Re-evaluating the effect of the Finnish comprehensive school reform on intergenerational mobility*

Andrea Albanese† Vesa-Matti Heikkuri‡ Stefano Lombardi§

Abstract

We re-evaluate the 1972-1977 Finnish comprehensive school reform with modern difference-in-differences estimators. The reform eliminated the old two-track school system and introduced a nine-year comprehensive school, postponing the vocational vs. academic track choice from age 11 to 16. We find some differences compared to the seminal papers by Pekkarinen (2008) and Pekkarinen et al. (2009) that use the two-way fixed effects estimator, such as the fact that the reform did not reduce the probability of obtaining a tertiary degree for men. However, even though our estimates are generally smaller in magnitude compared to the existing ones, our analysis leads to similar overall conclusions: the reform reduced the intergenerational persistence of socioeconomic status.

JEL codes: I21, I24, I28, J24, J62, E24

Keywords: Intergenerational mobility, Comprehensive school, Education policy, Tracking

*We thank Adrian Adermon, Pietro Biroli, Matti Sarvimäki and Mikko Silliman for useful comments. Permission to use the Military test score data was approved by the Finnish Defence Forces (permit AQ9516).

†Luxembourg Institute of Socio-Economic Research (LISER), Ghent University, UCLouvain, and Institue of Labor Economics (IZA). E-mail: andrea.albanese@liser.lu.

‡Tampere University. E-mail: vesa-matti.heikkuri@tuni.fi

§VATT Institute for Economic Research, Institute for Evaluation of Labor Market and Education Policy (IFAU), IZA, Uppsala Center for Labor Studies (UCLS). E-mail: stefano.lombardi@vatt.fi.

1 Introduction

Understanding how differences in socioeconomic status in one generation are transmitted to the next is a cornerstone topic of interest in economics and social sciences (Becker and Tomes, 1979; Solon, 1992).[1] While a large body of research in economics has focused on how to precisely estimate absolute or relative measures of income mobility, since the early 2000s a growing number of studies have been focusing on the causal pathways that explain such intergenerational correlations and elasticities (for reviews, see Black and Devereux, 2011, and Cholli and Durlauf, 2022).

The analysis of the causes of the persistence of socioeconomic status has deep policy implications. The goal is to answer the question of what can be considered an optimal mobility level, which is a non-trivial task. As already highlighted by Solon (2004), even when equality of opportunity and meritocracy are considered a primary goal of society, perfect mobility would imply null returns to human capital accumulation, leading to a peculiar society where investments in education are not rewarded with higher earnings.

Alternative approaches can help identifying what causes the documented parent-child correlations in education and income. A prominent one analyzes how changes in parental education driven by comprehensive education reforms are transmitted to the offspring.[2,3] The study of education reforms in this context is particularly relevant for a few reasons. First, several European countries implemented similar comprehensive school reforms in the aftermath of the Second World War. By comparing results across countries it is then possible to build scientific knowledge that goes beyond single-country analysis. Second, in each country, these reforms affected the whole population and had the direct goal of improving equality of opportunity. Hence, they offer a formidable test for whether changes in the education system (the "great equalizer") had the intended effects on mobility, and whether these effects were uniform across groups of people.

The focus of our analysis is the comprehensive education reform that was implemented in Finland in 1972–1977 for the cohorts of students born in 1961-1966. In the pre-period, students had to choose between vocational and academic track at age 11. The reform eliminated the old two-track school system and introduced a nine-year comprehensive school, postponing the selection of students to vocational and academic tracks to age 16. The reform was analyzed

[1] Sociologists have been long interested in studying the persistence of occupational prestige, and more recently income, across generations (DiPrete, 2020).

[2] Seminal papers include Meghir and Palme (2005), Holmlund (2007), Aakvik et al. (2010) for Sweden, Pekkarinen (2008), Pekkarinen et al. (2009), Pekkala Kerr et al. (2013) for Finland, and Black et al. (2005) for Norway.

[3] A related literature has used alternative approaches based on twins, siblings or adoptive parents designs (see Black and Devereux, 2011). More recently, Chetty et al. (2014) has revitalized interest in the role of neighborhoods and social mechanisms underlying intergenerational mobility (Chetty et al., 2016; Chetty and Hendren, 2018a; Chetty and Hendren, 2018b; Chetty et al., 2020). See also Nybom and Stuhler (2024) for a recent theoretical framework aimed at understanding trends in the intergenerational transmission of income and education.

by Pekkarinen (2008), Pekkarinen et al. (2009), and Pekkala Kerr et al. (2013), who show that the policy change improved the intergenerational income mobility and boosted educational attainment of women and individuals from lower socioeconomic backgrounds. Overall, the policy change appears to have contributed to greater equality of opportunity in Finnish society.

The existing evidence of the causal effect of the Finnish reform exploits the rollout of the policy change. In Finland, as for similar other reforms in Europe, the new education system was gradually introduced across regions and for different cohorts over time. In the language of the evaluation literature, one can then use the outcomes of the individuals in untreated regions to build the counterfactual outcomes for the treated units, had they not been treated. Identification of the reform effects is achieved assuming no anticipation and parallel trends. In practice, the existing evidence on the reform effects is based on a two-way fixed effects (TWFE) estimator that controls for regional and cohort indicators (see e.g., Pekkarinen et al., 2009).

The modern difference-in-differences literature has shown that the TWFE estimator is a weighted average of difference-in-difference comparisons between unit-time period pairs, where one unit changed treatment status while the other did not. A known issue with the TWFE estimator occurs when the treatment is staggered and the treatment effect is heterogeneous across time or cross-section units. In this case, the estimated average treatment effect on the treated is biased since it combines clean comparisons (between treated and not-yet-treated or never-treated units) with forbidden comparisons between early- and later-treated units (Goodman-Bacon, 2021). In other words, some of the comparison pairs will use early-treated units as control group, which can lead to negative weighting issues resulting in severe bias and even reversal of the estimated effect sign (de Chaisemartin and D'Haultfoeuille, 2020).

We argue that the effect of the Finnish comprehensive education reform is unlikely to be consistently estimated via a standard TWFE estimator, precisely because comparison units are treated later and the regions that implemented the reform at different times exhibit substantial heterogeneity, making it unlikely that the treatment effect is homogeneous across groups. This is confirmed by the presence of negative weights in our setting and by the fact that the treatment effect of the younger birth cohorts treated in the more disadvantaged areas is larger. For this reason, we re-evaluate the reform effects in Pekkarinen (2008) and Pekkarinen et al. (2009) on education and intergenerational income mobility by using modern difference in differences methods (Borusyak et al., 2024; Callaway and Sant'Anna, 2021).

To compare our results to those in the existing literature, our analysis starts with the TWFE estimator. This also allows us to update the results of the 2008 and 2009 Finnish papers with whole population data instead of a 10% sample. We then implement two alternative difference-in-differences estimators that avoid making the type of forbidden comparisons mentioned above. These are the imputation method (IM) by Borusyak et al. (2024) (our preferred estimators in this setting for reasons explained in detail in the text), and the Callaway and Sant'Anna

(2021) (CS) estimator, which we use to substantiate the results and test for their robustness. We also experiment with different ways to relax the parallel trends assumption, allowing it to hold conditionally predetermined parental characteristics and regional trends.[4]

When we use the TWFE estimator to analyze the reform effect on tertiary education, we replicate the Pekkarinen (2008) negative effect on men and positive effect for women.[5] Next, the IM and CS estimates show robust evidence of a positive effect on women of about 2–3 percentage points and somewhat weaker evidence of a smaller positive effect on men. We consider these our preferred estimates, as they are based on functional forms that control for parental background, thereby adjusting for observed differences in pre-trends. This leads to estimated effects that are smaller in magnitude compared to the unconditional estimates.

An important analysis of the 2008 paper was the analysis by parental education. We show across several specifications that for both men and women the effect is concentrated among the disadvantaged individuals—those whose fathers did not have a degree from secondary education—and generally insignificant otherwise. While this result is different from Pekkarinen (2008) (who shows a larger effect on children with academic fathers), it is consistent with the findings of Pekkarinen et al. (2009), who show that the reform reduced the intergenerational elasticity of income.

We also extend the analysis of the reform effect on tertiary education by studying heterogeneities jointly by father education and child's ability. We proxy ability by using brother's test score on the logical reasoning part of the Basic Skills test of the Finnish Army. The results show that the positive reform effect on men with low-educated father tends to be stronger for high-ability men. Moreover, the reform had a positive impact for the women with low ability brothers, whereas it had no differential effect by father background. Overall, the results are consistent with the analysis on intergenerational income mobility, which was positively affected only for men (see next paragraph).

We then turn to the analysis of the reform effect on the intergenerational income elasticity for men (Pekkarinen et al., 2009). Our data covers the full-population; it also allows us to exploit a much longer time window (until 2022 instead of 2000), therefore diminishing the influence of life cycle bias on the coefficients. We also extend the analysis by computing rank-rank coefficients and the inverse hyperbolic sine transformation to deal with the zeros. In contrast with the 2009 paper, we do not find the TWFE estimator to be significantly different from 0. However, the IM estimator shows a reduction in intergenerational elasticity coefficient in line with Pekkarinen et al. (2009), although not as large in magnitude (with estimates between -0.030 and -0.022, compared to -0.069). The effect is negative and statistically significant also

[4]In ongoing analysis we also try alternative ways to treat zero income rows, which is a known important aspect to operationalize when estimating intergenerational income mobility estimates.

[5]We also show that such a result is sensitive to the model functional form, as the estimates become insignificant when we allow municipality fixed effects to differ by gender.

for the percentile transformation, and results are robust when we implement a trend-adjusted version of the DiD estimator. In summary, the reform reduced income persistence, although to a lesser extent as compared to what was initially estimated.[6]

2 Comprehensive school reform in Finland

In the 1970s, Finland implemented a major school reform that delayed tracking from age 11 to age 16.[7] Before the reform, all students entered primary school at the age of 7. After four years of primary school, at age 11, students could apply to general secondary school, admission to which was based on school grades and teacher evaluation. Admitted students attended junior secondary school for five years, and they could continue to upper secondary school for three more years. At the end of general secondary school, students took the matriculation examination which granted admissibility to university education. Those who were not admitted to general secondary school at age 11 continued in primary school until age 15, after which they could apply to vocational school or enter the labor market. The last two years of the primary school were also called continuation classes, or civic school, and they concentrated on vocational skills.

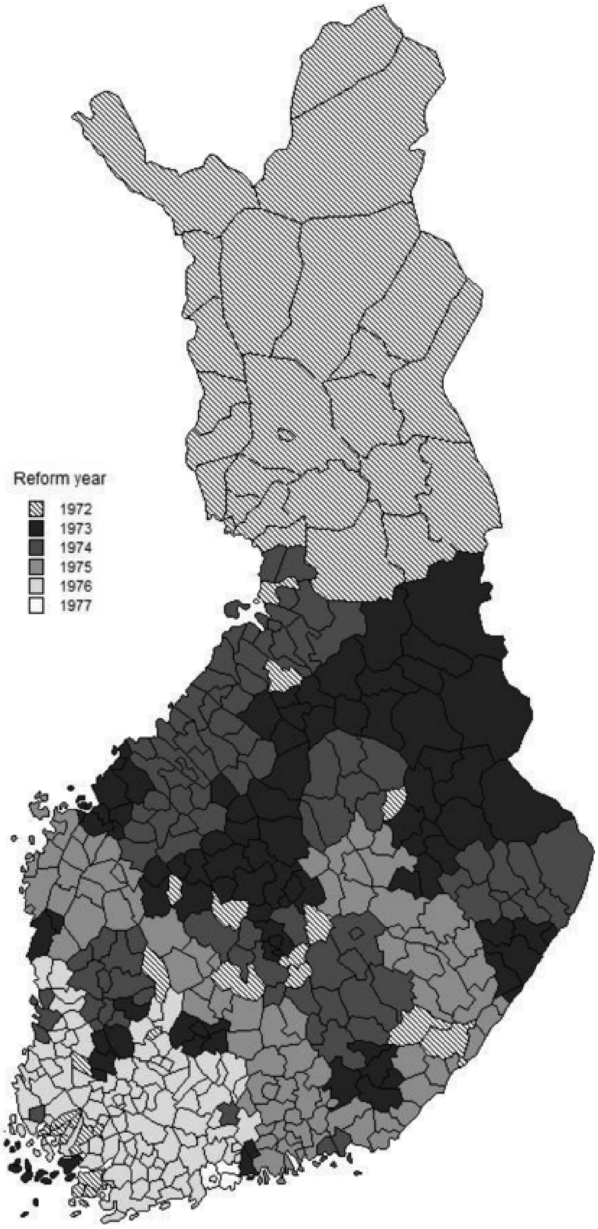
After the reform, the two-track system was replaced by a comprehensive school. Students attended the same school from age 7 until age 15, after which they could apply to either upper secondary school or vocational school, or enter labor market without further education. The reform did not change compulsory years of education. Admission to both upper secondary school and vocational school were based on the grades from comprehensive school.

The reform was implemented in different regions in different years according to a regional implementation plan. Figure 2 depicts the year each municipality implemented the reform. The first municipalities to adopt comprehensive school in 1972 were mostly located in the northernmost province of Lapland. The capital region of Helsinki was the last region to implement the reform in 1977. In the year of the reform, the first four grades of primary school switched to comprehensive school while students in grade five and above finished their education in the old system. The staggered rollout of the reform enables the identification of its effects based on the difference-in-differences methodology.

[6] As in Pekkarinen et al. (2009), our analysis also shows a smaller reduction in intergenerational elasticity for women (but only for the percentile transformation and with significant pre-trends).

[7] See Pekkarinen et al. (2009) for a more detailed discussion of the school reform.

Figure 1: Implementation of the school reform across regions in different years (source: Pekkarinen et al., 2009)



3 Data and sample

3.1 Data sources and sample selection

We draw the full population born between 1951 and 1966 in Finland, consisting of 1,271,604 individuals. We implement the following cleaning steps on this target population. First, we exclude 9,700 individuals for whom we could not identify the year of reform in the municipality where they lived. Second, we remove 16,442 individuals who passed away before 1987, the year our main dataset on educational achievement and income starts. Third, we exclude

42,876 individuals for whom we cannot observe educational attainment at the age of 30 or later. Fourth, we remove 117,116 individuals lacking information on the father's identifier and an additional 25,883 individuals without any information on their father's education and occupation, which are key controls in our regressions.

After these cleaning steps, we are left with 1,059,587 individuals. The sample covers years 1951-1960 to have a long-enough pre-treatment period for the IM estimator. If we limit the sample to those born between 1960 and 1966 and who did not move between regions during the period of the reform, as in previous empirical studies, we retain 446,963 individuals (228,086 men and 218,877 women). As expected, this sample size is roughly ten times larger than the 10% samples used in these studies.[8]

3.2 Data description

We present the descriptive statistics of our data. In Table 1, we illustrate the staggered implementation of the reform and the evolution of regional group sizes across different birth cohorts. Figure 2 depicts the descriptive evolution of the outcome, a dummy for having obtained a tertiary degree, across the six regional groups over the birth cohorts.

Tertiary educational achievement was lower in the early-treated regions, which may explain the earlier implementation of the reform in these areas. The pre-reform trend is relatively flat for men but positive for women. After the implementation of the reform, women in the earlier-treated regions catch up, and the gap is even reversed when compared to the latest-treated region, Helsinki. For men, there is also evidence of a catch-up. However, it appears that for women the trends were not parallel before the reform, which poses a potential threat to our identifying assumption.

This difference in trends might be driven by heterogeneous changes in characteristics of more recent birth cohorts across regions. To investigate this, in the Appendix (Figures A-1 - A-2) we descriptively show the evolution of predetermined characteristics (parental education, occupation, and household income measured in 1970) across regions and birth cohorts, which confirms the presence of differential trends. These findings underscore the importance of controlling for compositional changes in our repeated cross-sectional analysis. The descriptive characteristics and their evolution are also presented in Table 1.

Another important observation, when examining the explanatory variables and the pre-existing differences in levels for the outcomes, is the significant disparity between the earlier-treated regions of the North and the latest-treated region, Helsinki. As the parallel trend assumption is less likely to hold when using Helsinki as a control group, we follow the previous studies and exclude this regional group in an robustness analysis.

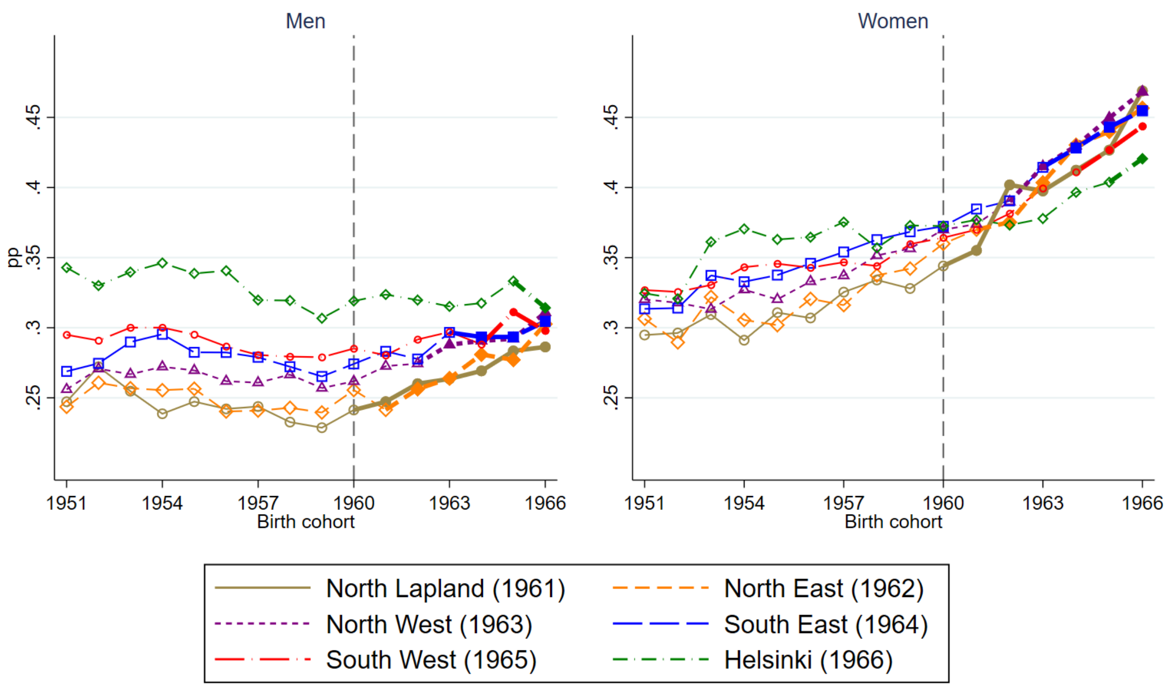
[8]For example, Pekkarinen (2008) relied on a final sample of 39,088 individuals, while Pekkarinen et al. (2009) used a sample of 20,824 men.

Table 1: Regional Rollout: Sample size

Region (Reform Year)	1951–1960	1961	1962	1963	1964	1965	1966	Total
North Lapland (1972)	63,918	6,925	6,763	6,675	6,397	5,980	5,844	102,502
North East (1973)	95,023	9,684	9,586	9,597	9,200	8,711	8,656	150,457
North West (1974)	142,625	14,621	14,484	14,338	13,900	13,640	13,508	227,116
South East (1975)	137,882	13,990	14,164	14,360	14,165	13,997	13,615	222,173
South West (1976)	139,979	14,530	14,675	15,098	15,148	14,617	15,025	229,064
Helsinki (1977)	73,349	8,311	8,707	9,074	9,413	9,611	9,810	128,275
Total	652,776	68,061	68,379	69,134	68,223	66,556	66,458	1,059,587

Notes: group size by regional group and over birth cohort. The shaded cells are the birth cohorts affected by the reform.

Figure 2: The evolution of tertiary education across birth cohorts by regions



Notes: The figure shows the evolution of tertiary degree achievement across birth cohorts by regional groups. The staggered implementation of the reform between regional groups is depicted by using empty markers for birth cohorts not affected by the reform in a given regional group, and full markers with thicker lines for cohorts after the reform has been implemented in that group. The vertical dashed line depicts the last completely untreated birth cohort (1960). Left panel: men, right panel: women.

4 Empirical approach

As in the original papers, we identify the causal effect of the education reform on the outcomes of interest by following a difference-in-differences (DiD) approach. We exploit the staggered implementation of the reform across Finland, where pupils in primary schools in different regions were exposed to the treatment at different times. The identifying assumption relies on the parallel trends in the potential outcomes in the absence of treatment, between the different six regional groups g (the earlier- and later-treated areas) and over the different birth cohorts t (constituting the "time" component of the DiD estimator). Consistent with the prior literature, we use outcomes such as achieving a tertiary degree (by the age of 30) and taxable income in 2000, as well as across different age brackets.

We begin the analysis by replicating previous findings using a two-way fixed effects estimator, controlling for the municipality of birth m and the birth-cohort t fixed effects.[9] This is estimated using the following linear regression:

$$Y_{it} = \beta_m + \gamma_t + \delta r_{mt} + \epsilon_i \tag{1}$$

where Y_{it} represents the outcome of individual i, who is born in year t in municipality m. The municipality and birth-cohort fixed effects are captured by β_m and γ_t, respectively. The treatment effect is identified by the coefficient δ of the dummy variable r_{mt} that is equal to one if the individual is born in a municipality m that has already implemented the reform for the birth cohort t. Standard errors are clustered at the municipality level m in all the empirical analyses.

As in Pekkarinen (2008) and Pekkarinen et al. (2009) we fully interact model (1) variables with the heterogeneity variable H_i (which is a gender dummy in the education analysis and father's income in the intergenerational income elasticity analysis).[10] The interaction allows for differential municipality (α_m) and birth-cohort (τ_t) fixed effects, as well as the heterogeneous treatment effect (θ). We also estimate the more restrictive model of Pekkarinen (2008, p. 819), in which the regional fixed effects are assumed to be homogeneous by gender and are therefore not interacted with the heterogeneity variable H_i.[11] This is implemented through the following linear regression:

$$Y_{it} = \beta_m + \gamma_t + \delta r_{gt} + (\alpha_m + \tau_t + \theta r_{mt}) \cdot H_i + \epsilon_i \tag{2}$$

As highlighted in recent econometric literature (e.g. Goodman-Bacon, 2021; Sun and Abraham, 2021), these two-way fixed effects (TWFE) estimators identify the treatment effect of the reform only under the assumption of homogeneous treatment effects across the six g re-

[9] Unlike Pekkarinen (2008), Pekkarinen et al. (2009) aggregate the municipality fixed effects into six regional group fixed effects β_g, which represent the calendar time of the reform's implementation. Since our setting is a repeated cross-section, this approach is not equivalent. In our main specification, we follow the more flexible approach Pekkarinen (2008) and control for β_m to account for changes in municipality composition within regional groups. However, the results are robust to using regional fixed effects and are available upon request.

[10] For the educational analysis, we also investigate the presence of heterogeneous effects based on the father's level of education. These analyses are conducted separately for each gender, and H_i is a dummy variable indicating whether the father had a secondary degree diploma.

[11] Following Pekkarinen et al. (2009), in the intergenerational income analysis, the interaction of the group fixed effects with the continuous variable of father's income is aggregated across the six regional groups (α_g) to avoid an overly complex model allowing for a different level of intergenerational income elasticity for each municipality m.

gional groups (the groups of municipalities m implementing the reform at the same time) and birth cohorts t. Given that the regions implementing the reform earlier are the most disadvantaged areas of Finland (i.e. the northern areas), this assumption appears to be restrictive. After demonstrating the presence of negative weights in our setting, which are problematic in the presence of heterogeneous effects over t and g, as described by de Chaisemartin and D'Haultfoeuille (2020), we adopt the imputation method proposed by Borusyak et al. (2024), which relaxes this homogeneity assumption.[12]

The treatment effect is therefore estimated using a two-step approach. First, we estimate the coefficients using a linear regression similar to equations (1) and (2), but excluding observations from the moment they become treated. By relying solely on untreated observations, we do not impose any functional form on the treatment effect to correctly identify the other coefficients. As shown in Section 3.2, we observe different changes in observable characteristics over t across g. We account for this by including a set of control variables, X_i, capturing parental background effects in levels (coefficients μ). These controls are further interacted with time dummies to account for differential time trends conditional on X (coefficients π_t), thereby relaxing the parallel trends assumption to hold only conditional on observable characteristics. This first-step is formalized in the following linear regression:[13]

$$Y_{it} = \beta_m + \gamma_t + (\alpha_m + \tau_t) \cdot H_i + \mu X_i + \pi_t X_i + \epsilon_{it} \qquad if \quad r_{mt} = 0 \tag{3}$$

After estimating the coefficients in equation (3), we predict the counterfactual outcome in the absence of treatment, $\hat{Y}_{it}(0)$, for the excluded treated observations. These predictions are based on the coefficients identified from the untreated observations and therefore are robust to treatment effect heterogeneity. The predicted counterfactual outcomes are then compared to the observed outcomes under treatment, enabling the identification of the individual treatment effect.

$$\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0) \tag{4}$$

To gain precision, we aggregate the estimated individual treatment effects $\hat{\delta}_{it}$ to calculate the average treatment effect for each birth cohort ($\hat{\delta}_t = \sum_i \hat{\delta}_{it}$) or across all treated observations $\hat{\delta} = \sum_t \sum_i \hat{\delta}_{it}$). Heterogeneous treatment effects θ are obtained in a third step by regressing the treatment effects $\hat{\delta}_{it}$ on the heterogeneity variable H_i.

Standard errors are obtained by estimating individual error terms $\hat{\varepsilon}_{it}$ and using the plug-in estimator $\hat{\sigma}^2_w = \sum_i \left(\sum_{t_i, t \in \Omega} v_{it} \tilde{\varepsilon}_{it} \right)^2$. To separate the error component from the individual treat-

[12]Related estimators to the imputation methods have also been proposed in the literature, including Wooldridge (2021), Liu et al. (2022), and Gardner (2022).

[13]In a sensitivity analysis, we rely on a trend-adjusted DiD version by including a group-specific linear trend for the 6 regional groups g as follows: $Y_{it} = \beta_m + \gamma_t + (\alpha_m + \tau_t) \cdot H_i + \mu X_i + \pi_t X_i + \iota_g t_t + \epsilon_{it} \qquad if \quad r_{mt} = 0$

ment effect, it is necessary to impose restrictions on the heterogeneity of individual treatment effects. We rely on the default assumption of homogeneous treatment effects within the same group g, which is obtained by averaging the individual treatment effects within g. As explained by Borusyak et al. (2024), these standard errors account for potential misspecification of individual treatment effects, and a correctly specified auxiliary parsimonious model implies tighter confidence intervals.

Placebo tests on the pre-treatment period are implemented by running equation (3), including a set of placebo dummies as in an event study framework (but using only untreated observations) for P periods prior to the start of treatment, which we set equal to 5.[14] The joint significance of these placebo dummies is tested using a Wald test. Note that the reference "pre-treatment" periods differ between the treatment analysis and the placebo analysis: the treatment analysis uses the full pre-treatment period, whereas the placebo analysis focuses on the initial pre-treatment period preceding the placebo dummies.

Finally, by relying on the full pre-treatment periods to estimate the coefficients of eq. (3), the estimator achieves efficiency in the absence of serial correlation. Borusyak et al. (2024) demonstrate through simulations that this estimator remains more efficient than alternative methods as long as serial correlation is not excessively high. Moreover, under spherical errors, it avoids the pretesting issues highlighted in Roth (2022). However, using the full pre-treatment period imposes a stronger parallel trend assumption, which must hold across all pre-treatment periods.

We test the sensitivity of our results and the specificity of the imputation estimator by employing the method proposed by Callaway and Sant'Anna (2021) (CS). The CS method decomposes the data into numerous 2x2 DiD comparisons for each treated regional group and time period t, excluding the already treated units. Similar to the imputation method, the group-time-specific treatment effects are aggregated using weighted averages to calculate the treatment effect over t or the average effect across all treated observations.

Compared to the imputation method, the CS estimator differs in several key ways. First, the CS estimator uses only the period immediately preceding the treatment as the reference to estimate treatment effects. Hence, it requires the parallel trend assumption to hold only from the last pre-treatment period, making it more robust to violations of the parallel trend assumption that may accumulate over time, particularly in periods distant from the treatment. Second, differential trends conditional on X are accounted for semi-parametrically in a doubly robust manner, simultaneously incorporating trends in X through the outcome regression model (Heckman et al., 1997) and the probability of treatment given X via an inverse probability weighting estimator (Abadie, 2005). Third, in repeated cross-sections, the estimator

[14] In the trend-adjusted version of the imputation model we set P=3 to allow a sufficiently long time for identifying differential trends during the pre-placebo period.

assumes stationarity, meaning there are no compositional changes in observable characteristics over t.[15] Fourth, inference is conducted using a simple multiplier bootstrap method, which accounts for multiple testing. Finally, the standard CS method does not directly allow for interactions to test for heterogeneous effects conditional on H. Therefore, when exploring such heterogeneity, such as in the intergenerational elasticity model, we rely exclusively on the imputation method.

5 Results

5.1 Tertiary education

We first replicate the finding of Pekkarinen (2008), who relied on a 10% sample, by implementing the same interactive TWFE estimator by gender. As shown in column (1) of Table 2, the results are in line and show a statistically significant negative effect for men (-1.6 pp) and a positive one for women (+2.2 pp).[16] However, if we implement a less restrictive model and allow the municipality fixed effect to differ by gender, the effects become statistically insignificant for both groups (column (2)), which is mechanically replicated in the subgroup analysis where we do not interact the explanatory variables by gender but implement the regression on each gender group separately (columns (3) and (4)). This suggests that if we use a more flexbible model specification, the TWFE estimator does not return a reform effect significantly different from zero.

However, this finding relies on the identifying assumption of homogeneous treatment effects by regional group and birth cohort. Indeed, if we calculate the weights used by the TWFE estimator as suggested by de Chaisemartin and D'Haultfoeuille (2020), the treatment effects of 6 different regional-cohort groups (out of 21) receive negative weights, and the sum of these negative weights is -0.2464 (see Table A-2 in the Appendix). These are relative to the three regions treated earlier (Lapland, the north-east, and the north-west) for some of the cohorts born later (1964, 1965, 1966). This would not represent a problem if the treatment effect of these regions and cohorts were the same as the other ones. However, as we will show later, the treatment effect of the younger birth cohorts treated in the more disadvantaged areas is larger, which downward biases the treatment effect estimated by the restrictive TWFE estimator.

We next implement the imputation method of Borusyak et al. (2024), which allows for heterogeneous effects by regional group and birth cohorts. In doing so, we expand the pre-treatment period to a longer period (for birth cohorts born after 1951 instead of only from 1960). This longer time span should increase precision in the treatment effect, if serial correlation is not excessive (De Chaisemartin and d'Haultfœuille, 2023; Borusyak et al., 2024), and allow

[15] This assumption is relaxed in Sant'Anna and Xu (2023).

[16] The original paper showed an effect of -0.3 pp on men and +1.5 pp on women.

Table 2: Replication of TWFE estimates on achieving tertiary education

	(1)	(2)	(3)	(4)
	Pooled	Pooled interacted	Subgroup	Subgroup
Men	–0.016***	–0.000	–0.000	.
se	0.004	0.004	0.004	.
p-value	0.000	0.944	0.944	.
Women	0.022***	0.005	.	0.005
se	0.006	0.004	.	0.004
p-value	0.001	0.219	.	0.219
Women - Men	0.038***	0.006	.	.
se	0.009	0.005	.	.
p-value	0.000	0.286	.	.
N	446,963	446,963	228,086	218,877

Notes: The table shows the estimates obtained by a TWFE estimator controlling for municipality of birth and birth-cohort fixed effects. Data: individuals born between 1960 and 1966, restricting to those who did not change regional group during the reform period. The outcome variable is achieving a tertiary education degree by the age of 30. Models (1) and (2) present results based on Equation (1) (interactive model), while Models (3) and (4) provide results from subgroup analyses: (3) for men and (4) for women. Model (1) does not include an interaction between municipality fixed effects and the gender variable, unlike Model (2). We report the absolute effect, standard errors, and p-values. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

us to better observe pre-existing differences in trends during the pre-treatment period and possibly to correct for these by relying on a trend-adjusted DiD estimator. Furthermore, we remove the individuals born after 1965 as from the 1966 cohorts onward, all regions are treated, making it impossible to identify the causal effects on the younger individuals. Finally, we do not drop individuals who changed regional groups during the period of reform (which might induce an endogenous sample selection) and consider the region where the individual lived in 1970, two years before the reform started to be implemented in the first municipalities. We do this to avoid introducing endogenous selection into the subgroup of non-movers. The trade-off is that compliance is not perfect. This means that some units assigned to the (earlier) treated regions may not have actually received the treatment, while some units in the control regions may have. Since the treatment differences between the groups are not sharp, our estimates represent an intention-to-treat effect, which, under standard assumptions, can be interpreted as a lower bound of the actual treatment effect. Overall, departing from the original sample selection does not significantly affect the estimates.

As shown in columns (1) and (2) of Table 3, the effect is 1.8 pp on men and 4.2 pp on women, which are statistically significant at the 10% level and 1% level, respectively. As the counterfactual share of graduates in the absence of the treatment during the treatment period is 26.4% and 37.8%, respectively, this represents a sizable increase of 6.8% and 11.2% in the share of men and women obtaining a tertiary degree due to the reform. However, placebo tests fail for women, suggesting that the parallel trends assumption is unlikely to be satisfied. This can also be observed in the evolution of the estimates for the older birth cohorts that were not affected by the reform as they already completed primary school by the time of the regional implementation. As shown in Figure A-3 in the Appendix, women in the earlier treated regions exhibited a steeper increase in educational levels across these older birth cohorts, suggesting that the estimated treatment effect might reflect preexisting trends and, therefore, represent an upper bound of the actual intention to treat.

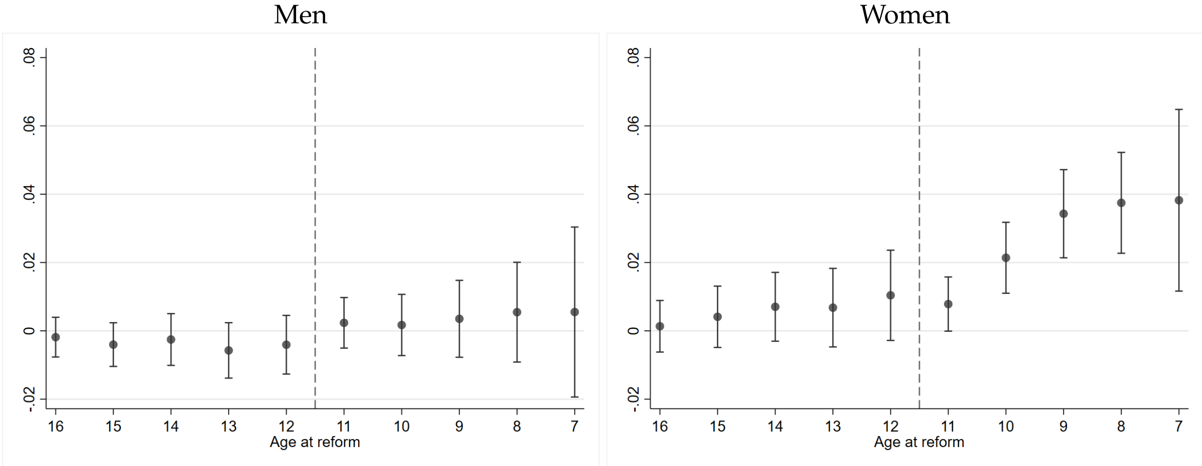
Table 3: Heterogeneous DiD estimates on achieving tertiary education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men	Women	Men	Women	Men	Women	Men	Women
δ	0.018*	0.042***	0.004	0.028***	0.009	0.033***	0.013**	0.021***
se	0.010	0.007	0.005	0.006	0.008	0.005	0.006	0.008
p-value	0.068	0.000	0.487	0.000	0.271	0.000	0.033	0.006
p-value placebo	0.877	0.011	0.790	0.721	0.000	0.000	0.659	0.802
N	511,928	481,201	491,407	462,342	491,407	462,342	491,407	462,342
Method	Imput	Imput	Imput	Imput	CS	CS	Imput	Imput
Parental background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	No	No	No	No	Yes	Yes

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1965 and without mobility restriction. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men: columns (1), (3), (5), (7); women: columns (2), (4), (6), (8)). Columns (1) and (2) implement the imputation method (Borusyak et al., 2024). In columns (3) and (4) we implement a conditional DiD estimator by including to the imputation model control variables on the parental background (father's and mother's education, occupation, and household income in 1970), which explain both the level and differential trends. Columns (5) and (6) implement the doubly robust conditional difference-in-differences as in Callaway and Sant'Anna (2021), which assumes no compositional changes over birth cohorts. In columns (7) and (8), we rely on a conditional trend-adjusted DiD estimator by implementing the imputation method, controlling for both parental background and linear differences in trends by region. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

We therefore add, both in levels and trends, a set of control variables including parental education, occupation, and income in 1970. This allows us to relax the parallel trends assumption to hold only conditionally and to control for compositional changes over birth cohorts in those predetermined characteristics. Columns (3) and (4) show a reduced positive effect of 0.4 pp on men and 2.8 pp on women (+1.3% and +7.1% in relative terms), with only the latter remaining significant. Placebo tests show a p-value above 0.700, indicating that the treatment effect estimates are more reliable, as also shown in Figure 3, which now displays smaller and insignificant pre-trends. As anticipated already in Section 3.2, conditioning on parental background is an important factor for retrieving more credible causal estimates.

Figure 3: Imputation estimates on achieving tertiary education: aggregation by age of children at time of reform: controlling for parental background



Notes: The figure shows the event-study estimates obtained by the imputation estimator (Borusyak et al., 2024) controlling for municipality of birth and birth-cohort fixed effects and parental background (father and mother educational level, household taxable income and father and mother profession in 1970), controlling for differences in level and trends. Data: individuals born between 1951 and 1965 and without mobility restriction. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men, left; women, right). The effects are aggregated by the individual's age at the time of the reform's implementation, corresponding to their birth cohort. Older birth cohorts, aged above 11 at the time of the reform's regional implementation, were not affected by the treatment as they had already completed primary school. The benchmark pre-treatment period of the DiD varies for the treatment effect (ages 7–11) and the placebo estimates (ages 12–16): the full pre-reform period is used for the treatment effect estimates, while the five years preceding the placebo period are used for the placebo estimates.

We then use the Callaway and Sant'Anna (2021) estimator, which is robust to heterogeneous effects and allows for controlling parental background in a doubly robust manner. In columns (5) and (6), we observe a larger effect of 0.9 percentage points for men and 3.3 percentage points for women. However, these effects do not pass the placebo tests. This may be due to the estimator's assumption of stationarity in repeated cross-sections—meaning no compositional changes across birth cohorts—a stronger assumption compared to the imputation method, which accounts for such changes. Finally, as Figure 3 still showed some increasing (though insignificant) trends for earlier treated regions before the treatment, we move to the trend-adjusted version of the imputation method and include six regional linear trends. As expected, estimates are partially reabsorbed for women: +2.1 pp (or 5.2%) which is still statistically significant at the 1% level. For men they increase to +1.3 pp (or 4.7%), which is statistically significant at the 5% level.

Overall, we find robust evidence of a positive effect on women of about 2–3 percentage points and some weaker evidence of a smaller effect on men.

We then examine heterogeneous effects based on fathers' education. As shown in Tables A-3 and A-4 in the Appendix, the effect is concentrated among more disadvantaged individuals—specifically, those whose fathers did not have a degree from secondary education. In contrast, the effect is rarely statistically significant for individuals whose fathers had at least a secondary degree.

Children from more advantaged families already had relatively high rates of tertiary degree attainment (around 42% for men and 52% for women) compared to children from more disadvantaged families, whose rates were approximately 20% for men and 34% for women. This suggests that postponing the selection of children into academic and non-academic tracks reduced intergenerational inequalities in educational achievement. The results remain robust even after excluding Helsinki from the data. As shown in the descriptive statistics, the Helsinki region appears to differ significantly from other regions in terms of pretreatment outcomes (Figure 2) and predetermined control variables (Figures A-1-A-2 in the Appendix). Following previous literature, we exclude this region in a robustness analysis, effectively limiting the analysis to cohorts up to 1964. As shown in Tables A-5-A-6 in the Appendix, the results remain robust to this exclusion, showing a positive effect for men and women whose fathers did not have a secondary education degree, with no effect on the rest of the population.

While this finding is not in line with Pekkarinen (2008), who shows a larger effect on children with academic fathers (results, however, based on the restrictive interactive model not interacting the municipality fixed effect and the assumption of effect homogeneity over t and g), it aligns with the findings of Pekkarinen et al. (2009), who showed that the reform reduced the intergenerational elasticity of income. We analyze this outcome in the next subsection.

Finally, we investigate heterogeneity by child's ability and father's level of education. We use brother's test score on the logical reasoning part of the Basic Skills test of the Finnish Army as a proxy for ability.[17] We use brother's test score instead of one's own since this allows us to extend the analysis for women and for birth cohorts born before 1962.[18] Moreover, brother's test score is less likely to be affected by the reform.[19]

Columns (1)-(4) in Table 4 show that the school reform had a positive effect on the probability of obtaining a tertiary degree for men with low educated fathers. The effect is stronger for high ability men, although the difference between the effects of columns (1) and (3) is not statistically significant. For men with fathers with at least a secondary degree, the effect of the school reform is insignificant. For women, we find that the school reform had a positive impact for those with low ability brothers whereas father background does not seem to matter. The results are consistent with our findings in the next section, where we show that the school reform increased intergenerational income mobility for men but not for women.

[17] In particular, we use either brother's test score for individuals with only one observed brother or the average of brothers' test scores for those with several brothers.

[18] We observe test scores for few men born before 1962 because the test score data starts in 1982 and the typical age at military service is 20.

[19] This concern is also alleviated by the results of Pekkala Kerr et al. (2013), who show that the Finnish school reform did not have a statistically significant effect on the logical reasoning test score.

Table 4: Heterogeneity by ability and father background

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men LA/LF	Men LA/HF	Men HA/LF	Men HA/HF	Women LA/LF	Women LA/HF	Women HA/LF	Women HA/HF
δ	0.021**	0.003	0.031***	0.021	0.038***	0.044***	-0.001	0.011
se	0.009	0.014	0.010	0.014	0.008	0.016	0.020	0.013
p-value	0.026	0.836	0.002	0.114	0.000	0.004	0.968	0.407
p-value placebo	0.976	0.099	0.693	0.696	0.771	0.207	0.114	0.161
N	38,988	13,069	27,162	19,827	37,845	13,227	27,182	20,353

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1965 and without mobility restriction. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender, ability proxied by brother's military test score (LA = below median, HA = above median), and father's education (LF = no secondary education, HF = at least secondary education). All columns use the imputation method by Borusyak et al. (2024) and control for mother's education, parental occupation, and log household income in 1970. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality level (using municipality of residence in 1970). * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

5.2 Intergenerational income elasticity

In this section we estimate the effect of the reform on the intergenerational elasticity of income.

We first replicate the analysis of Pekkarinen et al. (2009) by regressing the log of children's taxable income in 2000 (Y_{it}) on the log of the average taxable income (H_i) of the father during the period 1970–1990 (top-trimmed at the 99th percentile).[20] Column (1) of Table 5 shows an intergenerational income elasticity of 28.5%, which is very similar to the previously reported value of 27.6%.[21]

We then progressively add the TWFE components. Column (2) indicates that the reform dummy (r_{gt}) is associated with a 6% lower level of taxable income in 2000, which is statistically significant at the 1% level. Additionally, this negative association increases by 0.013 percentage points for individuals whose father's income is 1% higher; however, this effect is not statistically significant at the 10% level. While the coefficient of the reform is identical to that reported by Pekkarinen et al. (2009), the interaction term with father's income is smaller (-0.013 vs -0.055).

When we include the regional group and cohort fixed effects, the estimates change considerably, as we no longer find any statistically significant coefficients for the reform's impact on children's income or on the intergenerational income elasticity. This is in contrast with the results of Pekkarinen et al. (2009) which relied on a 10% sample.

As the evaluation setting involves negative weights, these results are unbiased only if the effect of the reform is homogeneous across regional groups and birth cohorts—a strong as-

[20] As in the original paper we consider only the years with positive earnings in the average.
[21] Note that Pekkarinen et al. (2009) uses a 10% sample of the population whereas we have the full population.

Table 5: Replication of TWFE estimates on intergenerational income elasticity (men)

	(1)	(2)	(3)
α	0.285***	0.288***	0.362***
se	0.007	0.009	0.014
p-value	0.000	0.000	0.000
δ	.	–0.063***	–0.002
se	.	0.004	0.006
p-value	.	0.000	0.694
θ	.	–0.013	0.007
se	.	0.009	0.011
p-value	.	0.131	0.518
N	219,709	219,709	219,709
Treatment dummy	No	Yes	Yes
Group and cohort FE	No	No	Yes

Notes: The table shows the estimates obtained by a TWFE estimator as in Pekkarinen et al. (2009). Data: Men born between 1960 and 1966, restricting to those who did not change regional group during the reform period. The outcome variable is the log of the imposable income in 2000. The heterogeneity variable is the log of the average father's income between 1970 and 1990, top-trimmed at the 99%. We report the absolute effect, standard errors, and p-values. α: intergenerational income elasticity between children and father's income ; δ: effect of the reform on children income; θ: effect of reform given father's income. Models: (1) includes only the father's income. Model (2) also adds the reform dummy r, which is interacted with father's income H. Model (3) is the TWFE estimator, which includes the regional group g and cohort t fixed effects. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

sumption. To address this, we implement the imputation method. Similar to the analysis of educational outcomes, we retain a longer pre-treatment period, condition on education and occupation of parents, and do not impose any mobility selection. Unlike Pekkarinen et al. (2009), we have access to income data up to 2022. This longer time window limits life cycle bias in the estimates and allows us to regress children's and fathers' taxable income over a comparable age window. Specifically, we rely on the average income level in the age brackets 35–55 or 45–50. The outcomes are transformed using the logarithmic transformation, the inverse hyperbolic sine, or the percentile rank within the birth cohort.

As shown in columns (1)–(6) of Table 6, we find evidence of a statistically significant positive effect on the average income level, which is different from the negative effect found in Pekkarinen et al. (2009). The interaction with the father's income variable is negative and statistically significant, which indicates a reduction in intergenerational elasticity, as found in Pekkarinen et al. (2009), though the effect is less than half of what they find (between -0.030 and -0.022 compared to -0.069). The effect is negative and statistically significant also for the percentile transformation. Placebo tests tend not to be rejected at the 5% level, apart from the average effect for the log and the inverse hyperbolic sine transformation for the age bracket 35–55.

To correct for pre-existing trends, we also implement the trend-adjusted version of the DiD estimator. As shown in columns (7)–(12), the overall effect of the reform is reduced after controlling for differences in trends, though the negative effect on the intergenerational elasticity remains virtually the same. This suggests that while the reform did not have a positive effect on the average income, it managed to reduce intergenerational income inequalities.

Table 6: Imputation method: estimates on intergenerational income elasticity (men)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct
δ	0.025***	0.034***	0.025***	0.034***	0.740***	0.652***	0.004	0.018*	0.004	0.018*	0.062	0.224
se	0.000	0.000	0.000	0.000	0.000	0.001	0.676	0.068	0.675	0.064	0.874	0.487
p-value	0.006	0.006	0.006	0.006	0.208	0.192	0.010	0.010	0.010	0.010	0.388	0.322
Placebo:												
p-value	0.026	0.158	0.026	0.162	0.133	0.370	0.950	0.675	0.951	0.676	0.695	0.868
θ	-0.022***	-0.030***	-0.022***	-0.030***	-0.014**	-0.012*	-0.025***	-0.033***	-0.025***	-0.033***	-0.016***	-0.012*
se	0.008	0.007	0.008	0.007	0.006	0.007	0.008	0.007	0.008	0.007	0.006	0.007
p-value	0.008	0.000	0.008	0.000	0.021	0.097	0.001	0.000	0.001	0.000	0.008	0.087
Placebo:												
θ	-0.008	0.011*	-0.008	0.011*	-0.009	-0.007	-0.009	-0.006	-0.009	-0.006	-0.006	-0.012*
se	0.005	0.006	0.005	0.006	0.007	0.007	0.007	0.010	0.007	0.010	0.006	0.007
p-value	0.130	0.092	0.132	0.092	0.227	0.306	0.187	0.562	0.190	0.560	0.279	0.072
N	439,017	369,162	439,017	369,162	439,017	369,162	439,017	369,162	439,017	369,162	439,017	369,162
Linear trend	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents the estimates obtained using the imputation estimator. The data include men born between 1951 and 1965 and without mobility restrictions. The outcome variable is the child's average imposable income within a given age bracket and transformation. The heterogeneity variable is the father's average income in the same age bracket. For the treatment effect, we report the average effect (δ) and the heterogeneity by father's income (θ), along with their respective standard errors and p-values. For the placebo tests, we report the p-value for the joint significance of the placebo dummies (starting five periods earlier) for the overall effect and the p-value for the heterogeneity of the effects. The latter is calculated similarly to θ but by rerunning the same imputation method, excluding the units when they enter treatment, and starting the placebo five periods earlier. Age brackets: 35–55 for columns (1), (3), (5), (7), (9), and (11); 45–50 for columns (2), (4), (6), (8), (10), and (12). Income transformations: Logarithm for columns (1), (2), (7), and (8); inverse hyperbolic sine for columns (3), (4), (9), and (10); percentile position within the birth cohort for columns (5), (6), (11), and (12). Standard errors are clustered at the municipality-of-birth level. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Finally, we also run the analysis on women, which also shows a reduction in intergenerational elasticity, though smaller and significant only for the percentile transformation. The lower effect was also found in Pekkarinen et al. (2009). However, placebo tests reject the null of parallel trends during the pretreatment birth cohorts; therefore, we should interpret these estimates with caution. Full results are shown in Table A-7 in the Appendix.

Table 7: Imputation method: estimates on intergenerational income elasticity (men) - No Helsinki

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct
δ	0.006	0.009	0.006	0.009	0.350	0.262	-0.001	0.004	-0.001	0.004	0.047	0.004
se	0.244	0.201	0.245	0.197	0.152	0.340	0.951	0.678	0.948	0.668	0.900	0.992
p-value	0.005	0.007	0.005	0.007	0.244	0.275	0.008	0.010	0.008	0.010	0.375	0.395
Placebo:												
p-value	0.479	0.563	0.480	0.567	0.410	0.521	0.890	0.849	0.892	0.847	0.868	0.740
θ	-0.019*	-0.014	-0.019*	-0.014	-0.010	-0.001	-0.020*	-0.014	-0.020*	-0.014	-0.011	-0.001
se	0.011	0.010	0.011	0.010	0.009	0.010	0.011	0.010	0.011	0.010	0.010	0.010
p-value	0.074	0.165	0.073	0.165	0.274	0.932	0.062	0.146	0.061	0.146	0.247	0.898
Placebo:												
θ	-0.013	-0.007	-0.013	-0.007	-0.011	-0.016	-0.013	0.000	-0.013	0.000	-0.006	-0.006
se	0.010	0.010	0.010	0.010	0.009	0.011	0.009	0.009	0.009	0.009	0.009	0.010
p-value	0.198	0.448	0.197	0.450	0.215	0.147	0.139	0.989	0.140	0.993	0.484	0.549
N	358,553	298,616	358,553	298,616	358,553	298,616	358,553	298,616	358,553	298,616	358,553	298,616
Linear trend	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents the estimates obtained using the imputation estimator. The data include men born between 1951 and 1964, without mobility restrictions and removing the region of Helsinki. The outcome variable is the child's average imposable income within a given age bracket and transformation. The heterogeneity variable is the father's average income in the same age bracket. For the treatment effect, we report the average effect (δ) and the heterogeneity by father's income (θ), along with their respective standard errors and p-values. For the placebo tests, we report the p-value for the joint significance of the placebo dummies (starting five periods earlier) for the overall effect and the p-value for the heterogeneity of the effects. The latter is calculated similarly to θ but by rerunning the same imputation method, excluding the units when they enter treatment, and starting the placebo five periods earlier. Age brackets: 35–55 for columns (1), (3), (5), (7), (9), and (11); 45–50 for columns (2), (4), (6), (8), (10), and (12). Income transformations: Logarithm for columns (1), (2), (7), and (8); inverse hyperbolic sine for columns (3), (4), (9), and (10); percentile position within the birth cohort for columns (5), (6), (11), and (12). Standard errors are clustered at the municipality-of-birth level. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

6 Conclusions

Understanding how socioeconomic status persists across generations is crucial for designing policies that promote equality of opportunity. This paper re-evaluates the effects of Finland's comprehensive education reform (1972–1977) on intergenerational mobility using modern difference-in-differences estimators. Our analysis updates the results of prior studies that relied on two-way fixed effects (TWFE) estimators (Pekkarinen, 2008; Pekkarinen et al., 2009).

We find that the reform had a positive and significant effect on tertiary education attainment for women, with weaker evidence of an effect on men. These results contrast with previous estimates using TWFE, which suggested a negative impact on men. The effect is particularly pronounced for individuals from lower socioeconomic backgrounds, reinforcing the idea that comprehensive schooling reforms can enhance mobility among disadvantaged groups.

Regarding income mobility, our analysis extends previous work by incorporating full-population data and a longer time horizon, thereby reducing concerns related to life-cycle bias. Our preferred estimates suggest that the reform led to a reduction in intergenerational income elasticity, albeit to a lesser extent than initially reported in Pekkarinen et al. (2009). This finding confirms the view that education policies can mitigate income persistence across generations.

Overall, our study provides robust evidence that the Finnish comprehensive school reform improved intergenerational mobility, although to a lesser extent than what it was initially thought. These findings have important implications for policymakers considering similar educational reforms to promote equality of opportunity. Future research should further explore the long-term effects of such reforms and investigate complementary policies that could enhance their effectiveness in reducing socioeconomic disparities.

References

Aakvik, Arild, Kjell G. Salvanes, and Kjell Vaage (2010). "Measuring heterogeneity in the returns to education using an education reform". *European Economic Review* 54.4, 483–500.

Abadie, Alberto (2005). "Semiparametric difference-in-differences estimators". *The Review of Economic Studies* 72.1, 1–19.

Becker, Gary S. and Nigel Tomes (1979). "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility". *Journal of Political Economy* 87.6, 1153–1189.

Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility". *Handbook of Labor Economics*. Vol. 4. Elsevier, 1487–1541.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes (2005). "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital". *American Economic Review* 95.1, 437–449.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024). "Revisiting Event-Study Designs: Robust and Efficient Estimation". *The Review of Economic Studies*, rdae007.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2, 200–230.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2020). "Income Segregation and Intergenerational Mobility Across Colleges in the United States*". *The Quarterly Journal of Economics* 135.3, 1567–1633.

Chetty, Raj and Nathaniel Hendren (2018a). "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects". *The Quarterly Journal of Economics* 133.3, 1107–1162.

— (2018b). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates". *The Quarterly Journal of Economics* 133.3, 1163–1228.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz (2016). "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment". *American Economic Review* 106.4, 855–902.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *". *The Quarterly Journal of Economics* 129.4, 1553–1623.

Cholli, Neil A and Steven N Durlauf (2022). "Intergenerational Mobility". NBER Working Paper, No. 29760.

De Chaisemartin, Clément and Xavier D'Haultfoeuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". *American Economic Review* 110.9, 2964–2996.

De Chaisemartin, Clément and Xavier d'Haultføeuille (2023). *Difference-in-Differences for Simple and Complex Natural Experiments*. November 2023.

DiPrete, Thomas A (2020). "The Impact of Inequality on Intergenerational Mobility". *Annual Review of Sociology* 46, 379–398.

Gardner, John (2022). *Two-stage differences in differences*. _eprint: 2207.05943.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, 254–277.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd (1997). “Matching as an econometric evaluation estimator: evidence from evaluating a job training programme”. *The Review of Economic Studies* 64.4, 605–654.

Holmlund, Helena (2007). “Intergenerational Mobility And Assortative Mating Effects of an Educational Reform”. Mimeo.

Liu, Licheng, Ye Wang, and Yiqing Xu (2022). *A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data*. _eprint: 2107.00856.

Meghir, Costas and Mårten Palme (2005). “Educational Reform, Ability, and Family Background”. *American Economic Review* 95.1, 414–424.

Nybom, Martin and Jan Stuhler (2024). “Interpreting Trends in Intergenerational Mobility”. *Journal of Political Economy* 132.8, 2531–2570.

Pekkala Kerr, Sari, Tuomas Pekkarinen, and Roope Uusitalo (2013). “School tracking and development of cognitive skills”. *Journal of Labor Economics* 31.3. Publisher: University of Chicago Press Chicago, IL, 577–602.

Pekkarinen, Tuomas (2008). “Gender Differences in Educational Attainment: Evidence on the Role of Tracking from a Finnish Quasi-experiment”. *The Scandinavian Journal of Economics* 110.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9442.2008.00562.x, 807–825.

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, 965–973.

Roth, Jonathan (2022). “Pretest with caution: event-study estimates after testing for parallel trends”. *American Economic Review: Insights* 4.3, 305–322.

Sant’Anna, Pedro H. C. and Qi Xu (2023). *Difference-in-Differences with Compositional Changes*. _eprint: 2304.13925.

Solon, Gary (2004). “A Model of Intergenerational Mobility Variation over Time and Place”. *Generational Income Mobility in North America and Europe*. Corak, Miles. Cambridge: Cambridge University Press, 38–47.

— (1992). “Intergenerational Income Mobility in the United States”. *American Economic Review* 82.3, 393–408.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. *Journal of Econometrics* 225.2, 175–199.

Wooldridge, Jeffrey M. (2021). *Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators*. Published: Technical report, Mimeo, Department of Economics, Michigan State University.

A Extended replication

Table A-1: Descriptive statistics

	Mean	N	1960	1966	dif	p-value
Men	0.515	1059587	0.511	0.508	-0.003	0.274
Women	0.485	1059587	0.489	0.492	0.003	0.274
Not moving	0.909	1059587	0.959	0.920	-0.039***	0.000
Tertiary degree	0.322	1059587	0.318	0.377	0.058***	0.000
Av. taxable income: age 35-55	27004	1049386	27605	30660	3055***	0.000
Av. taxable income: age 45-50	29384	1021286	30325	32655	2331***	0.000
Av. taxable income fathers: age 35-55	19855	957602	20128	22917	2789***	0.000
Av. taxable income fathers: age 45-50	20895	834960	20941	23553	2612***	0.000
Household taxable income in 1970	19781	1058773	20572	20432	-141	0.100
Father's education: Below upper secondary	0.723	1059587	0.711	0.599	-0.113***	0.000
Father's education: Upper Secondary	0.142	1059587	0.145	0.217	0.072***	0.000
Father's education: Short tertiary	0.066	1059587	0.068	0.090	0.023***	0.000
Father's education: Bachelor	0.034	1059587	0.037	0.045	0.009***	0.000
Father's education Master+	0.036	1059587	0.039	0.048	0.009***	0.000
Mother's education: Below upper secondary	0.747	1042141	0.734	0.615	-0.118***	0.000
Mother's education: Upper Secondary	0.173	1042141	0.179	0.254	0.075***	0.000
Mother's education: Short tertiary	0.040	1042141	0.042	0.074	0.032***	0.000
Mother's education: Bachelor	0.029	1042141	0.033	0.038	0.006***	0.000
Mother's education: Master+	0.011	1042141	0.012	0.018	0.006***	0.000
Father's occupation in 1970: salaried employee	0.229	1046923	0.241	0.286	0.045***	0.000
Father's occupation in 1970: manual worker	0.443	1046923	0.450	0.495	0.044***	0.000
Father's occupation in 1970: employer	0.044	1046923	0.046	0.036	-0.010***	0.000
Father's occupation in 1970: self or assisting family business	0.227	1046923	0.221	0.161	-0.060***	0.000
Father's occupation in 1970: no occupation	0.057	1046923	0.041	0.021	-0.020***	0.000
Mother's occupation in 1970: salaried employee	0.302	1029272	0.318	0.389	0.071***	0.000
Mother's occupation in 1970: manual worker	0.371	1029272	0.374	0.390	0.016***	0.000
Mother's occupation in 1970: employer	0.027	1029272	0.027	0.023	-0.004***	0.000
Mother's occupation in 1970: self or assisting family business	0.258	1029272	0.251	0.179	-0.072***	0.000
Mother's occupation in 1970: no occupation	0.042	1029272	0.031	0.020	-0.011***	0.000

Notes: Mean, sample size, mean in 1960 and in 1966, difference and p-value for the difference.

Table A-2: Treatment effect weights by regional group and birth cohort used by the TWFE estimator treatment (de Chaisemartin and D'Haultfoeuille, 2020)

Regional group	Birth cohort	Natural weights	TWFE weights
1961	1961	.0319277	.0965172
1961	1962	.0311166	.067806
1961	1963	.0304707	.0284394
1961	1964	.028868	-.0088459
1961	1965	.0267845	-.0440105
1961	1966	.0259928	-.0653522
1962	1962	.0437973	.1367522
1962	1963	.0433699	.0813894
1962	1964	.0413739	.0263496
1962	1965	.0389892	-.0272864
1962	1966	.0382899	-.0601514
1963	1963	.0656668	.1820284
1963	1964	.0633891	.0971269
1963	1965	.0614804	.0120209
1963	1966	.0603245	-.0407541
1964	1964	.0652152	.1560288
1964	1965	.0643167	.0679064
1964	1966	.0617961	.0114143
1965	1965	.0663371	.1277774
1965	1966	.0674687	.0711848
1966	1966	.0430251	.0836586

Note: Weights assigned by the TWFE estimator to the treatment effect of different regional groups and birth cohorts when estimating the overall treatment effect coefficient δ (de Chaisemartin and D'Haultfoeuille, 2020). The natural weights are calculated as the relative size of the region-birth cohort groups during the treatment which ensure that each treated individual is weighted equally. The sum of each set of weights is equals to one.

Table A-3: Heterogeneous DiD estimates on achieving tertiary education: father without a secondary education degree

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men	Women	Men	Women	Men	Women	Men	Women
δ	0.008	0.035***	0.005	0.032***	0.006	0.036***	0.016***	0.029***
se	0.007	0.005	0.006	0.005	0.010	0.006	0.006	0.006
p-value	0.211	0.000	0.391	0.000	0.572	0.000	0.010	0.000
p-value placebo	0.991	0.375	0.908	0.871	0.000	0.000	0.526	0.588
N	374,224	351,595	357,679	336,296	357,679	336,296	357,679	336,296
Method	Imput	Imput	Imput	Imput	CS	CS	Imput	Imput
Parental background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	No	No	No	No	Yes	Yes

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1965 and without mobility restriction, having a father without a secondary education degree. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men: columns (1), (3), (5), (7); women: columns (2), (4), (6), (8)). Columns (1) and (2) implement the imputation method (Borusyak et al., 2024). In columns (3) and (4) we implement a conditional DiD estimator by including to the imputation model control variables on the parental background (father's and mother's education, occupation, and household income in 1970), which explain both the level and differential trends. Columns (5) and (6) implement the doubly robust conditional difference-in-differences as in Callaway and Sant'Anna (2021), which assumes no compositional changes over birth cohorts. In columns (7) and (8), we rely on a conditional trend-adjusted DiD estimator by implementing the imputation method, controlling for both parental background and linear differences in trends by region. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table A-4: Heterogeneous DiD estimates on achieving tertiary education: father with a secondary education degree

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men	Women	Men	Women	Men	Women	Men	Women
δ	0.011	0.023***	0.004	0.016	0.010	0.018*	0.004	–0.002
se	0.009	0.009	0.006	0.010	0.010	0.009	0.013	0.014
p-value	0.231	0.008	0.533	0.132	0.302	0.053	0.756	0.886
p-value placebo	0.771	0.018	0.741	0.608	0.000	0.000	0.855	0.755
N	137,704	129,606	133,728	126,046	133,728	126,046	133,728	126,046
Method	Imput	Imput	Imput	Imput	CS	CS	Imput	Imput
Parental background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	No	No	No	No	Yes	Yes

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1965 and without mobility restriction, having a father with at least a secondary education degree. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men: columns (1), (3), (5), (7); women: columns (2), (4), (6), (8)). Columns (1) and (2) implement the imputation method (Borusyak et al., 2024). In columns (3) and (4) we implement a conditional DiD estimator by including to the imputation model control variables on the parental background (father's and mother's education, occupation, and household income in 1970), which explain both the level and differential trends. Columns (5) and (6) implement the doubly robust conditional difference-in-differences as in Callaway and Sant'Anna (2021), which assumes no compositional changes over birth cohorts. In columns (7) and (8), we rely on a conditional trend-adjusted DiD estimator by implementing the imputation method, controlling for both parental background and linear differences in trends by region. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table A-5: Heterogeneous DiD estimates on achieving tertiary education: father without a secondary education degree - Removing Helsinki

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men	Women	Men	Women	Men	Women	Men	Women
δ	0.017***	0.026***	0.015***	0.022***	0.018**	0.018**	0.024***	0.018**
se	0.005	0.006	0.005	0.007	0.007	0.008	0.007	0.008
p-value	0.001	0.000	0.006	0.001	0.010	0.023	0.000	0.021
p-value placebo	0.955	0.756	0.960	0.946	0.002	0.555	0.319	0.941
N	322,557	301,207	308,670	288,495	308,670	288,495	308,670	288,495
Method	Imput	Imput	Imput	Imput	CS	CS	Imput	Imput
Parental background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	No	No	No	No	Yes	Yes

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1964 and without mobility restriction, having a father without a secondary education degree and not living in Helsinki. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men: columns (1), (3), (5), (7); women: columns (2), (4), (6), (8)). Columns (1) and (2) implement the imputation method (Borusyak et al., 2024). In columns (3) and (4) we implement a conditional DiD estimator by including to the imputation model control variables on the parental background (father's and mother's education, occupation, and household income in 1970), which explain both the level and differential trends. Columns (5) and (6) implement the doubly robust conditional difference-in-differences as in Callaway and Sant'Anna (2021), which assumes no compositional changes over birth cohorts. In columns (7) and (8), we rely on a conditional trend-adjusted DiD estimator by implementing the imputation method, controlling for both parental background and linear differences in trends by region. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table A-6: Heterogeneous DiD estimates on achieving tertiary education: father with a secondary education degree - Removing Helsinki

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Men	Women	Men	Women	Men	Women	Men	Women
δ	0.011	0.011	0.003	0.006	–0.003	0.003	–0.009	–0.012
se	0.010	0.009	0.008	0.008	0.013	0.013	0.014	0.012
p-value	0.278	0.194	0.673	0.470	0.814	0.820	0.509	0.330
p-value placebo	0.658	0.122	0.810	0.116	0.113	0.353	0.584	0.375
N	100,064	93,891	97,091	91,245	97,091	91,245	97,091	91,245
Method	Imput	Imput	Imput	Imput	CS	CS	Imput	Imput
Parental background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	No	No	No	No	Yes	Yes

Notes: Treatment effect estimates obtained by implementing several difference-in-differences estimators allowing for heterogeneous effects by regional group and birth cohorts. Data: individuals born between 1951 and 1964 and without mobility restriction, having a father with at least a secondary education degree and not living in Helsinki. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men: columns (1), (3), (5), (7); women: columns (2), (4), (6), (8)). Columns (1) and (2) implement the imputation method (Borusyak et al., 2024). In columns (3) and (4) we implement a conditional DiD estimator by including to the imputation model control variables on the parental background (father's and mother's education, occupation, and household income in 1970), which explain both the level and differential trends. Columns (5) and (6) implement the doubly robust conditional difference-in-differences as in Callaway and Sant'Anna (2021), which assumes no compositional changes over birth cohorts. In columns (7) and (8), we rely on a conditional trend-adjusted DiD estimator by implementing the imputation method, controlling for both parental background and linear differences in trends by region. We report the absolute effect, standard errors, and p-value of the effect, along with the p-value of joint significance from the placebo test on the five birth cohorts before the reform. Standard errors are clustered at the municipality-of-birth level. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table A-7: Imputation method: estimates on intergenerational income elasticity (women)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	35-55	45-50	35-55	45-50	35-55	45-50	35-55	45-50	35-55	45-50	35-55	45-50
	Log	Log	IHS	IHS	pct	pct	Log	Log	IHS	IHS	pct	pct
δ	0.017***	0.025**	0.017***	0.025**	0.623***	0.887***	0.006	0.016	0.006	0.016	0.498*	0.881***
se	0.001	0.013	0.001	0.014	0.006	0.000	0.297	0.170	0.295	0.172	0.058	0.004
p-value	0.005	0.010	0.005	0.010	0.228	0.227	0.006	0.011	0.006	0.012	0.262	0.306
Placebo:												
p-value	0.618	0.158	0.621	0.156	0.617	0.646	0.509	0.620	0.512	0.616	0.538	0.724
θ	-0.010	-0.008	-0.010	-0.008	-0.017***	-0.014**	-0.012*	-0.010	-0.012*	-0.010	-0.020***	-0.015***
se	0.006	0.011	0.006	0.011	0.005	0.006	0.007	0.011	0.007	0.011	0.006	0.006
p-value	0.115	0.464	0.119	0.479	0.002	0.016	0.062	0.363	0.065	0.378	0.000	0.009
Placebo:												
θ	-0.036***	-0.018**	-0.036***	-0.018**	-0.032***	-0.020***	-0.035***	-0.020***	-0.035***	-0.020***	-0.024***	-0.015**
se	0.007	0.009	0.007	0.009	0.006	0.006	0.006	0.005	0.006	0.005	0.006	0.007
p-value	0.000	0.048	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024
N	414,142	352,972	414,142	352,972	414,142	352,972	414,142	352,972	414,142	352,972	414,142	352,972
Linear trend	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

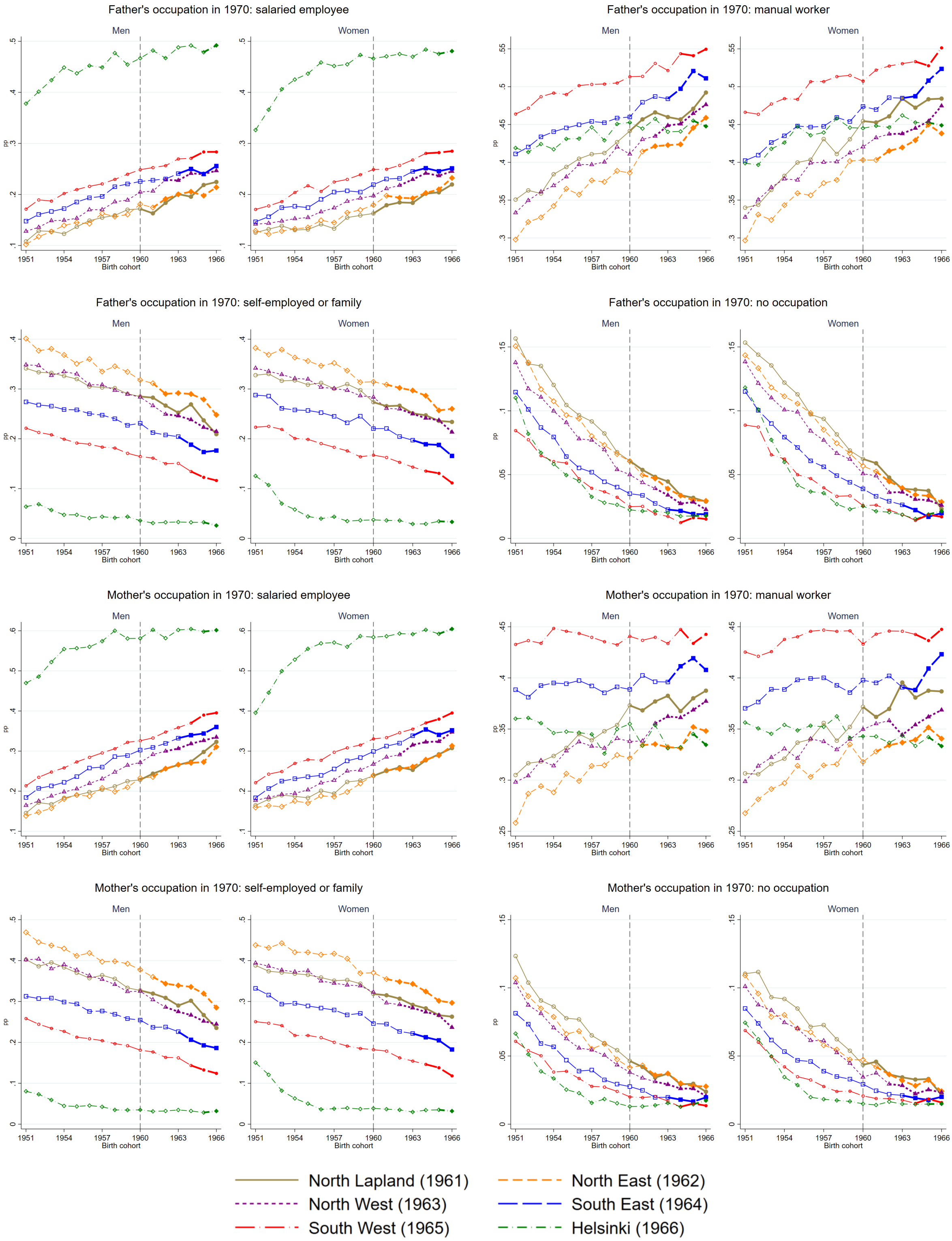
Notes: The table presents the estimates obtained using the imputation estimator. The data include women born between 1951 and 1965 and without mobility restrictions. The outcome variable is the child's average imposable income within a given age bracket and transformation. The heterogeneity variable is the father's average income in the same age bracket. For the treatment effect, we report the average effect (δ) and the heterogeneity by father's income (θ), along with their respective standard errors and p-values. For the placebo tests, we report the p-value for the joint significance of the placebo dummies (starting five periods earlier) for the overall effect and the p-value for the heterogeneity of the effects. The latter is calculated similarly to θ but by rerunning the same imputation method, excluding the units when they enter treatment, and starting the placebo five periods earlier. Age brackets: 35–55 for columns (1), (3), (5), (7), (9), and (11); 45–50 for columns (2), (4), (6), (8), (10), and (12). Income transformations: Logarithm for columns (1), (2), (7), and (8); inverse hyperbolic sine for columns (3), (4), (9), and (10); percentile position within the birth cohort for columns (5), (6), (11), and (12). Standard errors are clustered at the municipality-of-birth level. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table A-8: Imputation method: estimates on intergenerational income elasticity (women) - Removing Helsinki

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct	35-55 Log	45-50 Log	35-55 IHS	45-50 IHS	35-55 pct	45-50 pct
δ	-0.003	-0.004	-0.003	-0.005	-0.221	-0.038	-0.001	-0.002	-0.001	-0.002	0.127	0.173
se	0.536	0.454	0.540	0.443	0.336	0.874	0.840	0.807	0.843	0.797	0.642	0.558
p-value	0.005	0.006	0.005	0.006	0.230	0.242	0.006	0.008	0.006	0.008	0.273	0.295
Placebo: p-value	0.978	0.702	0.979	0.698	0.461	0.872	0.963	0.786	0.964	0.783	0.914	0.873
θ	0.012	0.012	0.012	0.012	0.011	0.011	0.012	0.012	0.012	0.012	0.011	0.011
se	0.008	0.008	0.008	0.008	0.007	0.008	0.008	0.008	0.008	0.008	0.008	0.009
p-value	0.136	0.108	0.134	0.110	0.138	0.199	0.144	0.114	0.142	0.115	0.141	0.194
Placebo: θ	-0.017*	-0.002	-0.017*	-0.002	-0.006	-0.002	-0.003	-0.010	-0.003	-0.010	-0.001	-0.008
se	0.009	0.010	0.009	0.010	0.010	0.010	0.008	0.010	0.008	0.010	0.008	0.009
p-value	0.054	0.832	0.053	0.824	0.542	0.809	0.675	0.323	0.673	0.319	0.864	0.359
N	336,981	284,446	336,981	284,446	336,981	284,446	336,981	284,446	336,981	284,446	336,981	284,446
Linear trend	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

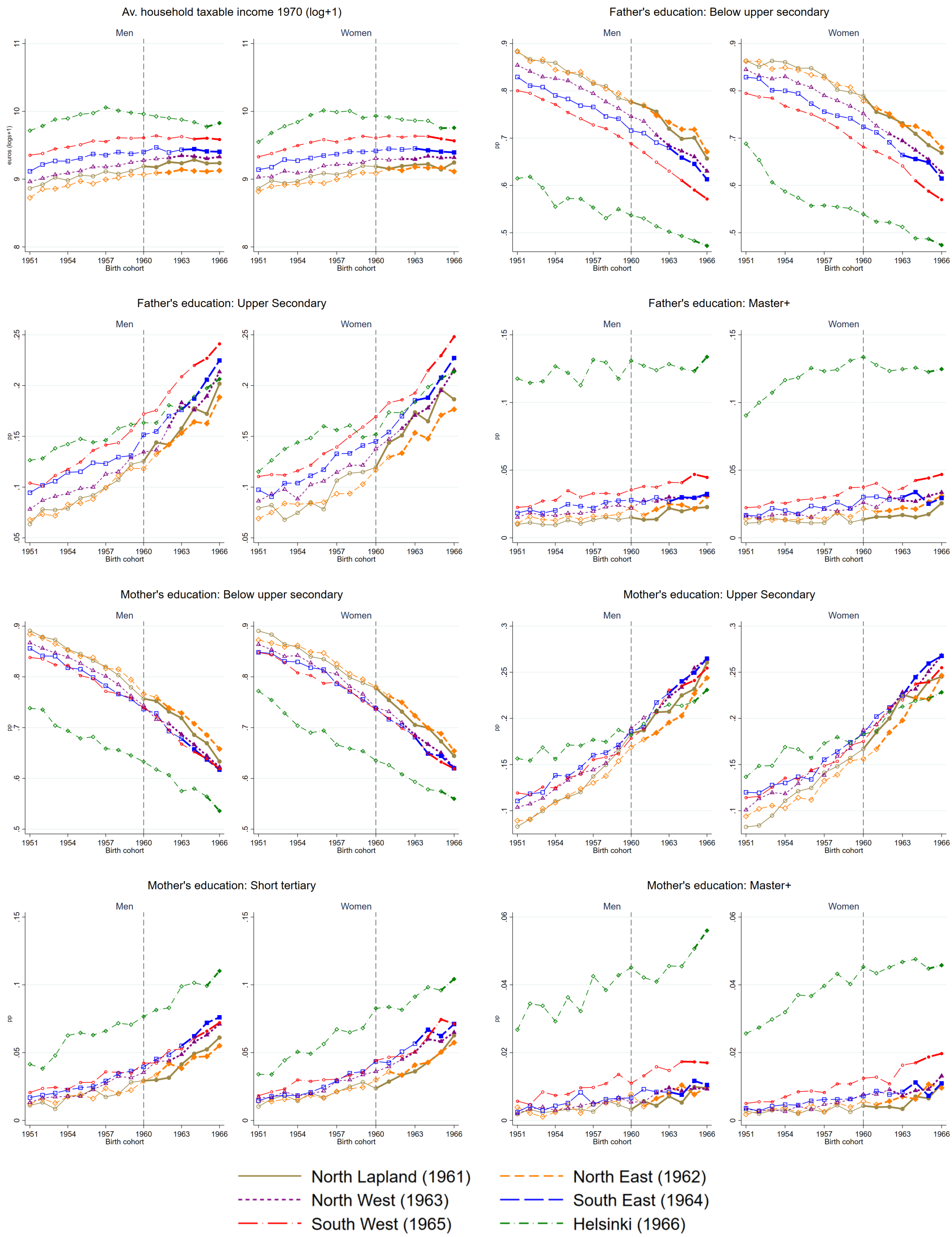
Notes: The table presents the estimates obtained using the imputation estimator. The data include women born between 1951 and 1964 and without mobility restrictions. The outcome variable is the child's average imposable income within a given age bracket and transformation. The heterogeneity variable is the father's average income in the same age bracket. For the treatment effect, we report the average effect (δ) and the heterogeneity by father's income (θ), along with their respective standard errors and p-values. For the placebo tests, we report the p-value for the joint significance of the placebo dummies (starting five periods earlier) for the overall effect and the p-value for the heterogeneity of the effects. The latter is calculated similarly to θ but by rerunning the same imputation method, excluding the units when they enter treatment, and starting the placebo five periods earlier. Age brackets: 35–55 for columns (1), (3), (5), (7), (9), and (11); 45–50 for columns (2), (4), (6), (8), (10), and (12). Income transformations: Logarithm for columns (1), (2), (7), and (8); inverse hyperbolic sine for columns (3), (4), (9), and (10); percentile position within the birth cohort for columns (5), (6), (11), and (12). Standard errors are clustered at the municipality-of-birth level. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Figure A-1: Evolution of parental occupation in 1970 over birth-cohort by regional groups



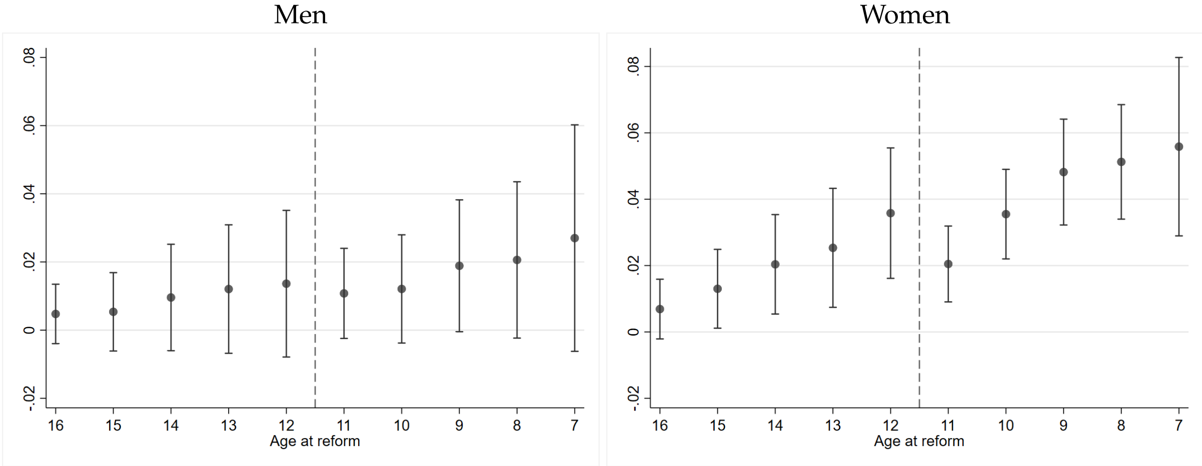
Notes: The figure shows the evolution of parental occupation across birth cohorts by regional groups. The staggered implementation of the reform between regional groups is depicted by using empty markers for birth cohorts not affected by the reform in a given regional group, and full markers with thicker lines for cohorts after the reform has been implemented in that group.

Figure A-2: Evolution of household income in 1970 and parental education over birth-cohort by regional groups



Notes: The figure shows the evolution of household income in 1970 and parental education across birth cohorts by regional groups. The staggered implementation of the reform between regional groups is depicted by using empty markers for birth cohorts not affected by the reform in a given regional group, and full markers with thicker lines for cohorts after the reform has been implemented in that group.

Figure A-3: Imputation estimates on achieving tertiary education: aggregation by age of children at time of reform (no covariates)



Notes: The figure shows the event-study estimates obtained by the imputation estimator (Borusyak et al., 2024) controlling for municipality of birth and birth-cohort fixed effects. Data: individuals born between 1951 and 1965 and without mobility restriction. The outcome variable is achieving a tertiary education degree by the age of 30. We provide results from a subgroup analysis by gender (men, left; women, right). The effects are aggregated by the individual's age at the time of the reform's implementation, corresponding to their birth cohort. Older birth cohorts, aged above 11 at the time of the reform's regional implementation, were not affected by the treatment as they had already completed primary school. The benchmark pre-treatment period of the DiD varies for the treatment effect (ages 7–11) and the placebo estimates (ages 12–16): the full pre-reform period is used for the treatment effect estimates, while the five years preceding the placebo period are used for the placebo estimates.

B Replication using Pekkarinen et al. (2009) sample

Table B-1: Descriptive statistics of Pekkarinen et al. (2009) sample

	Mean	N	1960	1966	Diff	p-value
Unmarried	.391	20824	.344	.466	.122	0
Married	.521	20824	.549	.472	-.077	0
Divorced	.087	20824	.106	.061	-.045	0
Widow	.001	20824	.001	.001	0	.948
Not moving	.816	20824	.868	.762	-.105	0
Tertiary degree	.103	20824	.102	.116	.014	.086
Years of education	13.042	20824	12.923	13.101	.177	.046
Academic degree	.403	20824	.378	.413	.035	.007
Avg. ln income (2000)	29778	20824	30032	33229	3197	.557
Avg. ln income 25-35	9.768	20804	9.717	9.843	.126	0
Avg. ln income (2000)	10.015	20824	10.019	10.003	-.016	.438
Mother's income	12083	20687	9995	14370	4375	0
Mother's avg. ln income 1970's	8.836	19218	8.789	8.866	.077	.003
Mother's avg. ln income 1980's	9.203	20169	9.102	9.293	.191	0
Mother's avg. ln income 70-80's	9.044	20395	8.96	9.115	.155	0
Father's income	14420	20824	11975	16868	4892	0
Father's avg. ln income 1970's	9.655	20742	9.599	9.692	.093	0
Father's avg. ln income 1980's	9.7	19715	9.605	9.782	.178	0
Father's avg. ln income 70-80's	9.678	20824	9.605	9.738	.133	0
Parents' mean ln income 1970's	9.309	20804	9.257	9.346	.089	0
Father's education: missing	.299	20824	.38	.227	-.153	0
Father's education: basic	.443	20824	.412	.46	.048	0
Father's education: upper secondary	.137	20824	.107	.176	.069	0
Father's education: low tertiary	.059	20824	.05	.064	.014	.021
Father's education: bachelor	.033	20824	.028	.038	.01	.033
Father's education: MSc+	.029	20824	.024	.036	.012	.01
Mother's education: missing	.116	20824	.165	.082	-.083	0
Mother's education: basic	.589	20824	.605	.552	-.053	0
Mother's education: upper secondary	.203	20824	.165	.241	.077	0
Mother's education: low tertiary	.048	20824	.033	.069	.036	0
Mother's education: bachelor	.031	20824	.025	.04	.015	.002
Mother's education: MSc+	.012	20824	.007	.016	.009	.002

Table B-2: Replication of Table 3 of Pekkarinen et al. (2009)

	(1)	(2)	(3)	(4)
Father's earnings	0.277***	0.297***	0.298***	0.296***
	(0.014)	(0.011)	(0.010)	(0.014)
Reform		-0.063**	-0.019	
		(0.012)	(0.021)	
Father's earnings * reform		-0.055**	-0.069*	-0.066
		(0.009)	(0.022)	(0.031)
Constant	10.015***	10.043***	10.021***	10.012***
	(0.012)	(0.010)	(0.010)	(0.000)
N	20824	20824	20824	20824
r2	0.05	0.05	0.05	0.06

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

Table B-3: Replication of Table 4 of Pekkarinen et al. (2009)

	(1) Without early reformers	(2) Without Helsinki	(3) 1995-2000 earnings	(4) No top-coding	(5) No bottom-coding	(6) No top or bottom-coding	(7) Labor earnings
Father's earnings	0.311***	0.302***	0.251***	0.288***	0.297***	0.288***	0.276***
	(0.022)	(0.012)	(0.009)	(0.010)	(0.011)	(0.010)	(0.016)
Reform	-0.004	-0.008	-0.024	-0.018	-0.019	-0.018	-0.015
	(0.025)	(0.026)	(0.019)	(0.021)	(0.021)	(0.021)	(0.022)
Father's earnings * reform	-0.092	-0.074*	-0.047	-0.070*	-0.070*	-0.070*	-0.067
	(0.039)	(0.022)	(0.018)	(0.021)	(0.024)	(0.022)	(0.036)
Constant	10.002***	10.009***	9.903***	10.020***	10.021***	10.021***	9.981***
	(0.014)	(0.014)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
N	12040	18206	20824	20824	20824	20824	19177
r2	0.05	0.05	0.06	0.05	0.06	0.05	0.04

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

Table B-4: Replication of Table 5 of Pekkarinen et al. (2009)

	(1) 1st quintile	(2) 2nd quintile	(3) 3rd quintile	(4) 4th quintile	(5) 5th quintile
Reform	0.036	0.038	-0.037	-0.051	-0.080
	(0.045)	(0.040)	(0.038)	(0.041)	(0.048)
Constant	9.770***	9.918***	10.037***	10.096***	10.294***
	(0.025)	(0.022)	(0.021)	(0.022)	(0.026)
N	4165	4165	4165	4165	4164
r2	0.00	0.00	0.01	0.00	0.01

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

Table B-5: Alternative Table 5 specification with treatment interacted with quintiles

	(1) Income in 2000
Reform	-0.019
	(0.028)
Reform * Q1	0.036
	(0.033)
Reform * Q2	0.049
	(0.033)
Reform * Q4	0.001
	(0.033)
Reform * Q5	-0.067*
	(0.033)
Constant	10.027***
	(0.017)
N	20824
r2	0.05

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

C Replication of Pekkala Kerr et al. (2013)

Table C-1: Replication of Table 2 Column 2 of Pekkala Kerr et al. (2013)

	(1) Original	(2) Replication	(3) Updated TWFE	(4) Updated TWFE	(5) Imputation	(6) CS
Reformed	0.010	0.013	0.003	0.024	0.041**	0.039***
school	(-0.017,0.038)	(-0.015,0.041)	(-0.024,0.031)	(-0.012,0.059)	(0.018,0.061)	(0.015,0.055)
Pretrend					0.007	0.001
					(-0.021,0.035)	(-0.031,0.032)
N	141,765	135,506	147,174	91,510	91,510	91,510

confidence intervals in parentheses

* p 0.05, ** p 0.01, *** p 0.001

Notes: The first column shows the values in Pekkala Kerr et al. (2013). The second column shows the replication with TWFE. In the third column, we include movers in the sample, control for municipality fixed effects and cluster standard errors at the municipality level, and do not control for age at test, which might be affected by the reform. The fourth column shows the results under the same specification as in column (3) but with the more restricted sample that is used with the imputation method and the Callaway and Sant'Anna (2021) estimator. In the fifth column, we estimate the effect with the imputation method by Borusyak et al. (2024) in an analogous specification to columns (3) and (4). The sample size is smaller in column (5) because the imputation method cannot estimate the reform effect in regions where the reform took place in 1972 or 1973 since all cohorts in those regions are treated. The sixth column estimates the effect of the reform using the estimator by Callaway and Sant'Anna (2021).

Table C-2: Replication of Table 3 Column 1 of Pekkala Kerr et al. (2013)

	(1) Original	(2) Replication	(3) Updated TWFE	(4) Updated TWFE	(5) Imputation	(6) CS
Math	0.002	0.004	-0.005	0.014	0.019	0.021
	(-0.025,0.030)	(-0.024,0.032)	(-0.032,0.023)	(-0.021,0.050)	(-0.001,0.039)	(-0.000,0.042)
Verbal	0.023	0.024	0.014	0.039*	0.048***	0.055***
	(-0.004,0.051)	(-0.004,0.052)	(-0.013,0.042)	(0.004,0.075)	(0.026,0.069)	(0.032,0.077)
Logical	0.006	0.005	-0.000	0.006	0.026*	0.029**
reasoning	(-0.022,0.033)	(-0.023,0.034)	(-0.028,0.027)	(-0.029,0.042)	(0.004,0.048)	(0.007,0.052)
N	141,765	135,506	147,174	91,510	91,510	91,510

confidence intervals in parentheses

* p 0.05, ** p 0.01, *** p 0.001

Notes: The first column shows the values in Pekkala Kerr et al. (2013). The second column shows the replication with TWFE. In the third column, we include movers in the sample, control for municipality fixed effects and cluster standard errors at the municipality level, and do not control for age at test, which might be affected by the reform. The fourth column shows the results under the same specification as in column (3) but with the more restricted sample that is used with the imputation method and the Callaway and Sant'Anna (2021) estimator. In the fifth column, we estimate the effect with the imputation method by Borusyak et al. (2024) in an analogous specification to columns (3) and (4). The sample size is smaller in column (5) because the imputation method cannot estimate the reform effect in regions where the reform took place in 1972 or 1973 since all cohorts in those regions are treated. The sixth column estimates the effect of the reform using the estimator by Callaway and Sant'Anna (2021).

Table C-3: Replication of Table 6 Column 1 of Pekkala Kerr et al. (2013)

	(1) Original	(2) Replication TWFE	(3) Updated TWFE	(4) Imputation Method
Reform	0.031	0.022	0.012	
	(-0.003,0.065)	(-0.011,0.054)	(-0.020,0.043)	
Reform × high educated parents	-0.035*	-0.024	-0.021	
	(-0.070,-0.001)	(-0.057,0.009)	(-0.054,0.012)	
Reform, low educated parents				0.040**
				(0.012,0.068)
Reform, high educated parents				0.016
				(-0.008,0.041)
N	126,977	130,269	141,171	87,368

confidence intervals in parentheses

* p 0.05, ** p 0.01, *** p 0.001

Notes: The first column shows the values in Pekkala Kerr et al. (2013). The second column shows the replication with TWFE. In the third column, we include movers in the sample, control for municipality fixed effects and cluster standard errors at the municipality level, and do not control for age at test, which might be affected by the reform. The fourth column shows the results under the same specification as in column (3) but with the more restricted sample that is used with the imputation method and the Callaway and Sant'Anna (2021) estimator. In the fifth column, we estimate the effect with the imputation method by Borusyak et al. (2024) in an analogous specification to columns (3) and (4). The sample size is smaller in column (5) because the imputation method cannot estimate the reform effect in regions where the reform took place in 1972 or 1973 since all cohorts in those regions are treated. The sixth column estimates the effect of the reform using the estimator by Callaway and Sant'Anna (2021). Columns (2)-(6) also differ from (1) in that we use register based parent-child links instead of linking parents with children using household codes from the census as in Pekkala Kerr et al. (2013)

Table C-4: Replication of Table 7 Column 1 of Pekkala Kerr et al. (2013)

	(1) Original	(2) Replication TWFE	(3) Updated TWFE	(4) Imputation Method
Reform	0.014	0.013	0.004	0.023*
	(-0.015,0.042)	(-0.015,0.041)	(-0.025,0.032)	(0.004,0.042)
Reform × parents' income	-0.034*	-0.040*	-0.030	-0.013
	(-0.066,-0.002)	(-0.068,-0.012)	(-0.061,0.000)	(-0.052,0.026)
N	126,891	127,783	127,783	85,640

confidence intervals in parentheses

* p 0.05, ** p 0.01, *** p 0.001

Notes: The first column shows the values in Pekkala Kerr et al. (2013). The second column shows the replication with TWFE. In the third column, we include movers in the sample, control for municipality fixed effects and cluster standard errors at the municipality level, and do not control for age at test, which might be affected by the reform. The fourth column shows the results under the same specification as in column (3) but with the more restricted sample that is used with the imputation method and the Callaway and Sant'Anna (2021) estimator. In the fifth column, we estimate the effect with the imputation method by Borusyak et al. (2024) in an analogous specification to columns (3) and (4). The sample size is smaller in column (5) because the imputation method cannot estimate the reform effect in regions where the reform took place in 1972 or 1973 since all cohorts in those regions are treated. The sixth column estimates the effect of the reform using the estimator by Callaway and Sant'Anna (2021). Columns (2)-(6) also differ from (1) in that we use register based parent-child links instead of linking parents with children using household codes from the census as in Pekkala Kerr et al. (2013)