Education, Language, and Identity
Evidence from Nation-Building in Northern Sweden

Alexandra Sandström*
Uppsala University

16 January, 2026

Abstract

How do minorities respond to state-led assimilation efforts? I study this question by evaluating a nation-building reform that introduced state-funded primary schools using Swedish rather than Finnish as the language of instruction in late nineteenth-century Tornedalen, a predominantly Finnish-speaking region of northern Sweden. Using linked full-population census data and exploiting differential exposure to school openings across space and cohorts, I estimate the causal effects of exposure to the reform. I find that individuals exposed to Swedish-only schools were less likely to give their children Swedish names and more likely to give them minority names, indicating reduced cultural assimilation. The effect is driven by treated males, who are also the individuals who increased educational attainment. These effects emerge despite ongoing cultural assimilation in the surrounding community, as reflected in naming practices at the time the schools opened, and are concentrated in places with no prior educational supply, measured by teacher presence.

Word count: 2676

*Department of Economics. Email: alexandra.sandstrom@nek.uu.se.

1 Introduction

Nations often use the education system to promote national identity and assimilate minorities. While such nation-building efforts may strengthen social cohesion and improve economic outcomes, this is not guaranteed (Carvalho et al., 2024; Fouka, 2022). Individuals and groups may resist, and rather than reinforcing a shared national identity, these policies can undermine it, potentially harming social cohesion and affecting outcomes such as cooperation, trust, public goods provision, and political polarization (Rohner and Zhuravskaya, 2023). These reforms may also have direct economic consequences through their effects on the uptake of education. Even where reforms expand access or improve quality, they may fail to deliver benefits if individuals choose not to attend or engage due to the cultural component, potentially exacerbating economic disparities.

In this paper, I examine effects on cultural assimilation of an educational reform introduced at the turn of the 20th century in Tornedalen, a region in northern Sweden bordering Finland. Prior to the reform, the language of instruction in the region was almost exclusively Finnish. Although parishes were legally required to provide six years of compulsory education for children aged 7 to 12, attendance was low due to poverty, teacher shortages, and long travel distances. To assimilate the poor and remote region economically and culturally, the state offered funded primary schools, with funding conditional on using Swedish as the language of instruction; Finnish was not to be used. The schools were introduced gradually stating in 1889 and over time the Swedish instruction spread to all types of primary schools. The Swedish-only schools were not dismantled until 1940 and the ban on Finnish was not lifted until 1957 (Regeringskansliet, 2023).

To identify the causal effect, I employ a difference-in-differences design that leverages two sources of variation: within cohorts, by comparing villages that received a state-funded school to those that did not; and within villages, by comparing affected cohorts (ages 4–12) to cohorts to old for primary school at the opening (ages 13–21). My design relies on parallel trends in outcomes between younger and older cohorts in the absence of treatment. While this assumption allows villages to differ in levels, it could be violated if treated villages are on different trajectories than control villages. Therefore, I select similar control villages using nearest-neighbor matching based on propensity scores. I further present event-study plots and conduct placebo tests using pre-treatment characteristics, showing that treatment does not predict a set of socioeconomic and cultural variables at the individual level.

I find that exposure to the reform decreased cultural assimilation, as measured by naming practices.

Exposed individuals gave their children less Swedish names and more minority names on average, as captured by a Cultural Name Score (CNS). The CNS is calculated based on the relative frequency of a given name in one cultural group relative to another and has been widely used to measure cultural assimilation (Fryer and Levitt, 2004; Fouka, 2020; Gagliarducci and Tabellini, 2024). The effect is driven by treated males and increase with time of exposure.

To understand the mechanisms underlying these results, I first examine whether the effect operates through school attendance or whether the mere opening of a school is sufficient. I test this by assessing whether educational attainment on the extensive margin increased for exposed individuals and find that it did. As before, the effect is driven by males, indicating that attendance rather than mere school presence is the relevant channel. Importantly, males and females had similar educational attainment prior to treatment, implying that the gendered response reflects differential reactions to the cultural component of schooling rather than baseline differences in educational propensity. Moreover, the effect is concentrated in locations with no prior supply of education, as measured by teacher presence before treatment. I interpret these locations as ones in which no alternative formal schooling options were available. As a result, individuals were less able to choose Finnish-language education—if that was their preference—and were more likely to attend the newly established schools.

Moreover, I assess whether the broader community exhibited increased assimilation following school openings, as measured by naming practices. While an increase in Swedish naming would indicate community-level resistance to assimilation through the schools, I find the opposite, indicating that the surrounding community culturally assimilated in response to the reform.

My paper contributes to the literature on nation-building through education. Some studies have isolated the effects of curriculum (Cantoni et al., 2017) or language of instruction (Fouka, 2020; Clots-Figueras and Masella, 2013), but nation-building policies are often implemented as bundles, and their combined effects need not equal the sum of their parts (Carlitz et al., 2025; Blanc and Kubo, 2025). The combination of additional schools and a change in the language of instruction studied in this paper is policy-relevant, since in reality policies are often introduced into a preexisting school system.

Furthermore, I contribute to the literature on resistance to assimilation policies in the educational system, which can take several forms, such as the opening of alternative schools (Bazzi et al., 2025) or withdrawal from, or non-engagement with, formal education (Meyersson, 2014). Resistance may also arise even when individuals attend schools, and some studies find that such backlash effects are stronger in communities with smaller minority populations (Fouka, 2020; Maruthiah, 2025), pointing to an important role for group-level resistance and for cultural transmission by the surrounding environment in reaction to the policy (Bisin and Verdier, 2001; Carvalho et al., 2024), or to stronger demand for signals of commitment to the group (Fouka, 2022). However, resistance may also operate at the individual level as a psychological response to coercive policies (Schøyen, 2021), or through identity concerns (Shayo, 2009). I contribute to this literature by providing empirical evidence consistent with individual-level rather than community-level resistance.

Finally, I contribute to the literature on education policy in developing settings by examining how language of instruction and schooling expansion shape long-run outcomes. While historical, the analysis is relevant for contemporary low-income contexts where basic education remains a policy priority (Aghion et al., 2025). I show that language policy affects educational uptake not only through learning quality (Lloyd and Yang, 2025; Taylor and von Fintel, 2016) but also through its cultural content, generating gendered responses. Meyersson (2014) shows that females can be particularly responsive to the cultural components of education, documenting that women's educational attainment in Turkey increased following exposure to more religious schools. I show effects for a more gender-neutral cultural component; language.

2 Background

Finland was part of Sweden for nearly 700 years before becoming an autonomous region of the Russian Empire in 1809. The new border, drawn along rivers, divided Tornedalen, a culturally connected Finnish-speaking region. When Finland gained independence in 1917, the border remained. Today, Finnish is mainly spoken on the Finnish side and Swedish on the Swedish side, though cultural ties persist. Some residents on the Swedish side still speak Finnish as a second language, and the local Finnish variety, Meänkieli, is recognized as a protected minority language in Sweden.

Sweden's educational policy played a key role in shaping language use on the Swedish side. Before 1889, Finnish was the main language of instruction in Tornedalen, reflecting its dominance in the region, and there was no strong political will to change this (Hyltenstam and Salö, 2023). However, starting in 1889, the state introduced publicly funded primary schools to culturally and economically assimilate the region: four schools opened in 1889, sixteen in 1894, and two additional schools in 1895 and 1897 (Tenerz, 1960; Borgh, 1889; Lindell, 1896). State funding required instruction exclusively in Swedish. These schools were not forcibly imposed; opening one was voluntary, but every parish was required to provide basic education, and many parishes struggled economically to do so (Andersson and Berger, 2019). Swedish-only instruction gradually spread to parish-run schools (Elenius, 2023), and by 1912 Finnish had been entirely removed from the curriculum, both as a subject and as a language of instruction (Hyltenstam and Salö, 2023).

The policy of forbidding Finnish in schools was locally known as the "All-Swedish Method," which required Finnish-speaking children to use only Swedish from their first day of schooling. The alternative approach used was bilingual, relying on students' native language to support the learning of Swedish through textbooks with parallel Swedish–Finnish texts (Elenius, 2023).

3 Data

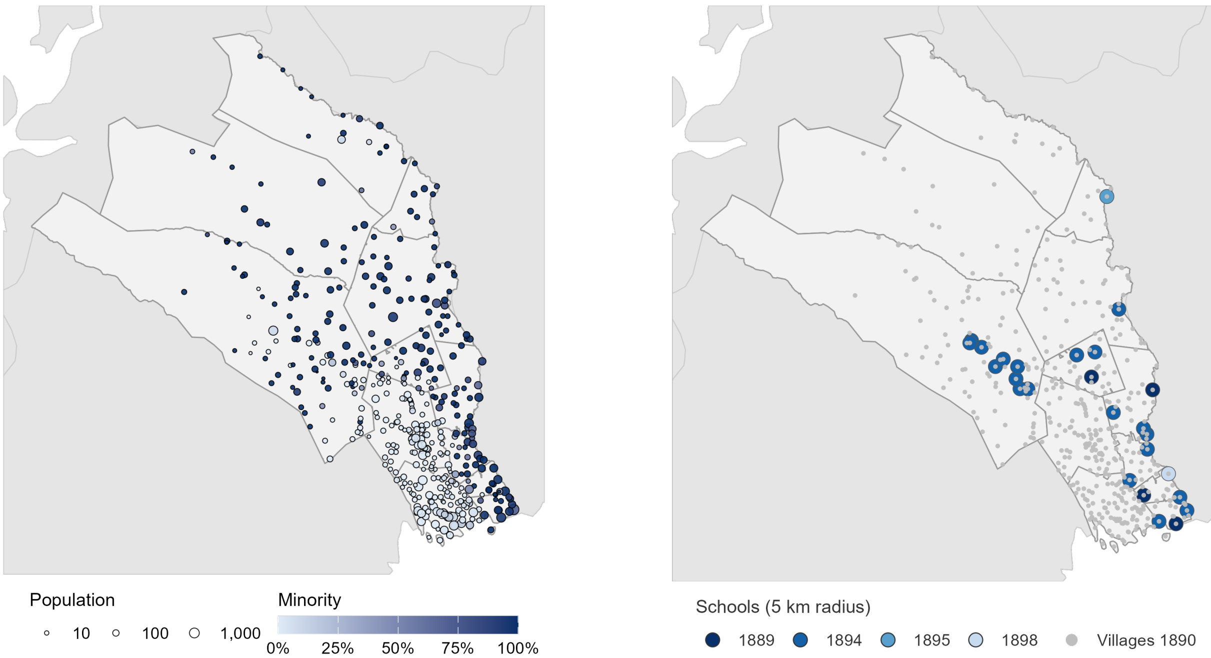
My primary data source is the Swedish population censuses which I link over time following Abramitzky et al. (2021); Wisselgren et al. (2014); Berger et al. (2025). I obtain backwards linkage rates of approximately 68% for both genders when linking the 1890 census to the 1910 and 1930 censuses, respectively. Details are presented in Appendix A.2.

The census cover the entire Swedish population and are available from IPUMS for the years 1880, 1890, 1900, and 1910. The 1930 census is partially digitized and available from the Swedish National Archives and adds information on income from tax registers and self-reported education. Approximately two thirds of this census have been digitized. My region of interest is fully covered, but to address potential bias from incomplete digitization and selective migration, I weight observations by the probability of inclusion in the sample when relevant.

My main outcome variable is the Swedish CNS, which ranges from 0 to 100. A higher value indicates a ore Swedish name. Details are provided in Appendix A.1.

Information on Swedish-only school openings comes from primary and secondary sources (Tenerz, 1960; Borgh, 1889; Lindell, 1896) while information on other forms of schooling is inferred from occupational occupation in the 1890 census. I use teachers village of residence to determine local educational supply in 1890. This information is supplemented with teacher registers from 1889 and 1904 (Borgh, 1889; Lindell, 1896). These registers are not used exclusively because they cover only folkskolor and exclude småskolor, which were widespread during this period Andersson and Berger (2019).

I define a village as treated if it is located within 5 km of a state-funded school, a cut-off chosen based on historians' assessments of typical travel distances (Pers et al., 2022) as well as observations from my own archival work. To calculate distances to schools, I geocoded all villages by hand for all census waves. Figure 1 shows the spatial distribution of the villages in 1890 and the Swedish only schools as well as the 5-km radius around them.



(a) Villages pre-treatment

(b) Swedish-only schools

Figure 1: Villages in 1890 and Swedish-only schools

Notes: The left panel (a) shows the spatial distribution of villages prior to the opening of Swedish-only schools in 1890, with marker sizes weighted by population and shading indicating the share of individuals classified as Finnish minority in the census. The right panel (b) shows the location of schools in the region of interest in relation to the villages, together with a 5-km radius.

I select control villages using propensity score matching, choosing each village's five closest neighbour as its matched control while restricting control villages to those located more than 10 km from any school opening. I estimate the propensity scores using information on economic, demographic and cultural characteristics in 1890. Descriptive statistics at the village level in 1890 are shown in 1 for untreated villages, matched control villages, and treated villages. Matching clearly improves comparability.

4 Empirical strategy

4.1 Empirical specification

To estimate the individual-level effect of exposure to the Swedish-only schools, I use a difference-in-difference design. I compare cohorts that are too old for primary school $(13 - 21)$ with individuals at

Table 1: Comparability of treated and control villages in 1890

	Not treated	Matched control	Treated
Economic and demographic			
Population (mean)	76.07	148.90	175.66
Population (log)	3.48	4.28	4.42
Age (mean)	25.91	26.25	26.22
Males (share)	0.52	0.49	0.51
Females in the labor force (share)	0.08	0.06	0.07
Males in the labor force (share)	0.49	0.53	0.54
Farm housholds (share)	0.65	0.60	0.58
Household members (mean)	6.16	5.53	5.55
Education			
Any teacher 1880	0.07	0.10	0.13
Any teacher 1890	0.11	0.29	0.28
Culture			
Classified as Finnish minority (share)	0.46	0.71	0.72
Finnish name score (mean)	0.18	0.20	0.20
Swedish name score (mean)	0.19	0.18	0.18
Tornedalian name score (mean)	0.28	0.32	0.32
Villages	405	122	47

Notes: The table reports village-level characteristics measured in the 1890 census. Treated villages are defined as those located within 5 km of a Swedish-only school opening. Matched control villages are located more than 10 km from any school opening and consist of the five nearest neighbours selected using propensity score matching based on all variables reported in the table, except population size, for which the log of population is used.

or below schooling age ($4-12$) in villages located within or outside a 5-km radius of a Swedish-only schools. My main specification is

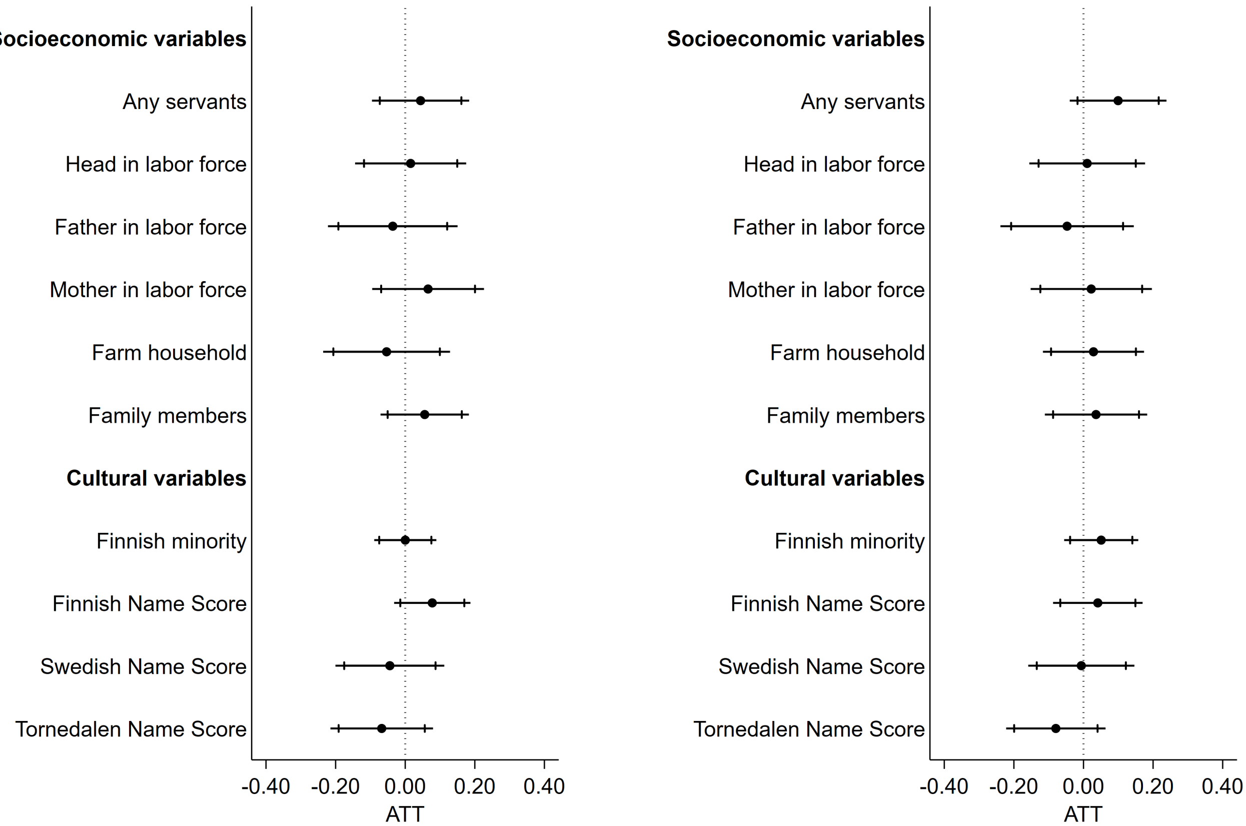
$$y_{i,l,c} = \beta\,\text{Treated}_{l,c} + \alpha_l + \alpha_c + \epsilon_{i,l,c} \tag{1}$$

where $y_{i,l,c}$ is an outcome for individual i from village l and birth cohort c, and $\text{Treated}_{l,c}$ is an indicator equal to one for individuals who were exposed to a school opening within 5km while they were of school age or younger. α_l are village fixed effects and α_c are cohort fixed effects. The coefficient of interest is β.

In an alternative specification, instead of using the indicator $\text{Treated}_{l,c}$, I estimate the effect of years treated, allowing the effect to vary with treatment intensity. One year corresponds to a school opening when an individual is age 12, while the maximum of nine years corresponds to exposure beginning at age 4.

Because Swedish-only schools were introduced at different points in time, relying on a two-way fixed-effects model to estimate Equation 1 would bias the estimates if treatment effects are heterogeneous (Goodman-Bacon, 2021; Sun and Abraham, 2021; de Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021). To alleviate this concern, I use a stacked difference-in-differences approach (Wing et al., 2024).

To increase confidence in the identifying assumption of parallel tends in outocmes between treated and control cohorts, I conduct placebo tests by estimating Equation 1 on pre-treatment characteristics. If cohorts are evolving in parallel, I should find no effects on characteristics determined before treatment. Figure 2 shows that this is indeed not the case.



(a) 1910 and 1930 combined

(b) 1930

Figure 2: Placebo

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and the 1910 and/or 1930 census in Panel (a), or only the 1930 census in Panel (b), and restricted to the matched village sample. The figure reports estimates from a stacked difference-in-differences specification, where the treatment is an indicator equal to one if an individual is exposed to a Swedish-only school between ages 4 and 12, and the outcomes are pre-treatment characteristics measured in 1890. All regressions include village and cohort stack-specific fixed effects. Coefficients are standardized to have mean zero and unit standard deviation. Estimates are shown with 95% and 90% confidence intervals based on standard errors clustered at the village level.

5 Results

5.1 Main results

Table 2 presents the results on treated individuals' assimilation choices. I find that treated individuals, on average, give their children less Swedish names. The estimated effect is -1.276, corresponding to a reduction of approximately 6% relative to the unconditional mean. The effect is concentrated among males: treated males give their children first names with a CNS that is 1.732 points lower than those of non-treated males. The estimate for females is smaller (-0.799) and is not statistically significant. Instead of giving their children Swedish names, treated individuals are more likely to give their children Finnish or Tornedalian names, as shown in Appendix Table B3. Appendix Figure B1 shows event-study plots. Although the estimates are noisy, there are no indications of systematic pre-trends.

Table 2: Swedish-only school exposure and their children's Swedish CNS

	Female & Male		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated (age 4-12)	-1.276**		-0.799		-1.732**	
	(0.518)		(0.952)		(0.833)	
Years treated		-0.217**		-0.071		-0.356**
		(0.097)		(0.157)		(0.153)
Village-Stack FE	✓	✓	✓	✓	✓	✓
Age-Stack FE	✓	✓	✓	✓	✓	✓
Dep. var. mean	20.124	20.124	20.097	20.097	20.152	20.152
Observations	7485	7485	3785	3785	3700	3700
R-squared	0.207	0.208	0.230	0.230	0.244	0.246

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and the 1910 and/or 1930 census, and restricted to the matched village sample. *Treated* is an indicator equal to one if an individual is exposed to a Swedish-only school between ages 4 and 12. *Years treated* is a continuous variable ranging from 1 to 9, capturing the number of years an individual is exposed to the Swedish-only school system (equal to 1 if the school opens when the individual is 12 years old and 9 if it opens when the individual is 4 years old). Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.2 Mechanism

5.2.1 School attendance vs. school opening

To assess whether the effect operates through actual school attendance, I examine whether educational attainment increased on the extensive margin, measured by any reported education in 1930. Table 3 reports results from the baseline specification and from a specification that interacts treatment with an indicator for pre-existing educational supply, measured by the presence of a teacher within 5 km prior to the school opening. In locations with pre-treatment educational supply, no effect on educational attainment is expected. The results show that educational attainment increases for men but not for women, and that the effect is driven by locations without prior educational supply. Baseline attainment is similar across genders, indicating differential responses to the reform rather than pre-treatment differences. Results are robust to a continuous specification (Table B4), and event-study estimates for areas without prior supply support the identifying assumptions (Figure B2).

Table 3: Swedish-only school exposure and education (extensive margin)

	Female & Male		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated (age 4-12)	0.043*	0.108***	0.010	0.045	0.083**	0.176***
	(0.026)	(0.032)	(0.031)	(0.050)	(0.040)	(0.061)
Treated (age 4-12) × Teacher before		-0.119***		-0.063		-0.173**
		(0.036)		(0.052)		(0.070)
Village-Stack FE	✓	✓	✓	✓	✓	✓
Age-Stack FE	✓	✓	✓	✓	✓	✓
Dep. var. mean	0.853	0.853	0.858	0.858	0.848	0.848
Unique 1890 IDs	4637	4637	2236	2236	2401	2401
R-squared	0.127	0.129	0.179	0.179	0.139	0.144

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 and 1930 censuses and restricted to the matched village sample. *Treated* is an indicator equal to one if an individual is exposed to a Swedish-only school between ages 4 and 12. *Teacher before* is an indicator equal to one if there was a teacher within 5 km of the village of residence in 1890. Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4 provides additional evidence for the mechanism operating through actual school attendance rather than school openings alone. The likelihood that a treated individual gives their child a less Swedish name is driven by locations that were less likely to have formal educational supply prior to treatment, which I interpret as individuals being less likely to have an alternative school to attend when the Swedish-only school opened and therefore being able to choose a Finish school if they where opposed to assimilating.

Table 4: Exposure by pre-treatment education supply and Swedish CNS

	Female & Male	Female	Male
	(1)	(2)	(3)
Treated (age 4-12)	-2.121***	-1.718*	-2.465***
	(0.585)	(0.955)	(0.852)
Treated (age 4-12) × Teacher before	1.297*	1.428	1.106
	(0.753)	(1.420)	(1.252)
Village-Stack FE	✓	✓	✓
Age-Stack FE	✓	✓	✓
Dep. var. mean	20.124	20.097	20.152
Observations	7,485	7,485	7,485
R-squared	0.208	0.231	0.244

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 and the 1910 and/or 1930 census censuses and restricted to the matched village sample. *Treated* is an indicator equal to one if an individual is exposed to a Swedish-only school between ages 4 and 12. *Teacher before* is an indicator equal to one if there was a teacher within 5 km of the village of residence in 1890. Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.2.2 Community-level vs. individual-level resistance

I examine community-level responses to the opening of Swedish-only schools by analyzing naming practices around school openings. Specifically, instead of using age at school opening as the treatment-defining variable, I examine names given to children born before versus after a school opening. I include sibling fixed effects to ensure that the results are not driven by selective migration into or out of villages, but instead reflect within-family comparisons between siblings born before and after the opening.

The event-study estimates from this analysis are shown in Figure 3, and results from a continuous difference-in-differences specification are reported in Table B5. The results indicate that children are given more Swedish names after school openings, suggesting community-level support for cultural assimilation rather than resistance and therefore pointing to individual-level resistance as the underlying mechanism.

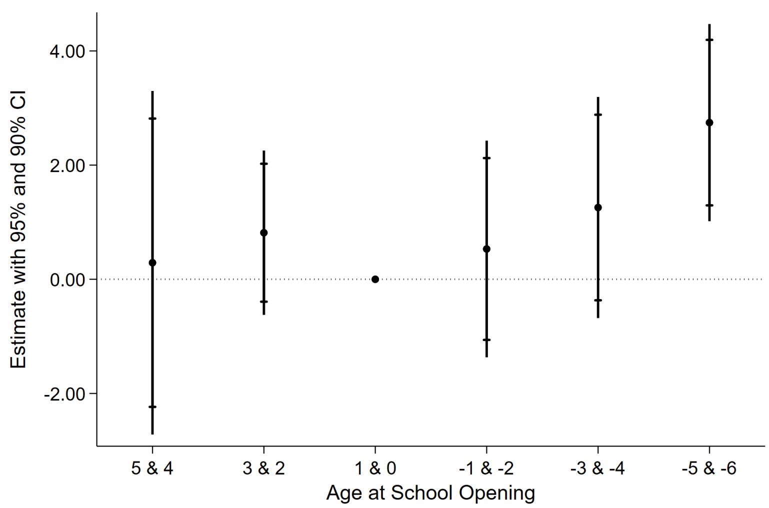


Figure 3: Swedish CNS around school opening

Notes: The sample consists of individuals born between six years before and five years after the opening of a Swedish-only school, observed in the 1900 census and restricted to the matched village sample. The figure reports estimates from a difference-in-differences specification, where the treatment is an indicator equal to one if an individual is born in a village experiencing a Swedish-only school opening, and the outcome is the Swedish CNS. The specification includes village, cohort, and sibling fixed effects. Estimates are shown with 95% and 90% confidence intervals based on standard errors clustered at the village level.

6 Conclusion

In this paper, I show that education with a cultural component—in this case, language—can affect cultural assimilation and educational uptake, and that these effects can depend strongly on gender, with implications for gender inequality. Furthermore, I show that cultural resistance emerged even though the community at large appeared supportive of cultural assimilation, suggesting that resistance can arise from individuals' own experiences with assimilation policies even when broader societal preferences favor cultural assimilation.

References

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., and Perez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3):865–918.

Aghion, P., Almås, I., and Meghir, C. (2025). Human capital and development. *NBER*, (w34602).

Andersen, L. and Bentzen, J. (2025). In the name of god! Religiosity and the production of science.

Andersson, J. and Berger, T. (2019). Elites and the expansion of education in nineteenth-century sweden. *Economic History Review*, 72(3):897–924.

Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2020). Frontier culture: The roots and persistence of "rugged individualism" in the United States. *Econometrica*, 88(6):2329–2368.

Bazzi, S., Hilmy, M., and Marx, B. (2025). Religion, education, and the state. *Review of Economic Studies*.

Berger, T., Karadja, M., and Prawitz, E. (2025). Cities and the rise of working women.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Blanc, G. and Kubo, M. (2025). The making of France.

Borgh, P. (1889). *Utförlig och fullständig matrikel öfver Sveriges folkskollärare, organister, kantorer och klockare samt seminarii-lärare, folkskoleinspektörer och lärare vid högre folkskolor*. M. W. Wallbergs Förlag, Norrköping.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Curriculum and ideology. *Journal of Political Economy*, 125(2):338–392.

Carlitz, R., Morjaria, A., Mueller, J., and Osafo-Kwaako, P. (2025). State building in a diverse society. *Review of Economic Studies*, 92(6):3704–3740.

Carvalho, J.-P., Koyama, M., and Williams, C. (2024). Resisting education. *Journal of the European Economic Association*, 22(6):2549–2597.

Cermeño, A. L., Enflo, K., and Lindvall, J. (2022). Railroads and reform: How trains strengthened the nation state. *British Journal of Political Science*, 52(2):715–735.

Clots-Figueras, I. and Masella, P. (2013). Education, language and identity. *The Economic Journal*, 123(570):F332–F357.

de Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Elenius, L. (2001). *Både finsk och svensk: Modernisering, nationalism och språkförändring i Tornedalen 1850–1939*. Umeå Univerity.

Elenius, L. (2023). Statlig minoritetspolitik för språklig assimilering: Historisk bakgrund till skol- och språkpolitik i Tornedalen samt genomgång av tidigare gjord forskning. In *Som om vi aldrig funnits: Tolv tematiska forskarrapporter*, volume 2023:68 of *Statens offentliga utredningar (SOU)*, pages 203–260. Kulturdepartementet, Regeringskansliet, Stockholm. Rapport för Sannings- och försoningskommissionen för tornedalingar, kväner och lantalaiset.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in US schools after World War I. *The Review of Economic Studies*, 87(1):204–239.

Fouka, V. (2022). Assimilation in historical political economy. In Jenkins, J. A. and Rubin, J., editors, *The Oxford Handbook of Historical Political Economy*. Oxford Academic.

Fryer, R. G. and Levitt, S. D. (2004). The causes and consequences of distinctively Black names. *The Quarterly Journal of Economics*, 119(3):767–805.

Gagliarducci, S. and Tabellini, M. (2024). Faith and assimilation: Italian immigrants in the US. *Economic Journal (forthcoming)*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Hyltenstam, K. and Salö, L. (2023). Språkideologi och det ofullbordade språkbytet: Den språkliga försvenskningen av det meänkielitalande området. In *Som om vi aldrig funnits. Tolv tematiska forskarrapporter*, volume 2023:68 of *Statens offentliga utredningar (SOU)*, pages 9–152. Kulturdepartementet, Regeringskansliet, Stockholm. Rapport i forskarrapportdelen av SOU 2023:68.

Jayes, J., Molinder, J., and Enflo, K. (2025). Power for progress: The impact of electricity on individual labor market outcomes. *Explorations in Economic History*, 98:101702.

Knudsen, A. S. (2025). Those who stayed: Selection and cultural change in the age of mass migration.

Lindell, P. (1896). *Sveriges folkskolelärarematrikel 1895*. K. L. Beckmans Boktryckeri, Stockholm.

Lloyd, T. and Yang, D. (2025). The long shadow of early education: Evidence from a natural experiment in the Philippines. *NBER*, (w33600).

Maruthiah, C. (2025). Assimilation policy, integration, and identity: Evidence from American Indian boarding schools.

Meyersson, E. (2014). Islamic rule and the empowerment of the poor and pious. *Econometrica*, 82:229–269.

Pers, A. S., Lagerqvist, M., and Björklund, A. (2022). Walk a mile in someone else's shoes: The difficult school route and how it was managed during the emergence of the Swedish folkskolan, 1840–1930. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, 76(1):1–13.

Regeringskansliet (2023). Som om vi aldrig funnits – exkludering och assimilering av Tornedalingar, Kväner och Lantalaiset.

Rohner, D. and Zhuravskaya, E., editors (2023). *Nation Building: Big Lessons from Successes and Failures*. CEPR Press, London.

Schøyen, Ø. (2021). What limits the efficacy of coercion? *Cliometrica*, 15(2):267–318.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review*, 103(2):147–174.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Taylor, S. and von Fintel, M. (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review*, 50:75–89.

Tenerz, H. (1960). *Folkupplysningsarbetet i Norrbottens finnbygd från äldsta tid till sekelskiftet 1900*. Föreningen för svensk undervisningshistoria, Stockholm.

Tillförordnade Folkskoleinspektörer (1896). *Berättelser om folkskolorna i riket*. Kungl. Boktryckeriet. P. A. Norstedt & Söner, Stockholm.

Wing, C., Freedman, S. M., and Hollingsworth, A. (2024). Stacked difference-in-differences. (w32054).

Wisselgren, M. J., Edvinsson, S., Berggren, M., and Larsson, M. (2014). Testing methods of record linkage on Swedish censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47(3):138–151.

Appendix

A Data

A.1 Measuring cultural assimilation

To capture individuals' own decisions to assimilate, I use names reported in the 1910 and 1930 census. Naming practices are commonly used to proxy for the vertical transmission of culture from parents to children and, consequently, for individuals' own assimilation choices (Fouka, 2022).

Each name is assigned a Cultural Name Score (CNS), calculated following (Fryer and Levitt, 2004; Gagliarducci and Tabellini, 2024; Fouka, 2020):

$$CNS_{\text{name, culture, }t} = \frac{P_t(\text{name} \mid \text{culture})}{P_t(\text{name} \mid \text{culture}) + P_t(\text{name} \mid \text{other culture})}. \tag{2}$$

The score is calculated by comparing the prevalence of names across cultural groups. The cultural groups of interest are Tornedalen, Finnish, and Swedish. The Swedish group is defined as individuals born in Sweden outside the northernmost region (Norrbotten, which includes Tornedalen) and who are not classified as Sami or Romani. The Tornedalen group consists of individuals born in Tornedalen who are not Sami or Romani, while the Finnish group includes individuals born in Finland. Each name receives a score based on its relative prevalence across these groups using equation 2.

In addition, when constructing the score, I include an "other" category, which consists of individuals born in other countries as well as members of minority groups, specifically the Sami and Romani. This category is included to reduce noise in the estimation. Moreover, following the literature, for each year t I only use names given in the 20 years prior to year t. This approach ensures that the score reflects what is considered a culturally distinct name in more recent periods.

In Sweden, children are often given multiple names. When calculating the CNS, I use all recorded given names. However, when calculating an individual's CNS, I rely only on the first given name, as the first name is assumed to carry the strongest cultural signal.

A.2 Census linking

To avoid inducing bias from the linkage process, I link individuals only on characteristics that should be stable over time. I require exact matches on year of birth, parish of birth, and sex, while allowing for some variation in names. The Swedish census—unlike, for example, the U.S. census—reports year of birth reliably, so I require an exact match on this variable. Because parish borders change over time, I harmonize parishes using a historical shapefiles available using the R package *histmaps* which is available on GitHub and written by Johan Junkka.

I also match on first and last names, but I do not require exact matches. Instead, I require names to be sufficiently similar based on Jaro–Winkler distances, as is common in the literature (Abramitzky et al., 2021). In a first matching round, I require both the first and last name/names to be unique within a Jaro–Winkler distance of 0.15, using all reported first names. However, individuals often have multiple first names, so I conduct a second round for those unmatched in the first round. In this round, I retain only the first given name and require uniqueness within a Jaro–Winkler distance of 0.10.

The main challenge with the Swedish census is that children's surnames are most often not reported. Instead, surnames must be inferred from the parents. At the time, there were two practices for how children inherited surnames. Either a child inherited the surname of the household head (usually the father, if one existed), or the child received a patronymic surname (the head's given name plus "-dotter" for females or "-son" for males). The patronymic practice was less common and was disappearing during this period.

Because these alternative practices coexisted, each child is assigned both possible surnames: one patronymic and one inherited. If the household head has more than one given name, I use the first. However, if the household head has two reported surnames, I assume that the patronymic practice is no longer used in the family and instead assign those two surnames as alternatives. If more than two surnames are reported, I use the first and the last.

In general, if anyone in the household has a missing surname, it is assumed that they share the surname(s) of the household head and I assign up to two alternatives if the head has more than one surname (the first and the last). For individuals with reported surnames, I separate multiple surnames into up to two alternatives rather than matching only on the combined form. However, I also allow matches on the combined surname or on missing surnames when relevant.

As a result, each individual can have up to three surnames used for matching: the reported surname (whether single, multiple, or missing) and up to two imputed alternative surnames. I match on missing, as an individual might stay a child or without a surname for other reasons between census rounds. I only allow a match if it is unique within the alternatives.

I have created links between every decade and when I want to link an individual back to a specific decade for an analysis, I allow for every possible connection there, but only keep unique matches. For example, when I wan to analyse outcomes in 1910 and need information from 1890, I allow the direct link between the decades, but I also allow the individual to link first from 1890 to 1900 and then to 1910 and only keep a link if it is conclusive.

My outcomes are measured using information from the 1910 and 1930 censuses combined, or from the 1930 census alone. I therefore report linkage rates to both censuses in Table A1. When possible, I combine the two censuses to increase the amount of available information and to enlarge the sample. Doing so allows me to include individuals linked to either the 1910 or the 1930 census, as well as those linked to both. The table shows high linkage rates by international standards for both genders. I also report measures of representativeness in Table A2. Overall, the samples are comparable to the underlying population.

Table A1: Backwards linkage rates

Year	Male	Female	All
1910	68.76%	67.63%	68.22%
1930	67.27%	71.83%	69.38%

Notes: The table reports backward linkage rates from the 1890 census to the 1910 and 1930 censuses for individuals aged 0–22 in 1890.

Table A2: Means in 1890 census

	Population	1910 Sample	1930 Sample
Female	0.48	0.46	0.46
Father in HH	0.92	0.93	0.93
HH head in labor force	0.80	0.81	0.80
Age	9.13	9.28	9.31
Nr. of people in HH	6.98	7.10	7.11
Finnish minority	0.49	0.48	0.48
Sami minority	0.05	0.05	0.05
Observations	19,138	10,300	8,429

Notes: The table compares characteristics in the full 1890 population (ages 0–22) to the subsets successfully linked to the 1910 and 1930 censuses.

B Additional results

Table B3: Swedish-only school exposure and their children's Tornedalian and Finnish CNS

	Female & Male		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Tornedalian CNS						
Treated (age 4-12)	1.012		1.251		1.181	
	(0.730)		(1.044)		(1.006)	
Years treated		0.208*		0.255		0.199
		(0.110)		(0.165)		(0.164)
Dep. var. mean	23.417	23.417	23.177	23.177	23.663	23.663
Observations	7485	7485	3785	3785	3700	3700
R-squared	0.204	0.205	0.244	0.244	0.243	0.243
Panel B: Finnish CNS						
Treated (age 4-12)	1.256*		-0.113		2.658**	
	(0.676)		(1.088)		(1.287)	
Years treated		0.215*		-0.068		0.495**
		(0.109)		(0.152)		(0.216)
Dep. var. mean	18.193	18.193	17.972	17.972	18.419	18.419
Observations	7485	7485	3785	3785	3700	3700
R-squared	0.085	0.085	0.111	0.111	0.120	0.121
Village-Stack FE	✓	✓	✓	✓	✓	✓
Age-Stack FE	✓	✓	✓	✓	✓	✓

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and the 1910 and/or 1930 census, and restricted to the matched village sample. *Treated* is an indicator equal to one if an individual is exposed to a Swedish-only school between ages 4 and 12. *Years treated* is a continuous variable ranging from 1 to 9, capturing the number of years an individual is exposed to the Swedish-only school system (equal to 1 if the school opens when the individual is 12 years old and 9 if it opens when the individual is 4 years old). Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B4: Swedish-only school exposure and education (extensive margin)

	Both		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Years treated	0.003	0.015***	-0.002	0.005	0.009	0.025**
	(0.003)	(0.005)	(0.005)	(0.008)	(0.005)	(0.010)
Years treated × Teacher before		-0.020***		-0.012		-0.028**
		(0.005)		(0.008)		(0.011)
Village-Stack FE	✓	✓	✓	✓	✓	✓
Age-Stack FE	✓	✓	✓	✓	✓	✓
Dep. var. mean	0.853	0.853	0.858	0.858	0.848	0.848
Observations	4,637	4,637	2,236	2,236	2,401	2,401
R-squared	0.127	0.129	0.179	0.179	0.138	0.143

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and the 1930 census, and restricted to the matched village sample. *Years treated* is a continuous variable ranging from 1 to 9, capturing the number of years an individual is exposed to the Swedish-only school system (equal to 1 if the school opens when the individual is 12 years old and 9 if it opens when the individual is 4 years old). *Teacher before* is an indicator equal to one if there was a teacher within 5 km of the village of residence in 1890. Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

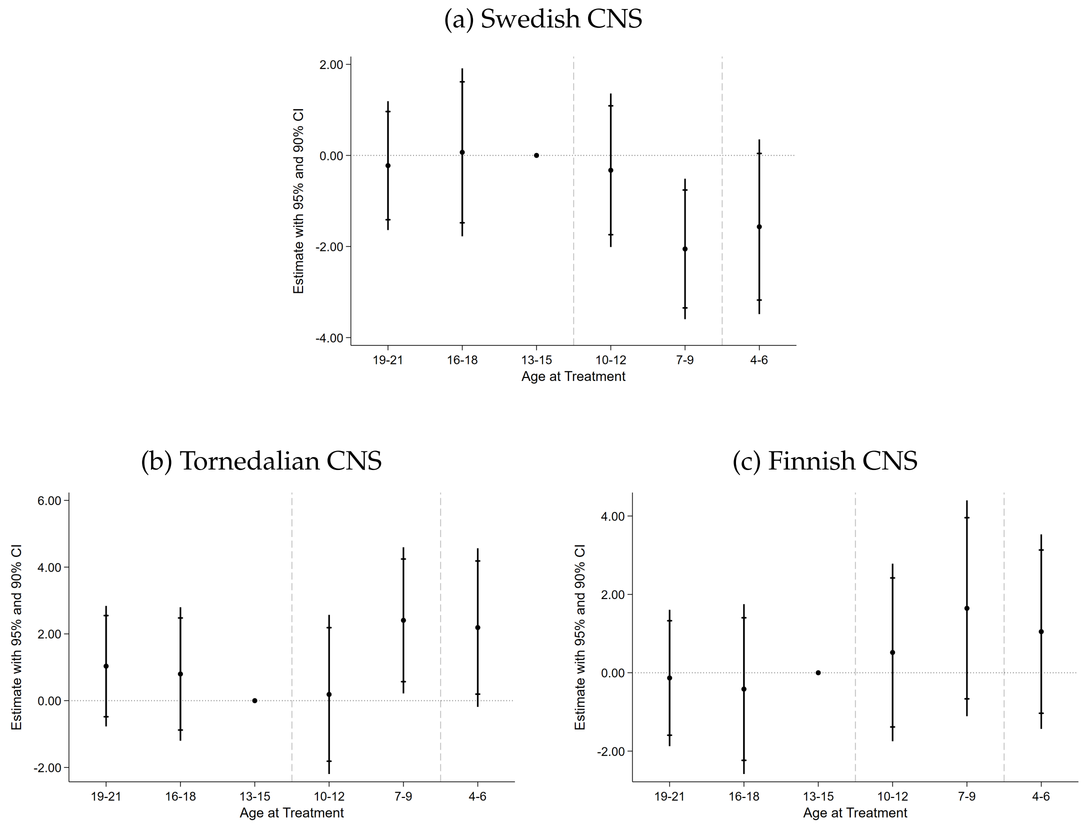


Figure B1: Swedish-only school exposure and their children's CNS

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and the 1910 and/or 1930 census and restricted to the matched village sample. The figure reports estimates from an event-study specification that interacts an indicator equal to one if an individual is born in a village experiencing a Swedish-only school opening with age-at-exposure indicators grouped in three-year bins. The omitted category is ages 13–15, corresponding to individuals just above school age at the time of treatment. The outcomes are the various Cultural Name Scores (CNS). The specification includes village and cohort stack specific fixed effects. Estimates are shown with 95% and 90% confidence intervals based on standard errors clustered at the village level.

Table B5: Swedich CNS around school opening

	Female & Male		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Years since opening	0.188**	0.392***	0.089	0.191	0.281**	0.256
	(0.086)	(0.107)	(0.173)	(0.220)	(0.113)	(0.200)
Village FE	✓	✓	✓	✓	✓	✓
Birth-year FE	✓	✓	✓	✓	✓	✓
Sibling FE		✓		✓		✓
Dep. var. mean	18.926	18.926	19.323	19.323	18.545	18.545
Observations	11102	9131	5431	3566	5671	3810
R-squared	0.117	0.390	0.122	0.488	0.133	0.459

Notes: The sample consists of individuals born between six years before and five years after the opening of a Swedish-only school, observed in the 1900 census and restricted to the matched village sample. *Years since opening* is a continuous variable ranging from 0 to 6, indicating the number of years since a Swedish-only school opened in the village. The specification includes village and cohort fixed effects. Standard errors are clustered at the village level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

(a) Female (b) Male

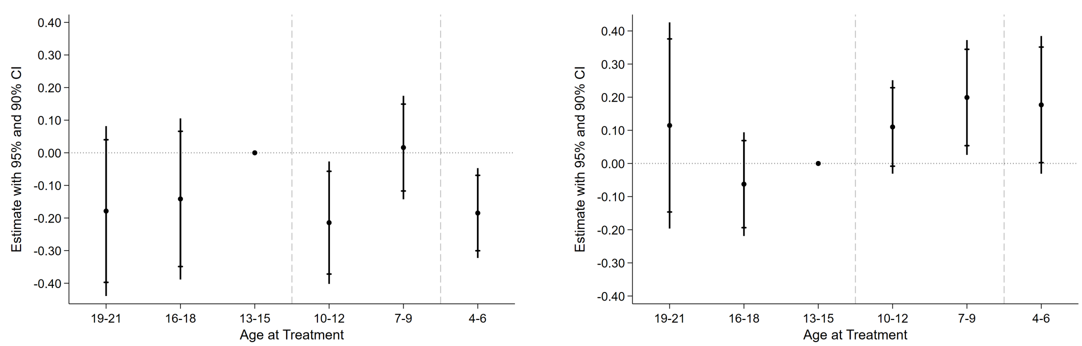


Figure B2: Education (extensive margin) in areas without pre-treatment school supply

Notes: The sample consists of individuals aged 4 to 21 at the opening of a Swedish-only school, linked between the 1890 census and 1930 census, restricted to the matched village sample, and residing in villages without any pre-treatment educational supply, defined as having no teacher present within 5 km in 1890. The figure reports estimates from an event-study specification that interacts an indicator equal to one if an individual is born in a village experiencing a Swedish-only school opening with age-at-exposure indicators grouped in three-year bins. The omitted category is ages 13–15, corresponding to individuals just above school age at the time of treatment. The outcome is any reported education in the 1930 census. The specification includes village and and cohort stack fixed effects. Estimates are shown with 95% and 90% confidence intervals based on standard errors clustered at the village level.