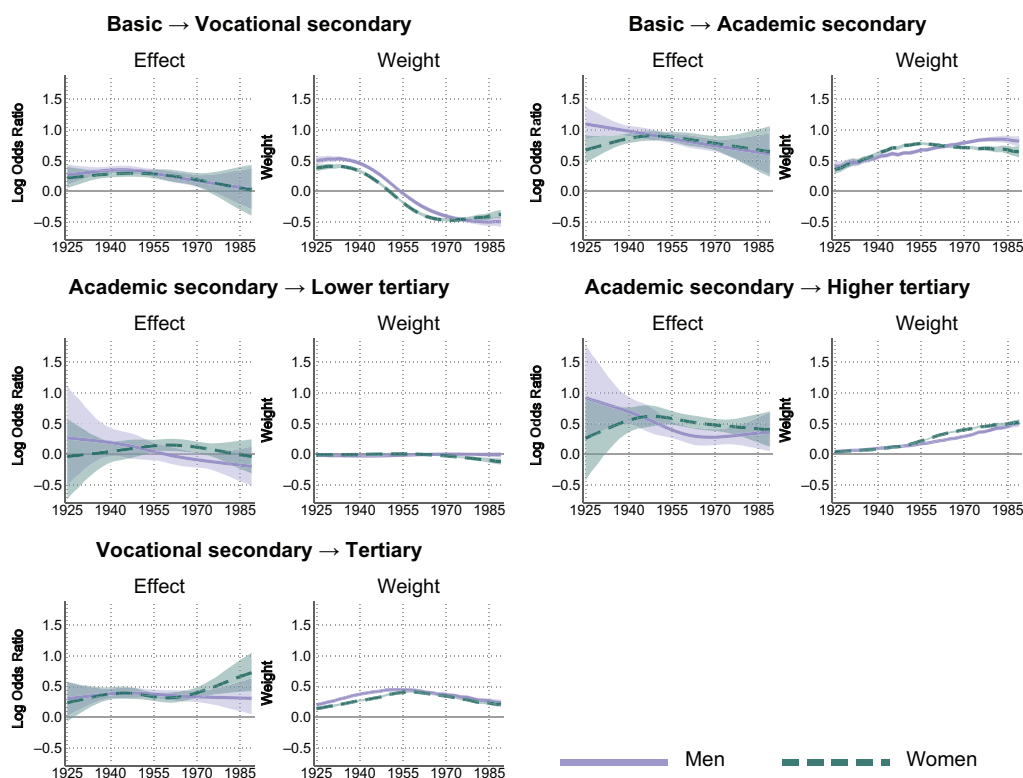


Fig. 4 Decomposition of the total contributions of specific educational transitions into the effect and weight components. Effect parameters are cohort- and gender-specific log odds ratios. “Effect parameter” is a technical term in the multiple sequential model and does not imply causality. Weight parameters are the product of three components (*atrisk* × *differentiate* × *gain*), which are presented in Figure A3 in the online appendix. The total contributions of each transition (i.e., *effect* × *weight*) are found in Figure A2. Shaded areas are 95% confidence intervals.

Among later cohorts, the convergence back toward zero was a result of the decline of the effect parameter. That is, the PGI did not predict completion of vocational education relative to those with only basic education.

Figure 4 also shows that the decline in the relative contribution of academic secondary education after the turn of the 1950s was a result of an almost linear decline in its effect parameter. Its weight, in turn, grew among men, which was a result of increasing differentiation (see online appendix Figure A3). Among women, the weight parameter declined only slightly owing to a decline in the gain associated with it. The increase in the contribution of higher tertiary education stems from its steadily increasing weight rather than its effect parameter. Among men, the effect parameter of the PGI was higher among earlier cohorts, and among women the pattern of the effect parameter was curvilinear, peaking in the 1949–1950 cohort. The increase in the weight parameter was attributable to the increased proportion of the population at risk of this transition (see Figure A3). Finally, the transition between academic secondary and lower tertiary education had both small effect and weight parameters

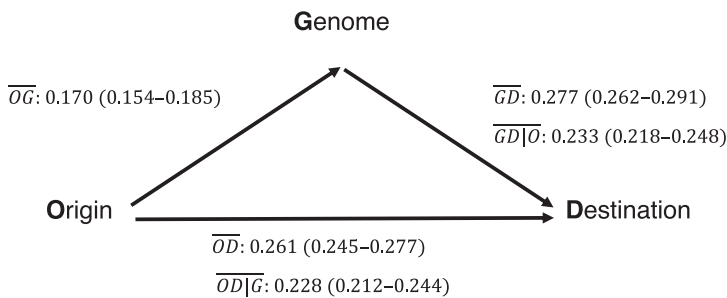


Fig. 5 Pathways of origin–genome–destination (*OGD*) triangle decomposition, cohorts 1955–1984 ($N = 15,143$). Estimates are partial Spearman’s rank-order correlations, with 95% confidence intervals in parentheses. All correlations are partial for five-year birth cohorts, gender, 10 first principal components of the genome, region of residence, and combination of data collection round and genotyping batch. *OG*: origin–genome; *GD*: genome–destination; *GD | O*: genome–destination adjusted by origin; *OD*: origin–destination; *OD | G*: origin–destination adjusted by genome. “Origin” refers to childhood household education, “destination” to individual’s education, and “genome” to the PGI of education.

across all cohorts, resulting in a negligible total contribution that was already observed in Figure 3. The transition between vocational and tertiary education had a nonnegligible and relatively steady effect parameter. There was modest curvilinearity in the weight parameter, which peaked among the 1950s cohorts.

Origin–Genome–Destination Decomposition

Figure 5 presents the decomposition analysis of the *OGD* triangle among cohorts born between 1955 and 1984. The weakest pathway was between education of origin and PGI of education (*OG*, Spearman’s partial $\rho = 0.17$, 95% CI, 0.15–0.19). The partial correlations between education of origin and education of destination (*OD*, $\rho = 0.26$, 95% CI, 0.25–0.28) and between PGI and education of destination were close to each other (*GD*, $\rho = 0.28$, 95% CI, 0.26–0.29). Adjusting for the PGI attenuated the *OD* pathway by 13%, while adjusting for parental education attenuated the *GD* pathway by 16%.

Figure 6 displays the corresponding analysis as in Figure 4 but in five-year birth cohorts. The strongest trend was an increase of the partial correlation between origin and PGI, which almost doubled from 0.14 (95% CI, 0.11–0.17) among the 1955–1959 cohort to 0.25 (95% CI, 0.20–0.30) among the 1975–1979 cohort. There was also an increase in the partial correlation between education of origin and destination, from 0.24 (95% CI, 0.20–0.27) in the 1955–1959 cohort to 0.30 (95% CI, 0.25–0.35) in the 1975–1979 cohort. It should be noted, however, that these trend estimates are relatively noisy. A substantial part of this increase was attenuated when adjusting for the PGI. Quantifying the extent of this attenuation is not straightforward and depends on measurement. The decline of correlation estimates between highest and lowest cohorts (1975–1979 vs. 1955–1959) was one quarter ($[0.253 - 0.204] / [0.302 - 0.236] = 0.742$). The decline of linear slopes fit between the point estimates in adjusted and unadjusted models was two thirds ($0.001971 / 0.005542 = 0.356$). As

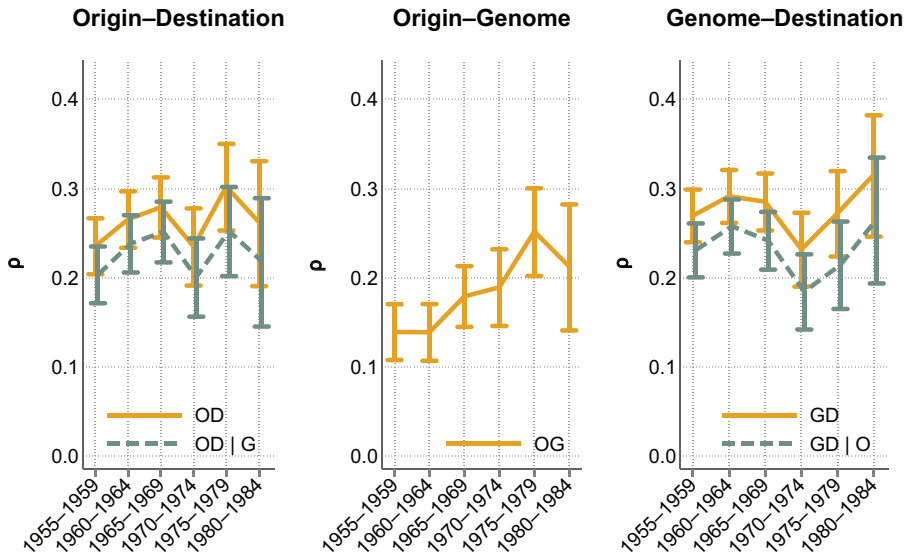


Fig. 6 Pathways of origin–genome–destination (*OGD*) triangle decomposition by five-year birth cohorts 1955–1984. The y-axis indicates partial Spearman correlations stratified in five-year birth cohorts, as presented on the x-axis. All correlations are adjusted by gender, 10 first principal components of the genome, region of residence, and combination of data collection round and genotyping batch. Whiskers represent 95% confidence intervals. Pathways: *OG*: origin–genome; *GD*: genome–destination; *GD | O*: genome–destination adjusted by origin; *OD*: origin–destination; *OD | G*: origin–destination adjusted by genome. “Origin” refers to childhood household education, “destination” to individual’s education, and “genome” to the PGI of education.

in the first analysis, the PGI–education association did not show a consistent trend among the cohorts 1955–1984.

Discussion

Genetic Prediction of Education

In this study, we have investigated the level and changes in polygenic prediction of education across Finnish cohorts from 1925 to 1989 using population-representative, genetically informed epidemiological surveys linked to administrative registers. We observed that a standard-deviation-larger PGI of educational attainment is associated with 0.74 years of additional education (see [Table 1](#)). When decomposing this overall association between PGI and education, the transition between comprehensive and secondary level education, particularly the transition to academic secondary track, was the most important. Stronger stratification in earlier educational transitions is in line with the general pattern observed in the literature studying the effects of various sociodemographic factors, including parental education ([Härkönen and Sirniö 2020](#); [Mare 1980](#)). However, a previous study by [Conley and Domingue \(2016\)](#) did not observe such a clear-cut relationship between the education PGI and educational

transitions in the United States, though this may be due to imprecision from a smaller sample size or less predictive earlier-generation PGI. The PGI in our study predicted the choice of academic track more strongly than more practically oriented tracks (vocational or lower tertiary) at both secondary and tertiary levels. This is suggestive of mechanisms via which the PGI can manifest in characteristics that are associated with educational attainment, pointing toward pursuing academic skills and interests. It further resonates with previous results relating the education PGI with genetic variants that are predominantly expressed in the brain on a cellular level and via executive functioning on a psychological level (for a review, see Harden 2021: chapter 7), or with findings that individuals with a higher PGI tend to choose more demanding mathematics curriculums in U.S. high schools (Harden et al. 2020).

In the analysis on the trends across cohorts, we observed a curvilinear pattern in polygenic prediction (see Figure 3). Among women, the increase in early cohorts was stronger than for men. This points to a possible explanation that societal obstacles hindering the educational achievement of women have gradually eased over time. Other studies have identified various potential obstacles for women, such as social norms and values favoring the education of male offspring or a lower return of education due to more limited labor market opportunities and more responsibilities in the (unpaid) domestic sphere (DiPrete and Buchmann 2013; Kaarninen 1995; Pekkarinen 2012). Correspondingly, an increase of polygenic prediction has been observed across cohorts among U.S. women (Herd et al. 2019).

The association between PGI and education peaked in the baby boomer cohorts born in the latter half of the 1940s. From 1950s cohorts onward, the total contribution was similar between men and women, decreasing at first and then remaining stable for cohorts born after the 1960s. Finnish cohorts born between 1960 and 1966 experienced a comprehensive school reform that was beneficial to the educational attainment of women and those from disadvantaged socioeconomic backgrounds (Pekkala Kerr et al. 2013; Pekkarinen 2008; Pekkarinen et al. 2009). Future studies could examine in more detail whether the school reform was associated with the polygenic prediction of education.

Among later cohorts, a higher PGI did not predict differences between vocational and basic education (see Figure 4). This may imply that among the minority of individuals with only basic qualifications, there exist structural (e.g., social exclusion of their childhood families) or tangible (e.g., a dramatic life event leading to school dropout) obstacles preventing the completion of further education. This finding may also reflect more limited educational opportunities for earlier-born cohorts, among whom PGI was a predictor of any kind of education. Among later cohorts with a wider range of educational opportunities, the prediction has shifted more specifically in academically oriented degrees.

For the earlier-born cohorts, the transition between basic and academic secondary school was clearly the most important in accounting for the association between the education PGI and educational attainment for both men and women. For the later cohorts of women, however, the transition between academic secondary and higher tertiary (university) education became almost as important. In addition, the transition from vocational school to tertiary education had a relatively large contribution to the total association between PGI and education for both men and women. This observation has policy implications as it suggests that avoiding dead ends in the educational

tracks can have substantial benefits in terms of ensuring equality of opportunity and the effective allocation of human capital (see also Pfeffer 2008; Schindler 2017).

The Role of Genetics in the Intergenerational Transmission of Education

For a subsample of cohorts born between 1955 and 1984, we also studied the role that polygenic prediction plays in the overall intergenerational transmission of education. The intergenerational partial rank-order correlation of education was 0.26, and the PGI–education partial rank-order correlation was 0.28 (see Figure 5). The intergenerational correlation of education is in line with findings from previous studies (Härkönen and Sirniö 2020; Hertz et al. 2008; Karhunen and Uusitalo 2017),⁷ and the relation in explanatory power between parental education and PGI is comparable to what was observed by Lee et al. (2018). The partial correlation between the higher educated parent and the PGI of offspring was 0.17, which is just above half of the PGI–(own) education correlation. This is roughly what should be expected given that individuals share on average half of their genes with each parent, plus some effect stemming from educational assortative mating and indirect genetic effects (Domingue et al. 2014; Mäenpää 2014; Morris et al. 2020). Adjusting for the PGI attenuated the association between education of origin and destination by one eighth, consistent with previous findings from the United States (Conley et al. 2015; Lin 2020; Liu 2018) and Norway (Isungset et al. 2021). After taking into account that this attenuation estimate is likely to be conservative because of “hidden heritability” (i.e., that the PGI explains only part of the total heritability of education), the results imply that common genetic factors contribute to the total intergenerational transmission of education in a way that is considerable, but far from overwhelming. Instead, the results suggest two important, and for a substantial part, independent pathways of inheritance of education: genetic and social. This is also consistent with twin studies, which have demonstrated both substantial genetic and shared environmental components in education (Branigan et al. 2013; Silventoinen et al. 2020).

Finally, in line with previous studies (Härkönen and Sirniö 2020; Karhunen and Uusitalo 2017; Lahtinen et al. 2022), we observed an increase in association between education of origin and destination across cohorts 1955–1984 in Finland (see Figure 6). As a novel contribution, we observed that a substantial part of this increase was attenuated after adjusting for the PGI. According to the view that high genetic heritability of education reflects high equality of opportunity (e.g., Conley 2016; Heath et al. 1985; Krapohl and Plomin 2016), such an observation is reassuring in that a strengthening intergenerational association of education does not necessarily indicate a lack of meritocratic achievement. A more detailed path decomposition revealed that this attenuation was due to the strengthening of the association between education of origin and the PGI. It suggests a strengthening of meritocratic achievement, although not in the generation of individuals studied but in their parents’ generation. This is consistent with our first analysis, which indicated a growing predictive power of PGI

⁷ Although in our case, the adjusted nature of correlation makes the coefficient slightly weaker than in previous studies.

among earlier cohorts. However, increasing educational homogamy has also been observed in Finland (Mäenpää and Jalovaara 2015), which can strengthen the origin–genome pathway. Such homogamy may be indicative of increasing social closure between educational groups and thus brings a qualification to the claim of improvement in the equality of opportunity.

Methodological Considerations

Strengths of this study include the use of genetically informed, population-representative samples with relatively high response rates. We used genetic measurements in conjunction with longitudinal register-based data, which provided accurate phenotype measurements. The quality of phenotypic measurements has been identified as an area in need of improvement in genetically informed analyses of health-related and social scientific outcomes (Mills and Tropf 2020).

A general weakness of PGIs is that they still capture only a part of the total heritability of outcomes. There is no particular reason to believe that the hidden heritability not captured in our PGI will have differing time trends than what we observed here, though a more predictive PGI would have provided us with greater statistical power. A related issue is that GWASes produce predictive scores in a “brute force” manner, and the mechanisms via which the genetic variants influence education remain ambiguous. The original GWAS used to construct the PGI was based on a meta-analysis of studies from many different national and temporal contexts (Conley 2016). While this may mean that the PGI picks up associations that are relatively robust to changes over time, it could also be weighted by factors that do not apply to Finnish society or to a given cohort studied (Tropf et al. 2017). A potentially fruitful avenue for future studies could be to conduct cohort-specific GWASes that allow for assessing whether the genetic variants that are important for education change over time and across cohorts. Observing such a pattern would suggest changes in traits valued by educational institutions or in wider society.

Our results may also be influenced by “genetic nurture,” or indirect genetic effects (Kong et al. 2018; Morris et al. 2020). Offspring inherit both their genes and their environments from their parents. Inclusions of environmental and social effects in PGIs can inflate correlations between genotype and phenotype, and this is seen particularly strongly for social phenotypes such as educational attainment. When comparing conventional and within-sibship estimates, Howe et al. (2022) observed 47% shrinkage of the estimate between education PGI and achieved education. We addressed this bias by randomly excluding close relatives and adjusting for principal components of genetic structure as well as the region of residence (Kerminen et al. 2017), but these procedures are likely to resolve bias only partly (Haworth et al. 2019). However, it is unclear whether such inflation stemming from genetic nurture, or other forms of genetic population stratification, biases the trend estimates. Future studies with genotypic and richly phenotyped data on siblings or families over time may be better equipped to study the extent to which environmental or social effects are being captured and offer novel possibilities investigating intergenerational associations in education.

Despite using methods that are relatively robust against distributional shifts, they have their limitations. For example, the lower predictive power of PGI among

earlier cohorts may partially stem from the fact that a large share of both parents and children had no qualifications beyond basic education. However, it is debatable how much this is a nuisance versus a substantial feature of the topic studied; PGI of education arguably concretely matters less when educational opportunities are poorly available.

Conclusion

In this study, we observed that the genetic predisposition of education as captured by a PGI, despite itself not being a subject of significant distributional change in our data, varies in its predictive power across cohorts in Finland. We detected low genetic predictiveness of education among early cohorts of women, as well as that, among later cohorts, those with only basic education did not have a lower PGI of education than those with vocational education, on average. In addition, we have shed light on a puzzle raised in the literature, namely, why the intergenerational correlation of education has grown again after the 1960s Finnish cohorts (Härkönen and Sirniö 2020; Karhunen and Uusitalo 2017; Lahtinen et al. 2022). According to our results, this can be partly attributed to the strengthened pathway between parental education and offspring's PGI.

In line with recent calls from many prominent sociologists (e.g., Conley 2016; Freese 2018; Herd et al. 2019; Mills and Troup 2020), our results demonstrate that—despite the relatively reluctant history in adopting biological explanations of behavior in the field—bringing genetic evidence explicitly to the table need not be at odds with sociological inquiry, but can be used precisely in elaborating and highlighting the importance of social relationships. For complex societal outcomes such as education, genetic predisposition necessarily manifests under societal and institutional constraints that change over time.

In addition, our analysis underlined that integrating genetic knowledge into social sciences is a two-way street. Just as social scientists would be ill-advised to ignore the important role of genetics in the intergenerational transmission of education, scholars specifically interested in genetic influences should not try to reinvent the wheel in addressing educational attainment, or other phenotypes related to demographic events, socioeconomic attainment, or inequalities in health. Instead, implementing theoretical and methodological insights refined in disciplinary traditions of social scientific research—such as multiple pathways sequential logit models and origin–genome–destination triangle decompositions introduced in this study—has much to offer for the burgeoning understanding of genetics in shaping social outcomes. ■

Acknowledgments The authors would like to give special thanks to Aysu Okbay for providing GWAS summary results excluding Finnish samples, Outi Sirniö for methodological guidance in the multiple pathways sequential logit decomposition analysis, Teemu Palviainen for methodological guidance in defining polygenic index, Anni Savinainen for assistance in the literature review, Katri Kantojärvi for guidance in using genetic data, Anne Juolevi for providing information regarding data collection, and Lempi Lahtinen-Parmala for assistance in the final proofreading. 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.

H.L. was supported by the Academy of Finland (grant 345219). P.M. was supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant 101019329), the Strategic Research Council 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 Erkkö Foundation, the Max Planck Society, the University of Helsinki, and the cities of Helsinki, Vantaa, and Espoo. The study does not necessarily reflect the European Commission's views and in no way anticipates the Commission's future policy in this area. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Ayorech, Z., Krapohl, E., Plomin, R., & von Stumm, S. (2017). Genetic influence on intergenerational educational attainment. *Psychological Science*, 28, 1302–1310.
- Becker, J., Burik, C. A. P., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., . . . Okbay, A. (2021). Resource profile and user guide of the Polygenic Index Repository. *Nature Human Behaviour*, 5, 12. <https://doi.org/10.1038/s41562-021-01119-3>
- Bernardi, F., & Ballarino, G. (Eds.). (2016). *Education, occupation and social origin: A comparative analysis of the transmission of socio-economic inequalities*. Cheltenham, UK: Edward Elgar Publishing.
- Bernelius, V., & Kauppinen, T. M. (2012). School outcomes and neighbourhood effects: A new approach using data from Finland. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Neighbourhood effects research: New perspectives* (pp. 225–247). Dordrecht, the Netherlands: Springer Science+Business Media.
- Björklund, A., & Jäntti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: A comparison of four approaches. *Research in Social Stratification and Mobility*, 70, 100455. <https://doi.org/10.1016/j.rssm.2019.100455>
- Björklund, A., & Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 201–247). Amsterdam, the Netherlands: North-Holland.
- Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York, NY: John Wiley & Sons.
- Borodulin, K., & Sääksjärvi, K. (2019). *FinHealth 2017 Study—Methods* (THL Report No. 17/2019). Helsinki, Finland: Finnish Institute for Health and Welfare.
- Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S., . . . Vartiainen, E. (2018). Cohort profile: The National FINRISK study. *International Journal of Epidemiology*, 47, 696–696. <https://doi.org/10.1093/ije/dyx239>
- Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces*, 92, 109–140.
- Breen, R. (2010). Educational expansion and social mobility in the 20th century. *Social Forces*, 89, 365–388.
- Breen, R., & Jonsson, J. O. (2005). Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility. *Annual Review of Sociology*, 31, 223–243.
- Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2009). Nonpersistent inequality in educational attainment: Evidence from eight European countries. *American Journal of Sociology*, 114, 1475–1521.
- Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2010). Long-term trends in educational inequality in Europe: Class inequalities and gender differences. *European Sociological Review*, 26, 31–48.
- Breen, R., & Müller, W. (Eds.). (2020). *Education and intergenerational social mobility in Europe and the United States*. Stanford, CA: Stanford University Press.
- Buis, M. L. (2017). Not all transitions are equal: The relationship between effects on passing steps in a sequential process and effects on the final outcome. *Sociological Methods & Research*, 46, 649–680.
- Bukodi, E., & Goldthorpe, J. H. (2016). Educational attainment—relative or absolute—as a mediator of intergenerational class mobility in Britain. *Research in Social Stratification and Mobility*, 43, 5–15.
- Colodro-Conde, L., Rijdsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in educational policy and the heritability of educational attainment. *PLoS One*, 10, e0143796. <https://doi.org/10.1371/journal.pone.0143796>

- Conley, D. (2016). Socio-genomic research using genome-wide molecular data. *Annual Review of Sociology*, 42, 275–299.
- Conley, D., & Domingue, B. (2016). The bell curve revisited: Testing controversial hypotheses with molecular genetic data. *Sociological Science*, 3, 520–539.
- Conley, D., Domingue, B. W., Cesarini, D., Dawes, C., Rietveld, C. A., & Boardman, J. D. (2015). Is the effect of parental education on offspring biased or moderated by genotype? *Sociological Science*, 2, 82–105.
- Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J., & Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour*, 2, 117–125.
- Debeer, D., Buchholz, J., Hartig, J., & Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *Journal of Educational and Behavioral Statistics*, 39, 502–523.
- DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York, NY: Russell Sage Foundation.
- Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among U.S. adults. *Proceedings of the National Academy of Sciences*, 111, 7996–8000.
- Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8, 551–561.
- Engzell, P., & Troup, F. C. (2019). Heritability of education rises with intergenerational mobility. *Proceedings of the National Academy of Sciences*, 116, 25386–25388.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge, UK: Polity Press.
- Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., . . . Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, 36, 4. <https://doi.org/10.1038/ng1333>
- Freese, J. (2018). The arrival of social science genomics. *Contemporary Sociology*, 47, 524–536.
- Gallego, A. (2007). Unequal political participation in Europe. *International Journal of Sociology*, 37(4), 10–25.
- Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLoS Genetics*, 13, e1006711. <https://doi.org/10.1371/journal.pgen.1006711>
- Gesthuizen, M., Solga, H., & Künster, R. (2011). Context matters: Economic marginalization of low-educated workers in cross-national perspective. *European Sociological Review*, 27, 264–280.
- Grätz, M., Barclay, K. J., Wiborg, Ø. N., Lyngstad, T. H., Karhula, A., Erola, J., . . . Conley, D. (2021). Sibling similarity in education across and within societies. *Demography*, 58, 1011–1037. <https://doi.org/10.1215/00703370-9164021>
- Hanmer, M. J., & Ozan Kalkan, K. (2013). Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models. *American Journal of Political Science*, 57, 263–277.
- Harden, K. P. (2021). *The genetic lottery: Why DNA matters for social equality*. Princeton, NJ: Princeton University Press.
- Harden, K. P., Domingue, B. W., Belsky, D. W., Boardman, J. D., Crosnoe, R., Malanchini, M., . . . Harris, K. M. (2020). Genetic associations with mathematics tracking and persistence in secondary school. *npj Science of Learning*, 5, 1. <https://doi.org/10.1038/s41539-020-0060-2>
- Härkönen, J., & Sirniö, O. (2020). Educational transitions and educational inequality: A multiple pathways sequential logit model analysis of Finnish birth cohorts 1960–1985. *European Sociological Review*, 36, 700–719.
- Haworth, S., Mitchell, R., Corbin, L., Wade, K. H., Dudding, T., Budu-Aggrey, A., . . . Timpson, N. J. (2019). Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications*, 10, 333. <https://doi.org/10.1038/s41467-018-08219-1>
- Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., . . . Nance, W. E. (1985). Education policy and the heritability of educational attainment. *Nature*, 314, 734–736.
- Herd, P., Freese, J., Sciscinski, K., Domingue, B. W., Harris, K. M., Wei, C., & Hauser, R. M. (2019). Genes, gender inequality, and educational attainment. *American Sociological Review*, 84, 1069–1098.
- Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *B.E. Journal of Economic Analysis & Policy*, 7, 2. <https://doi.org/10.2202/1935-1682.1775>

- Hout, M., & DiPrete, T. A. (2006). What we have learned: RC28's contributions to knowledge about social stratification. *Research in Social Stratification and Mobility*, 24, 1–20.
- Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., . . . Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, 54, 581–592.
- Huang, J., van den Brink, H. M., & Groot, W. (2009). A meta-analysis of the effect of education on social capital. *Economics of Education Review*, 28, 454–464.
- Isungset, M. A., Conley, D., Zachrisson, H. D., Yström, E., Havdahl, A., Njølstad, P. R., & Lyngstad, T. H. (2021). *Social and genetic effects on educational performance in early adolescence* (NBER Working Paper 28498). Cambridge, MA: National Bureau of Economic Research.
- Järnefelt, N. (2010). *Education and longer working lives: A longitudinal study on education differences in the late exit from working life of older employees in Finland* (Report No. 2010:1). Helsinki, Finland: Finnish Centre for Pensions.
- Kaarninen, M. (1995). *Nykyajan tytöt: Koulutus, luokka ja sukupuoli 1920-ja 1930-luvun Suomessa* [Modern girls: Education, class and gender in 1920s and 1930s Finland]. Helsinki, Finland: Suomen Historiallinen Seura.
- Karhunen, H., & Uusitalo, R. (2017). 50 vuotta koulutusmahdollisuuksien tasa-arvoa [50 years of equal educational opportunities]. *Yhteiskuntapolitiikka*, 82(3), 296–303.
- Kerminen, S., Havulinna, A. S., Hellenthal, G., Martin, A. R., Sarin, A.-P., Perola, M., . . . Pirinen, M. (2017). Fine-scale genetic structure in Finland. *G3: Genes, Genomes, Genetics*, 7, 3459–3468.
- Kivinen, O., Ahola, S., & Hedman, J. (2001). Expanding education and improving odds? Participation in higher education in Finland in the 1980s and 1990s. *Acta Sociologica*, 44, 171–181.
- Kivirauma, J., & Ruoho, K. (2007). Excellence through special education? Lessons from the Finnish school reform. *International Review of Education*, 53, 283–302.
- Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., . . . Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, 114, E727–E732. <https://doi.org/10.1073/pnas.1612113114>
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsen, B. J., Young, A. I., Thorgeirsson, T. E., . . . Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359, 424–428.
- Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry*, 21, 437–443.
- Lahtinen, H., Martikainen, P., & Tarkiainen, L. (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, 496–526.
- Lahtinen, H., Sirmä, O., & Martikainen, P. (2020). Social class and the risk of unemployment: Trends, gender differences and the contribution of education. *Acta Sociologica*, 63, 303–321.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzi, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50, 1112–1121.
- Lin, M.-J. (2020). The social and genetic inheritance of educational attainment: Genes, parental education, and educational expansion. *Social Science Research*, 86, 102387. <https://doi.org/10.1016/j.ssresearch.2019.102387>
- Linton, R. (1936). *The study of man: An introduction*. New York, NY: D. Appleton-Century Company.
- Liu, H. (2018). Social and genetic pathways in multigenerational transmission of educational attainment. *American Sociological Review*, 83, 278–304.
- Liu, Q., Li, C., Wanga, V., & Shepherd, B. E. (2018). Covariate-adjusted Spearman's rank correlation with probability-scale residuals. *Biometrics*, 74, 595–605.
- Lloyd-Jones, L. R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K. E., . . . Visscher, P. M. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature Communications*, 10, 5086. <https://doi.org/10.1038/s41467-019-12653-0>
- Lundqvist, A., & Mäki-Opas, T. (2016). *Health 2011 Survey—Methods* (Report No. 8/2016). Helsinki, Finland: National Institute for Health and Welfare. Retrieved from <https://www.julkari.fi/handle/10024/130780>
- Mackenbach, J. P., Kulhánová, I., Artnik, B., Bopp, M., Borrell, C., Clemens, T., . . . de Gelder, R. (2016). Changes in mortality inequalities over two decades: Register based study of European countries. *BMJ*, 353, i1732. <https://doi.org/10.1136/bmj.i1732>

- Mäenpää, E. (2014). Homogamy in educational level and parental social class in Finland: A log-linear analysis. *European Sociological Review*, 31, 253–267.
- Mäenpää, E., & Jalovaara, M. (2015). Achievement replacing ascription? Changes in homogamy in education and social class origins in Finland. *Advances in Life Course Research*, 26, 76–88.
- Mare, R. D. (1980). Social background and school continuation decisions. *Journal of the American Statistical Association*, 75, 295–305.
- Mare, R. D. (1981). Change and stability in educational stratification. *American Sociological Review*, 46, 72–87.
- Mills, M. C., Barban, N., & Tropf, F. C. (2020). *An introduction to statistical genetic data analysis*. Cambridge, MA: MIT Press.
- Mills, M. C., & Tropf, F. C. (2020). Sociology, genetics, and the coming of age of sociogenomics. *Annual Review of Sociology*, 46, 553–581.
- Morris, T. T., Davies, N. M., Hemani, G., & Davey Smith, G. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, 6, eaay0328. <https://doi.org/10.1126/sciadv.aay0328>
- Nielsen, F. (2006). Achievement and ascription in educational attainment: Genetic and environmental influences on adolescent schooling. *Social Forces*, 85, 193–216.
- OECD. (2019). *PISA 2018 results (Volume II): Where all students can succeed*. Paris, France: OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>
- Official Statistics of Finland. (2019). *Educational structure of the population* [Data set]. Helsinki: Statistics Finland. Retrieved from http://www.stat.fi/til/vkour/2019/vkour_2019_2020-11-05_tie_001_en.html
- Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., . . . Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54, 437–449.
- Pärn, K., Fontarnau, J. N., Isokallio, M. A., Sipilä, T., Kilpeläinen, E., Palotie, A., . . . Palta, P. (2018). *Genotyping chip data lift-over to reference genome build GRCh38/hg38* (Version 1) [FIMM HumGen sequencing informatics]. *Protocols.io*. <https://dx.doi.org/10.17504/protocols.io.nqtdddwn>
- Pärn, K., Isokallio, M. A., Nunez-Fontarnau, J., Palotie, A., Ripatti, S., & Palta, P. (2018). *Genotype imputation workflow* (Version 3.0.) [FIMM HumGen sequencing informatics]. *Protocols.io*. <https://dx.doi.org/10.17504/protocols.io.nmndc5e>
- Pekkala Kerr, S., Pekkarinen, T., & Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31, 577–602.
- Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment. *Scandinavian Journal of Economics*, 110, 807–825.
- Pekkarinen, T. (2012). Gender differences in education. *Nordic Economic Policy Review*, 2012(1), 165–194.
- Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93, 965–973.
- Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, 24, 543–565.
- Plomin, R. (2019). *Blueprint: How DNA makes us who we are*. Cambridge, MA: MIT Press.
- Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47, 702–709.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909.
- Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R. I., Selzam, S., Dale, P. S., . . . Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature Human Behaviour*, 2, 269–275.
- Savage, M. (2015). *Social class in the 21st century*. London, UK: Pelican.
- Schindler, S. (2017). School tracking, educational mobility and inequality in German secondary education: Developments across cohorts. *European Societies*, 19, 28–48.
- Selita, F., & Kovas, Y. (2019). Genes and Gini: What inequality means for heritability. *Journal of Biosocial Science*, 51, 18–47.

- Sillanpää, M. J. (2011). Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity*, 106, 511–519.
- Silventoinen, K., Jelenkovic, A., Sund, R., Latvala, A., Honda, C., Inui, F., . . . Kaprio, J. (2020). Genetic and environmental variation in educational attainment: An individual-based analysis of 28 twin cohorts. *Scientific Reports*, 10, 12681. <https://doi.org/10.1038/s41598-020-69526-6>
- Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., van der Most, P. J., de Vlaming, R., . . . Mills, M. C. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature Human Behaviour*, 1, 757–765.
- Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, 42, 83–88.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9, 160–164.
- Uusitalo, R. (1999). Return to education in Finland. *Labour Economics*, 6, 569–580.

Hannu Lahtinen (corresponding author)
hannu.lahtinen@helsinki.fi

Lahtinen • Population Research Unit, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland; Max Planck–University of Helsinki Center for Social Inequalities in Population Health, Helsinki, Finland; <https://orcid.org/0000-0003-0910-823X>

Korhonen • Population Research Unit, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland; Max Planck–University of Helsinki Center for Social Inequalities in Population Health, Helsinki, Finland; <https://orcid.org/0000-0001-8499-2008>

Martikainen • Population Research Unit, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland; Max Planck Institute for Demographic Research, Rostock, Germany; Max Planck–University of Helsinki Center for Social Inequalities in Population Health, Helsinki, Finland; <https://orcid.org/0000-0001-9374-1438>

Morris • Centre for Longitudinal Studies, Social Research Institute, University College London, London, UK; <https://orcid.org/0000-0001-8178-6815>