Impact of Residential Schools on Educational Attainment of Indigenous Women: Evidence from India

Aashay Tripathi†

October 2025

Abstract

While residential schools in North America have long been dismantled, India continues to expand its own residential school system, with a stated aim of "closing the gap" in education between Indigenous students and their peers. I provide the first causal evidence of the effect of enrollment in a residential school on the educational attainment of Indigenous women in India. Applying triple-difference and instrumental-variable strategies to a newly constructed dataset, I find that school exposure reduces educational attainment by up to 4 years. The result is driven by disruptions to family dynamics. Crowding out of day-school options by residential schools, along with mandatory residence at these schools, forces Indigenous girls to cancel enrollment, raising dropout rates due to domestic work obligations by 10 percentage points.

JEL Classification: I25, J15, O15

† Department of Economics, University of Calgary (aashay.tripathi1@ucalgary.ca). I am grateful to my advisors, Dr. Arvind Magesan, Dr. Sacha Kapoor, and Dr. Benjamin Crost, for their guidance and feedback. I also thank the seminar participants at the University of Calgary and IFMR Business School, as well as the audiences for the helpful comments and discussions at the Indigenous Economics Study Group, Economics Profession Data Committee, and the Embrace Day Graduate Poster sessions at the Canadian Economic Association 2025 annual conference.

1 Introduction

From the late 19^{th} to mid-20^{th} century, residential schools in the U.S. and Canada aimed to assimilate Indigenous populations by isolating children from their families, culture, and language. These schools exposed Indigenous children to poor nutrition, overcrowding, and cultural suppression, with many suffering abuse and inadequate healthcare (Meriam (1971), Feir (2016), Feir and Auld (2021)). The Truth and Reconciliation Commission of Canada (2015) (TRC) concluded that these schools sought to destroy Indigenous cultures, and the discovery of unmarked graves heightened calls for justice (Feir and Auld (2021), Jones (2021)).

In contrast, India's 2024 budget allocated 38,800 teachers and INR 63.99 billion ($\approx$ CAD 1.05 billion) for 740 *Eklavya Model Residential Schools* (EMRS), a 150% year-on-year increase, to improve education access for the Scheduled Tribes (STs)[1]. While North American studies have provided valuable insights into the effects of residential schools, limited evidence exists on the educational impact of operational residential schools, making India's case an important contemporary example. In this paper, I leverage a novel individual-level dataset to provide the first causal evidence of the impact of EMRS on the educational attainment of ST women in India.

The Eklavya Model Residential Schools (EMRS) were sanctioned in phases, with central and state governments determining locations and opening dates within Integrated Tribal Development Project (ITDP) areas[2]. These schools, which follow state or central education board curricula, admit students through competitive examinations, with provisions in place for tribal and first-generation students. Admitted students reside on school premises for the academic year, with all the related costs covered by the government.

[1]Scheduled Tribes (STs), often referred to as *Adivasis*, meaning Indigenous peoples or original inhabitants, are groups recognized and protected by the Indian Constitution. They predominantly reside in remote and under-served areas and receive special protections and benefits to address their historical, social and economic marginalization.

[2]These are contiguous administrative units where the ST population is at least 50 percent of the total.

To estimate causal effects of schooling, I first construct a novel deterministic link from EMRS to NFHS clusters and to village-level forest cover. First, I geo-code every EMRS to exact latitude–longitude coordinates. Second, I compute Haversine distances from each school to each of the 36,000 NFHS village clusters and assign each cluster to its unique nearest school, producing a complete school–cluster map. Third, I attach the forest cover by nearest-neighbor matches in two steps: (i) match each of 500,000 villages with forest cover to its nearest school, and (ii) match villages to their nearest NFHS cluster, so each cluster inherits local forest cover. This construction delivers respondent-level exposure to the nearest EMRS and a high-resolution forest measure, providing the measurement base for the causal analyses.

I implement a triple-differences design exploiting variation in (i) ST status, (ii) village EMRS presence, and (iii) cohort eligibility - whether an individual was of school-going age at school opening. The first difference contrasts ST and non-ST populations; the second contrasts villages with and without a school; the third contrasts age-eligible and ineligible cohorts. Their interaction absorbs pre-existing and potentially time-varying ST–non-ST gaps that can bias standard difference-in-differences estimates (reported in the Appendix). Identification requires that, absent EMRS, the change in the ST–non-ST gap at the age-eligibility cutoff should be the same in villages with and without a school. That is, the gaps would evolve in parallel across cohorts. The fact that, after opening, only ST in school villages drop while the other series continue their trends identifies the EMRS effect.[3]

The main results show that ST women who were of school-going age when a residential school became operational in their village completed about 1.25 fewer years of education relative to non-ST and older peers. Consistent with this finding, I find no detectable impact on the probability of completing secondary education. These results suggest that exposure

[3]Equivalently: the difference-in-differences between eligible and ineligible STs and non-STs would have trended similarly in the absence of schools.

to EMRS reduced overall educational attainment among ST women, implying that the loss concentrates in total years of schooling rather than terminal attainment.

I probe robustness with an instrumental variables strategy. As schools were predominantly sanctioned in areas with higher ST populations, and STs often reside in remote areas which tend to have higher forest cover, I use cross-sectional variation in Vegetation Continuous Fields (VCF)[4] as an exogenous variation for school presence. Interacting VCF with cohort eligibility yields a strong instrument: women living in high-VCF villages *and* who were age-eligible at the time of school opening in their village were more likely to be exposed to EMRS.

The IV estimates indicate a four-year decline in schooling for exposed ST women, with average education in affected villages also falling by roughly four years. For context, eligible ST women in EMRS villages have an average of 7.364 years of education; a four-year loss represents more than half of that total and effectively doubles the gap relative to eligible non-ST women, who average 8.907 years. These estimates rely on the exclusion restriction that the instrument affects education only through EMRS exposure. Results are stable across specifications with district fixed effects (for time-invariant district characteristics) and state-year fixed effects (for state-specific trends). The spatial and temporal variation introduced by forest cover and cohort eligibility remains intact, ensuring the instrument isolates the impact of residential school exposure. Placebo tests rule out alternative channels, reinforcing the validity of the instrumental variable results.

Finally, I provide evidence on the mechanism. Because EMRS requires students to reside on campus for the full academic year, families lose access to their children's labor for domestic and agricultural work. This disruption of family dynamics and the crowding out of day-schooling options by the residential schools raises the opportunity cost of sending children

[4] A MODIS product, VCF measures tree cover at 250m resolution from 2000 to 2019 using a machine-learning model trained on human-labeled data.

to EMRS. In response, ST girls cancel enrollment to meet domestic obligations, leading to a 10-percentage-point rise in dropout attributable to domestic work. While EMRS aims to expand formal education, these findings highlight how policies that overlook socioeconomic realities and cultural norms can unintentionally harm the communities they intend to support.

Historically, residential and boarding schools have been viewed as instruments of assimilation, with critics emphasizing their role in "cultural genocide" through the systematic erosion of language, kinship, and cultural practice (Truth and Reconciliation Commission of Canada (2015)). The historical record documents harsh material conditions and coercive separation from families, and a large body of work links these features to long-run cultural and socioeconomic harm (Adams (1995); Milloy (2017); Bombay, Matheson and Anisman (2014); Bougie and Senécal (2010)). At the same time, a competing narrative stresses selective gains for a small subset, with evidence for higher graduation, English acquisition, higher per capita income, lower poverty rates, smaller family sizes in the present day, and later political mobilization by an educated elite (Reyhner and Eder (2017); Szasz (2006); Glenn (2011); Gregg (2018)). Empirically, Feir (2016) reconciles the two historical positions: residential schooling raised high school graduation and employment probabilities but weakened cultural attachment, such as speaking an Aboriginal language at home and participating in traditional activities. Extending the horizon, Jones (2021) documents that residential school exposure is associated with lower educational attainment among descendants, challenging the canonical expectation that parents' education monotonically lifts children's attainment (Black, Devereux and Salvanes (2005); Oreopoulos, Page and Stevens (2006)).

My paper contributes directly to this debate by being the first to study a contemporary, state-run residential model in India, where entry is voluntary and exam-based rather than coerced. I show that, even without coercion and with full cost coverage, these schools can reduce completed schooling for ST girls. My work moves from retrospective legacies to *contemporaneous* effects in a fast-expanding system: I estimate immediate, causal impacts of today's EMRS on schooling outcomes and show that the mechanism runs through household dynamics and time constraints created by residential attendance.

This paper also connects to the literature on education finance and student aid, which shows that lowering direct costs generally raises attainment (Dynarski (2004); Deming and Dynarski (2009); Dynarski and Scott-Clayton (2013)). In related work on Indigenous students in the U.S., Jones (2023) finds that funding cuts reduce completion rates, especially on reserves with limited local access to post-secondary options. In the EMRS context, all direct costs of tuition, books, uniforms, food, and boarding are covered, and ST and non-ST households live in the same villages, removing classic reserve-based geographic barriers. My results nevertheless show no gain in secondary completion and a decline in years of schooling for ST girls. The contribution to this strand is to isolate a non-price margin. When the binding constraints are time and intra-household labor due to socioeconomic marginalization, a residential model can *raise* the opportunity cost of attendance and lower attainment despite generous subsidies.

A third strand studies state-backed missions and cultural change. Historical evidence from Catholic missions in Latin America and Christian missions in Africa shows that mission exposure often increased human capital while promoting assimilationist norms (Valencia Caicedo (2019); Jedwab, Meier zu Selhausen and Moradi (2022)). EMRS is a modern, secular, state-run analog with similar human-capital goals but applied to India's STs. My work shows that design features matter. By requiring residence and weakening day-school alternatives, EMRS risk pushing tribal communities into a standardized socio-cultural model and reducing schooling for the intended beneficiaries in the short run, highlighting the trade-off between standardized delivery and local socio-cultural and household realities.

Finally, the study relates to work on enrollment and attainment in developing countries that emphasizes pedagogy/remediation (Banerjee et al. (2007)), girl-friendly inputs (Kazianga et al. (2013)), demand-side incentives (Barrera-Osorio et al. (2011)), infrastructure expansions

(Duflo (2001); Breierova and Duflo (2004)), and scholarships (Kremer, Miguel and Thornton (2009)). Most of this literature evaluates day-school settings, as well as price and quality margins. My contribution is to identify a distinct, under-studied policy margin: a government-funded *residential* model for an Indigenous minority, and to measure its immediate, causal effects using a deterministic micro–spatial linkage of schools, NFHS clusters, and village-level forest cover.

2 Context and Data

2.1. Context

2.1.1. Scheduled Tribes

India hosts the world's second-largest Indigenous population: Scheduled Tribes (STs) constitute 8.6% of the population, over 104 million people across 705 notified groups per the 2011 Census (Census of India and Office of the Registrar General & Census Commissioner (2013); Subramanian et al. (2023)). Constitutionally, STs are communities specified by Presidential notification for each State/UT (Article 342). Article 46 directs the State to promote its educational and economic interests and protect them from social injustice and exploitation (Constitution of India (Commentary) (2024*a*,*b*)).

Despite these safeguards, STs remain among India's most socio-economically disadvantaged groups. Census benchmarks and official releases record a persistent literacy gap of 59% for STs versus 73% overall in 2011 (Press Information Bureau, Government of India (2020)). Spatially, ST populations are concentrated in a central–eastern belt (Madhya Pradesh, Chhattisgarh, Odisha, Jharkhand, Maharashtra, parts of Gujarat and Rajasthan) and across several North-Eastern states, often in remote, forested districts where access costs are high and services are thin (Census of India and Office of the Registrar General & Census Commissioner (2013)) The High Level Committee on Tribals documents structural sources of disadvantage, historical dispossession, displacement linked to extractive and infrastructure projects, weak local state capacity, and program design that overlooks language and cultural context—directly affecting schooling access and progression (Xaxa et al. (2014)). In education, distance to schools, instruction in non-mother-tongue languages, and the opportunity cost of children's domestic and agricultural labor remain salient barriers alongside poverty.

2.1.2. Eklavya Model Residential Schools (EMRS)

The Eklavya Model Residential School (EMRS) program, launched by the Ministry of Tribal Affairs (MoTA), aims to provide quality education to Scheduled Tribe (ST) children in remote areas. Sanctioned under Article 275(1) of the Constitution, it is funded by the MoTA. By August 2024, 728 schools were sanctioned, with 409 functional. The program gained momentum with the 2010 target to establish an EMRS in every Integrated Tribal Development Project (ITDP) area and a 2018 directive expanding it to all sub-districts. In 2018–20 the scheme was formally revamped: one EMRS per block with at least 50% ST population and 20,000 ST persons; minimum land requirement reduced to 15 acres; and each school designed for a sanctioned strength of 480 students from classes VI–XII (Press Information Bureau, Government of India (2025*a*); Government of India (2010)). As of 2022, 462 new schools were sanctioned in 564 sub-districts identified by the 2011 Census.

EMRS schools may affiliate with State or Central Boards and are managed by a society comprising local, State, and Central representatives. The National Education Society for Tribal Students (NESTS) was set up as an autonomous body under MoTA to establish, fund, and oversee EMRS operations nationwide (Ministry of Tribal Affairs (2025)). Some states use public–private partnership arrangements involving NGOs. Admission is competitive, with provisions for particularly vulnerable tribal groups and first-generation learners, and equal seats for girls and boys. The curriculum covers English, Hindi, and the student's mother tongue, with space for tribal culture, tradition, and history. All expenses, including tuition, books, uniforms, food, and boarding, are covered. The approved recurring grant is up to INR

1.09 lakh (≈ C$1700) per student per year for residential schools, which finances student expenses and school running costs (Press Information Bureau, Government of India (2023)).

However, the emphasis on parity with non-ST populations may overlook local vocational interests and socio-cultural ties. Limited per-student funding relative to actual residential costs and reserved seats for non-tribal students can dilute targeting, and the absence of formal mechanisms to sustain family and community contact risks weakening children's socio-cultural environment. Recent official updates note continuing expansion and operationalization under NESTS, but also show gaps between sanctioned and functional schools—suggesting capacity, staffing, and infrastructure remain binding constraints (Press Information Bureau, Government of India (2025*b*)).

2.2. Data

I assemble a new micro–spatial dataset that deterministically links EMRS, NFHS village clusters, and village-level forest cover. The construction proceeds in three stages.

(1) EMRS and geo-coding: I compile the universe of EMRS sanctioned across India through 2024, including complete postal addresses (state, district, block, village) and opening year. I geo-code each school to obtain exact latitude–longitude coordinates and retain one record per school. These coordinates anchor all school–location matches used below.

(2) NFHS and masked cluster coordinates: I use the National Family Health Survey (NFHS) to obtain individual-level outcomes and covariates across all states and union territories. Outcomes include years of education and the highest grade attained; covariates include income status and standard health indicators (malnutrition, anemia, hypertension, HIV, high blood glucose). NFHS locates respondents in GPS-based cluster centroids that are randomly displaced to protect anonymity: in rural areas, up to 5 km for all clusters and up to 10 km for 1% of clusters, with all points constrained to the original country, district, and survey region. I use these masked coordinates to link clusters to schools and to village-level forest cover.

(3) Deterministic spatial linkage: Using school and cluster coordinates, I calculate Haversine distances[5] between every EMRS and every NFHS cluster. Each cluster is assigned to its *unique nearest school*, producing a complete school–cluster map. I then incorporate the village-level forest cover. I geo-code ≈565,000 village names to coordinates and compute Haversine distances to assign each village to its nearest NFHS cluster, merging the forest cover, measured as tree cover at a 250-meter resolution, at the village level to the NFHS cluster using the cluster ID. Figures 1 and 2 document the spatial distributions of sanctioned EMRS, NFHS cluster ST shares, and forest cover. EMRS are predominantly located in regions with higher ST concentrations, and ST populations are concentrated in more forested areas.

Integrated analysis file: Finally, I merge the EMRS–NFHS–Forest Cover file with NFHS covariates using cluster IDs. The resulting analysis dataset contains, for each respondent: years of education and schooling attainment; income and health covariates; the identity and distance of the nearest EMRS; the school's opening year; the cluster-level forest cover; and the ST share of the cluster. To my knowledge, this is the first dataset to link individual outcomes to operational residential schools and high-resolution environmental measures with national coverage, connecting 36,000 NFHS clusters and 500,000 villages to the universe of EMRS through 2024. The deterministic school–cluster map defines exposure at the relevant micro scale, the village-to-cluster mapping delivers a consistent forest cover measure aligned with NFHS geography, and the combined file supports both the triple-difference design (ST status × EMRS presence × cohort eligibility) and the IV strategy that instruments exposure using VCF interacted with eligibility. The construction is reproducible and relies only on observed coordinates and minimum-distance rules, eliminating discretionary matching. Table I reports summary statistics for all variables used in the main analysis.

[5] $D(x, y) = 2 \arcsin \sqrt{\sin^2((x_1 - y_1)/2) + \cos(x_1)\cos(y_1)\sin^2((x_2 - y_2)/2)}$.

EMRS–NFHS–Forest Linkage Workflow

Sources (Admin & Survey)

EMRS Admin: full postal addresses (state, district, block, village) and opening year.

NFHS: individual outcomes (years of education, highest grade) and covariates (income status; malnutrition, anemia, hypertension, HIV, high blood glucose); GPS cluster points with standard rural displacement (up to 5 km; 1% up to 10 km; within district/region).

⇓

Geospatial

Village Forest Cover (VCF): percent tree cover at 250 m resolution for ∼565,000 villages.

Village Gazetteer: official village names and standardized spellings.

⇓

Preprocess

Geo-code EMRS: obtain exact latitude–longitude; one record per school.

Geo-code Villages: standardize names and find village coordinates.

⇓

Deterministic Linkage (Nearest-Neighbor)

(1) School → Cluster (Exposure): compute Haversine distances from each EMRS to every NFHS cluster; assign each cluster to its *unique nearest school* ⇒ complete school–cluster map.

(2) Village → Cluster (Forest Cover): compute Haversine distances from each village to NFHS clusters; assign each village to its *nearest cluster* to have village VCF for each cluster.

These links attach high-resolution forest cover consistent with NFHS geography.

⇓

Integrated Analysis File & Variables

For each respondent: years of education and attainment; income and health covariates; identity and distance of the nearest EMRS; school opening year; forest cover; and the cluster's ST share.

⇓

Construct Treatment & Instrument

Treatment (EMRS presence): cluster is treated if its unique nearest school is an EMRS.

Eligibility (cohort timing): equals 1 if the respondent was school-age when the nearest EMRS opened.

Instrument: $VCF \times Eligibility$, where VCF is the cluster-level forest cover built above.

⇓

Ready for Estimation

Dataset supports the triple-difference design (ST status $\times$ EMRS presence $\times$ cohort eligibility) and IV using $VCF \times Eligibility$.

Table 1: Summary Statistics

Villages with an EMRS						
Variable	**Eligible Cohort (ST)**		**Ineligible Cohort (ST)**		**Eligible Cohort (Non-ST)**	
	Mean	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Years of Education	7.364	4.617	3.551	4.551	8.907	4.434
Age	22.591	4.592	39.242	5.366	22.730	4.550
Hindu	0.6765	0.4678	0.6761	0.4680	0.8557	0.3514
Married	0.5364	0.4987	0.8681	0.3384	0.5830	0.4931
Male head of HH	0.8355	0.3707	0.8385	0.3679	0.8580	0.3491
Wealth Index	2.234	1.284	2.253	1.306	3.011	1.399

Villages with no EMRS						
Variable	**Eligible Cohort (ST)**		**Ineligible Cohort (ST)**		**Eligible Cohort (Non-ST)**	
	Mean	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Years of Education	7.747	4.505	3.999	4.647	9.184	4.430
Age	22.531	4.667	39.510	5.395	22.396	4.634
Hindu	0.6750	0.4683	0.6555	0.4751	0.8008	0.3993
Married	0.5383	0.4985	0.8852	0.3186	0.5524	0.4972
Male head of HH	0.8447	0.3621	0.8378	0.3685	0.8477	0.3592
Wealth Index	2.429	1.327	2.420	1.335	3.144	1.372

Note: Forest cover is the mean percentage of tree cover detected in the polygon. The wealth index spans from 0=poorest to 5=richest. The variables Hindu, Married, and Male head of HH are indicators that equal 1 if the individual is a Hindu, is married, and the head of that HH is a male. An individual belongs to the eligible cohort if they were of the school-going age when the school was operational in their village.

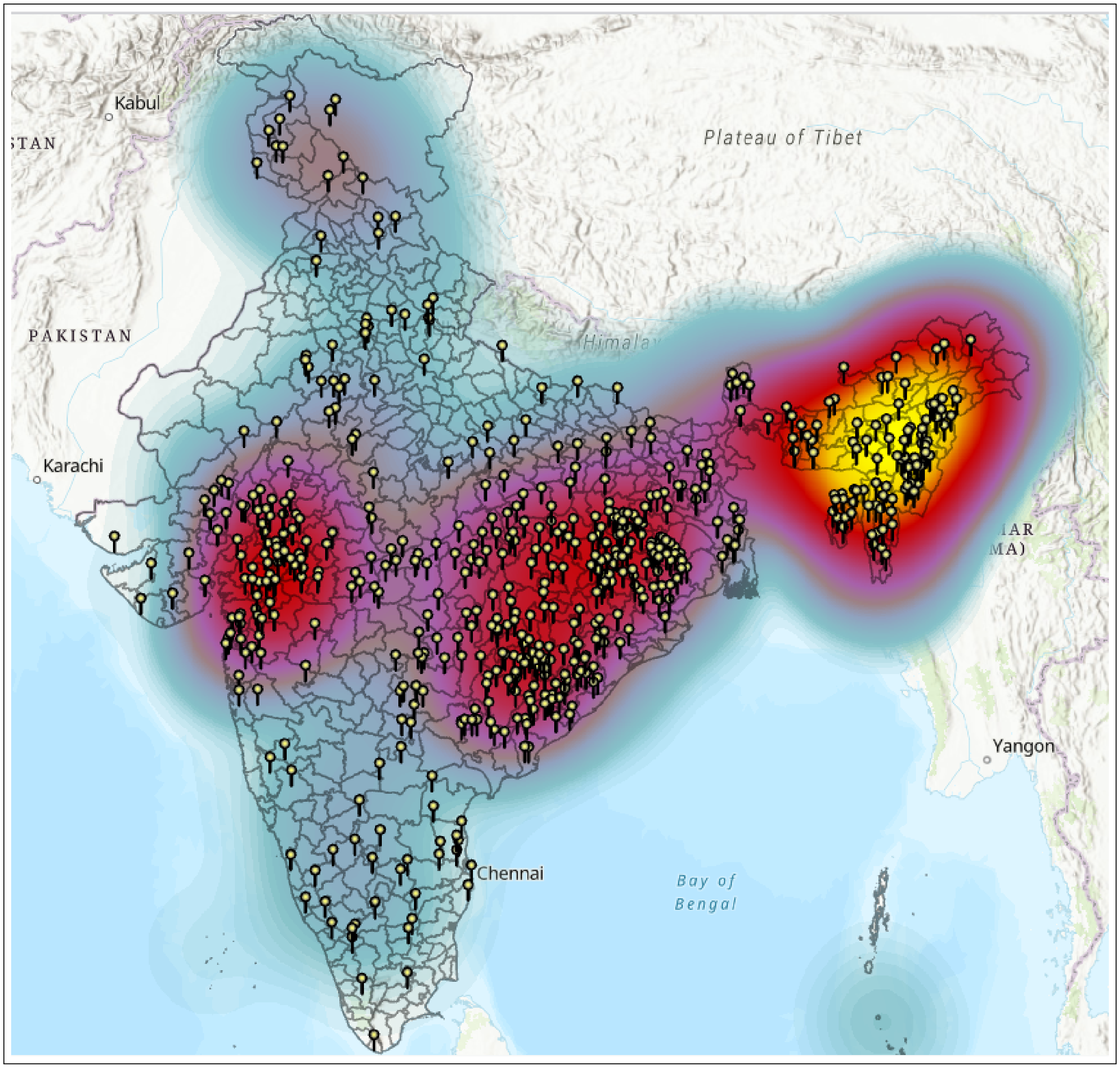

Figure 1: Overlap of EMRS on the percentage of ST population in the NFHS village clusters.

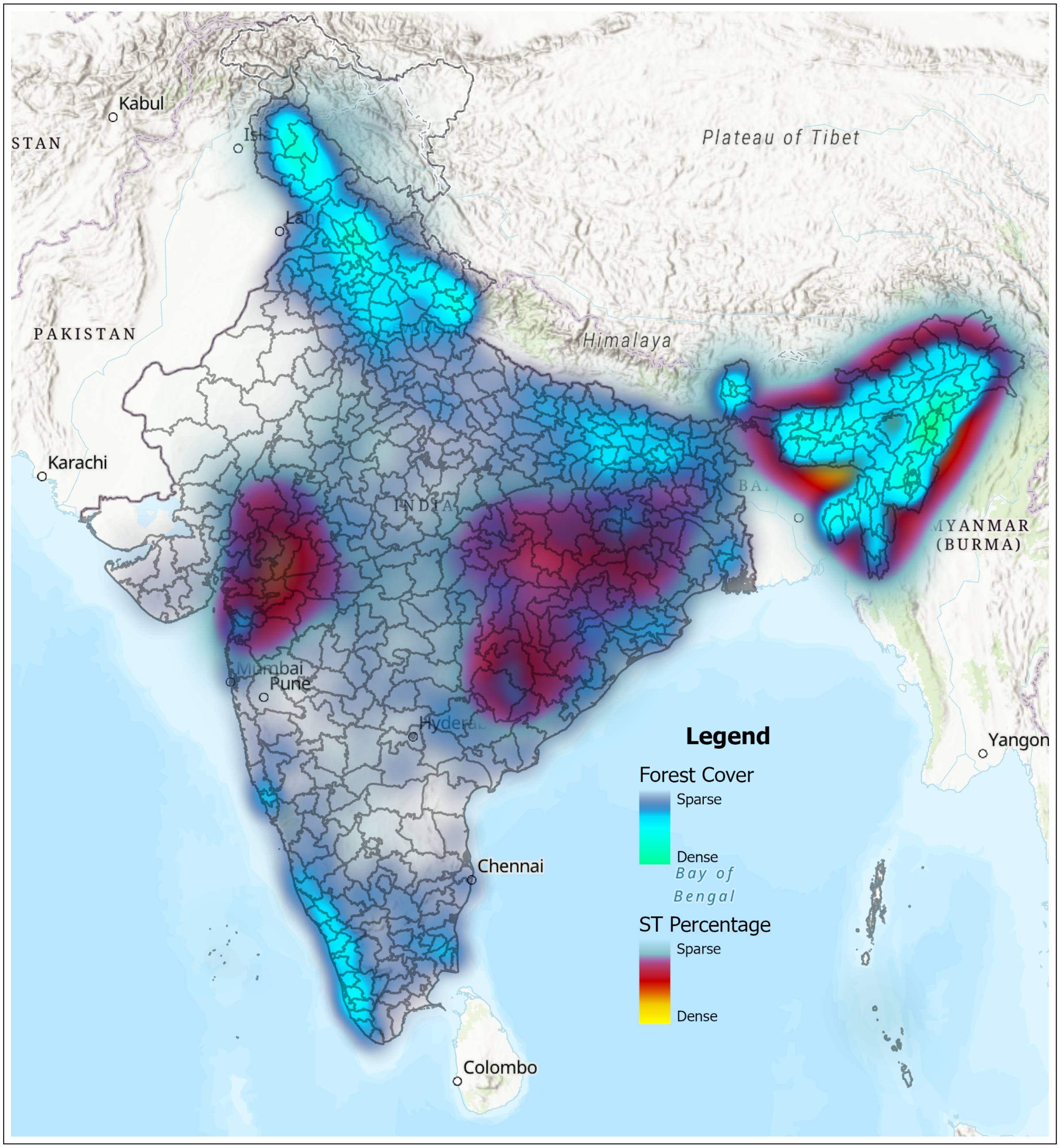

Figure 2: Overlap of Forest Cover on the percentage of ST population in the NFHS village clusters.

3 Identification I: Triple Difference

The empirical strategy exploits the phased expansion of government residential schools for Scheduled Tribes (ST) induced by the 2010 policy that targeted every Integrated Tribal Development Agency (ITDA) with an ST share of at least 50 percent. A natural approach is a difference-in-differences design contrasting villages with and without a school and comparing ST to non-ST residents. Such estimates require that unobserved determinants of schooling evolve similarly across treated and comparison groups. But, simple cross-sectional contrasts are uninformative because villages and groups differ systematically in geography, infrastructure, and access to programs. Furthermore, ITDAs span multiple blocks and villages. Although the selection of ITDAs may follow specific criteria, the choice of specific village sites and the timing of school opening within an ITDA are not known. Given these persistent ST–non-ST differences and because placement prioritized high-ST ITDA areas, with within-area opening dates following administrative readiness (land, construction, staffing, budgets), school presence and timing correlate with pre-existing trajectories. Given the possible violation of assumptions, I use a triple-difference design that adds within-village cohort eligibility as a third source of variation in exposure timing[6].

The triple-difference design uses the deterministic mapping from each NFHS cluster (village v) to its nearest EMRS and the year it became operational. Let $School_v \in \{0, 1\}$ indicate that the mapped school is operating; let $ST_i \in \{0, 1\}$ denote Scheduled Tribe status; and let $Eligible_{g(i)} \in \{0, 1\}$ indicate that individual i was of school-going age when the mapped EMRS opened. Eligibility varies by age within village, holding location fixed. Interacting these three indicators compares the ST–non-ST gap among age-eligible cohorts in school villages to the corresponding gap in non-school villages, and differences out the analogous comparison for age-ineligible cohorts.

[6] I report two-way difference-in-differences estimates for completeness in the Appendix.

Formally, for individual i in village v and cohort $g(i)$, the estimating equation is

$$\begin{aligned} Years\ Education_{iv} &= \beta_0 + \beta_1\ ST_i + \beta_2\ Eligible_{g(i)} + \beta_3\ School_v + \beta_4\ ST_i \times Eligible_{g(i)} \\ &+ \beta_5\ Eligible_{g(i)} \times School_v + \beta_6\ School_v \times ST_i \\ &+ \beta_7\ ST_i \times Eligible_{g(i)} \times School_v + \gamma_{s(v)t(g(i))} + \theta_{d(v)} + \epsilon_{iv} \end{aligned} \tag{1}$$

where $\gamma_{s(v)t(g(i))}$ are state-by-year fixed effects absorbing state-specific shocks and policies, and $\theta_{d(v)}$ are district fixed effects absorbing time-invariant district characteristics. The coefficient of interest is the triple interaction β_7. To make the estimand explicit, define the cell mean:

$$\mu(st, s, e) \ \equiv\ \mathbb{E}\Big[Years\ Education_{iv} \Big| ST_i = st,\ School_v = s,\ Eligible_{g(i)} = e\Big], \qquad st, s, e \in \{0, 1\}.$$

Then the triple-difference targeted by (1) is

$$\beta_7 + \Big\{[\mu(1,1,1) - \mu(0,1,1)] - [\mu(1,0,1) - \mu(0,0,1)]\Big\} - \Big\{[\mu(1,1,0) - \mu(0,1,0)] - [\mu(1,0,0) - \mu(0,0,0)]\Big\}. \tag{2}$$

Equation (2) is the difference in the ST–non-ST gaps for *eligible* cohorts between school and non-school villages, net of the same difference for *ineligible* cohorts. It is therefore the incremental change in the ST–non-ST difference for age-eligible cohorts attributable to school presence, relative to the same change for age-ineligible cohorts.

Identification requires a triple–difference parallel–trends condition. This means the change in the ST–non-ST gap at the eligibility cutoff would remain similar across school and non-school villages.

$$\Big\{[\mu_0(1,1,1) - \mu_0(0,1,1)] - [\mu_0(1,0,1) - \mu_0(0,0,1)]\Big\} = \Big\{[\mu_0(1,1,0) - \mu_0(0,1,0)] - [\mu_0(1,0,0) - \mu_0(0,0,0)]\Big\}. \tag{3}$$

Under (3), the triple-interaction coefficient β_7 from equation (1) recovers the causal effect of EMRS exposure for ST individuals who were school-age at opening. This is credible for four reasons.

First, eligibility varies by place due to biology, not choice. $Eligible_{g(i)}$ is determined mechanically by age at the locally mapped opening date. Birth timing cannot be adjusted to an opening year set years later by land availability, construction progress, staffing approvals, or budget releases. Hence, conditional on village, the eligible and ineligible cohorts are not endogenously aligned with EMRS timing.

Second, school placement and timing are place–level and not cohort–specific. Post-2010 targeting prioritized high-ST ITDA areas, and within those areas, village siting and opening years followed administrative readiness. These forces correlate with village traits (remoteness, infrastructure) but do not select a particular age band. Any time-varying village shocks that differentially affect ST and non-ST residents (e.g., roads, parallel programs) load similarly on the cohorts and therefore difference out in the triple difference estimand (2).

Third, broad ST–non-ST dynamics are absorbed and differenced. State-by-year fixed effects $\gamma_{s(v)\times t(g(i))}$ net out state-specific policies and macro shocks, and district fixed effects $\theta_{d(v)}$ absorb time-invariant geography. What remains are smooth cohort trends within a village. Unless there is a discrete, cohort-specific shock that coincides exactly with the eligibility boundary only in school villages, the counterfactual cell means $\mu_0(st, s, e)$ satisfy (3).

Fourth, alternative threats align poorly with the eligibility cutoff. Large-scale anticipatory migration of ST households targeted to a single cohort is implausible in rural settings, and other education programs typically operate at the school or village level (affecting all ages), not exclusively the just-eligible cohort. Because EMRS provides boarding rather than a new day-school option, there is little reason to expect a sharp improvement for exactly the eligible cohort absent EMRS opening.

4 Results I: Difference-in-Difference

I begin by reporting difference-in-differences estimates to benchmark magnitudes and to illustrate why a two–way design is fragile in this setting; the corresponding tables are relegated to the Appendix. Table A1(a) compares ST and non–ST women of school-going age across villages with and without an EMRS. The interaction $ST \times School$ in column (6) is -1.199, implying roughly 1.2 fewer years of completed schooling for eligible ST women relative to eligible non–ST women in villages with a school. Table A1(b) turns to secondary completion, resulting in an 8 percentage–point decline for the eligible ST cohort. Tables A2(a) and A2(b) restrict to villages with a school and implement a DiD across eligibility within those villages. The interaction $ST \times Eligible$ equals -1.424 years in A2(a), with a -7.2 percentage–point (insignificant) effect on completion in A2(b). These patterns are consistent with the adverse impacts of exposure. They are not causal: the two setups mix (i) non-random placement into high-ST, remote, low-infrastructure areas, (ii) ST–non-ST trends that change over time for reasons unrelated to EMRS, and (iii) cohort composition differences when eligibility is not used as a within-village contrast. For these reasons, the Appendix tables A1(a), A1(b), A2(a), and A2(b) are presented as descriptive benchmarks rather than preferred estimates.

The main estimates come from the triple–difference design in equation (1). Table 1 reports effects on years of education. The coefficient on the triple interaction $ST \times School \times Eligible$ is -1.248, indicating that ST women who were of school-going age when the mapped EMRS became operational completed, on average, 1.25 fewer years of education than their non-ST and non-eligible counterparts in villages without the school. Using the mean years of schooling of 7.364 for eligible ST women in EMRS villages, this corresponds to a decline of about 17 percent. State–year and district fixed effects are included throughout, so the estimate is identified off within–village cohort contrasts while absorbing state–specific time paths and time–invariant district characteristics.

Table 1: Education: Triple Difference

Dependent Variable:			**Years of Education**			
	(1)	(2)	(3)	(4)	(5)	(6)
ST × School × Eligible	-0.0950	-0.1462	-0.3913	-0.6477	-0.9559*	**-1.248****
	(0.6006)	(0.5816)	(0.5493)	(0.5643)	(0.5367)	(0.5409)
Observations	8,310	8,310	8,305	8,305	7,042	7,042
R-squared	0.1239	0.1994	0.2488	0.3287	0.4110	0.4658
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. Estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 turns to the extensive margin of secondary school attainment. The triple interaction indicates a -3.4 percentage–point change in the probability of completing secondary school; the estimate is statistically indistinguishable from zero. The contrast between Tables 1 and 2 makes clear that the policy primarily reduces total years of schooling, an intensive-margin loss, rather than shifting the probability of crossing the secondary threshold.

Taken together, the Appendix difference-in-difference results and the preferred triple difference estimates deliver a coherent picture. The difference-in-difference specifications show sizable negative associations but are vulnerable to bias from targeted placement and differential trends by group; the triple difference corrects these concerns by conditioning on within–village eligibility timing and comparing ST–non–ST gaps across adjacent cohorts, yielding a causal effect of roughly -1.25 years. That magnitude is significant in both levels and percentage terms, and it is obtained despite full cost coverage at EMRS.

The magnitude is policy-relevant. EMRS was intended to expand schooling for a marginalized population, yet the preferred estimate implies a large intensive-margin loss for the

Table 2: Completing Secondary Education: Triple Difference

Dependent Variable:	**Completing Secondary Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
$ST \times Eligible \times School$	0.0320	0.0317	0.0117	0.0059	-0.0230	-0.0341
	(0.0458)	(0.0448)	(0.0403)	(0.0429)	(0.0399)	(0.0424)
Observations	8,310	8,310	8,305	8,305	7,042	7,042
R-squared	0.0404	0.0573	0.1058	0.1608	0.1800	0.2144
Year FE	-	✓	-	-	-	-
State FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is an indicator which equals 1 if the individual completes secondary education. Estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

targeted group. The result is robust across specifications that include state–year fixed effects and district fixed effects, so it is identified from within-village contrasts across adjacent cohorts while absorbing state-specific shocks and time-invariant district characteristics. Contrary to the program's objective, EMRS exposure is associated with diminished attainment for eligible ST women. This pattern echoes evidence from North America: long-run studies find that residential school attendance increased formal credentials for some but ultimately depressed Indigenous educational attainment across generations (Feir (2016); Jones (2021)). Taken together, the Indian and North American evidence point to a recurring mechanism: residential models that remove students from family and community, and impose a standardized design, can be misaligned with local constraints and norms, producing significant losses in total years of schooling for the Indigenous populations even when direct costs are entirely covered.

5 Identification II: Instrumental Variable

The reduced–two forces can confound form relationship between residential schools and educational attainment: (i) *endogenous placement and timing:* EMRS were prioritized to high–ST, remote, forested areas and opened subject to land, construction, staffing, and budget cycles; and (ii) *selection on unobservables:* parental preferences, community attitudes toward schooling, or ability (relevant given entrance tests) may correlate with both exposure and outcomes. In the causal equation

$$YearsEducation_{iv} \;=\; \gamma_0 \;+\; \gamma_1\, D_v \;+\; \gamma_2 X_{iv} \;+\; \varepsilon_{iv},$$

γ_1 is therefore not identified under $[Y_0, Y_1 \perp D \mid X]$ or $\mathbb{E}[\varepsilon \mid X, D] = 0$, where Y_0, Y_1 are the potential outcomes, D is the treatment indicator, X are the observable characteristics, and ϵ are the unobservables.

To address these concerns, I use an *eligibility–by–environment* instrument that shifts exposure mechanically at the cohort margin only in places where schools were more likely to be sited. Specifically, define

$$Z_{ivt} \;=\; ForestCover_v \times Post_{it},$$

where $Post_{it} = 1$ if ST girl i was of school-going age when the mapped EMRS became operational, and $ForestCover_v = 1$ if village v's forest cover exceeds the average. The first factor, $ForestCover_v$, captures *cross-sectional siting propensity.* By design, EMRS were rolled out to ITDA/ITDP areas with high ST concentration, and ST populations in India are disproportionately located in remote, hilly, and heavily forested terrain. Consistent with this targeting rule, I show that villages with higher vegetation cover (VCF, 250m) are systematically more likely to host or be mapped to an operational EMRS. Thus, $ForestCover_v$ proxies the ex ante probability that a village lies on the EMRS siting margin generated by the 2010 policy and subsequent administrative implementation. The second factor, $Post_{it}$, converts that siting propensity into actual exposure. $Post_{it} = 1$ for individuals who were of school-going age when the mapped EMRS in their village became operational, and 0 otherwise. Because $Post_{it}$ varies within the village across cohorts and is mechanically determined by age at opening, it does not itself shift where schools are built; it only determines who could feasibly attend once a school exists. Accordingly, the interaction $ForestCover_v \times Post_{it}$ raises the probability of exposure *only* when (i) the individual resides in a location with high ex ante siting propensity *and* (ii) she belongs to the cohort that was age-eligible at the time of opening. This within-village, cohort-by-cohort variation strengthens the first stage, while the separate terms for forest cover and cohort absorb overall level differences and smooth age trends.

I estimate the following IV system:

$$\begin{array}{llcl}
\text{Causal:} & YearsEducation_{ivt} & = & \delta_0 \;+\; \delta_1\, SchoolExposure_{ivt} \;+\; u_{ivt}, \\
\text{First stage:} & SchoolExposure_{ivt} & = & \gamma_0 \;+\; \gamma_1\, Post_{it} \;+\; \gamma_2\, ForestCover_v \;+\; \gamma_3\, Z_{ivt} \;+\; v_{ivt}, \\
\text{Second stage:} & YearsEducation_{ivt} & = & \delta_0 \;+\; \delta_1\, \widehat{SchoolExposure}_{ivt} \;+\; u_{ivt}, \\
\text{Reduced form:} & YearsEducation_{ivt} & = & \beta_0 \;+\; \beta_1\, Post_{it} \;+\; \beta_2\, ForestCover_v \;+\; \beta_3\, Z_{ivt} \;+\; e_{ivt}.
\end{array}$$

The parameter of interest is δ_1, obtained as the Wald ratio of the reduced form to the first stage:

$$\widehat{\delta}_1 \;=\; \frac{\widehat{\beta}_3}{\widehat{\gamma}_3} \;=\; \frac{\mathbb{E}[\,Y_{iv1} - Y_{iv0} \mid Z_{ivt} = 1] - \mathbb{E}[\,Y_{iv1} - Y_{iv0} \mid Z_{ivt} = 0]}{\mathbb{E}[\,S_{iv1} - S_{iv0} \mid Z_{ivt} = 1] - \mathbb{E}[\,S_{iv1} - S_{iv0} \mid Z_{ivt} = 0]},$$

where Y is years of education and S is EMRS exposure. Figure 3 displays the spatial overlap of sanctioned EMRS with forest density; regions with thicker cover host more schools, supporting instrument *relevance* ($\gamma_3 \neq 0$).

For the instrument to be valid, the key condition is the exclusion restriction:

$$\mathrm{Cov}(Z_{ivt},\, u_{ivt}) = 0,$$

i.e., the instrument affects years of education only through EMRS exposure. Forest cover alone may correlate with remoteness, labor markets, or service access; likewise, eligibility alone is a cohort indicator that may pick up smooth age trends. The product term $ForestCover_v \times Post_{it}$ isolates the cohort-local increase in exposure that occurs only when an eligible cohort resides in a place with ex ante siting propensity. Because the specification also includes $ForestCover_v$ and $Post_{it}$ as separate terms, any effect of forests that is common to all cohorts, and any cohort effect that is common across places, is absorbed rather than attributed to EMRS. In Section 7, I probe alternative channels and placebo tests and find no evidence that Z_{ivt} moves years of education except through EMRS exposure.

The IV estimate $\widehat{\delta}_1$ is the causal effect of EMRS exposure for compliers: ST individuals whose exposure status changes when they become eligible *and* who live in higher–forest villages where schools were more likely to be placed. In practice, the first stage captures that women are more likely to be exposed to a school if they lived in a high forest village *and* were of school-going age when the local EMRS opened; the reduced form translates this induced exposure into years of completed schooling. Instrument relevance is visible in Figure 3 and borne out by strong first-stage coefficients on Z_{ivt}. Exogeneity is supported by (i) constructing Z_{ivt} as a cohort-by-place interaction that varies at the eligibility margin within villages, (ii) including the terms $ForestCover_v$ and $Post_{it}$ to purge level differences by forest and smooth cohort trends, and (iii) falsification tests (Section 7) showing no relationship between Z_{ivt} and outcomes in periods and cohorts that cannot be affected by EMRS.

Finally, another condition should be that before EMRS opens, outcome and treatment trends must be independent of the instrument. In practice, the gap in years of education

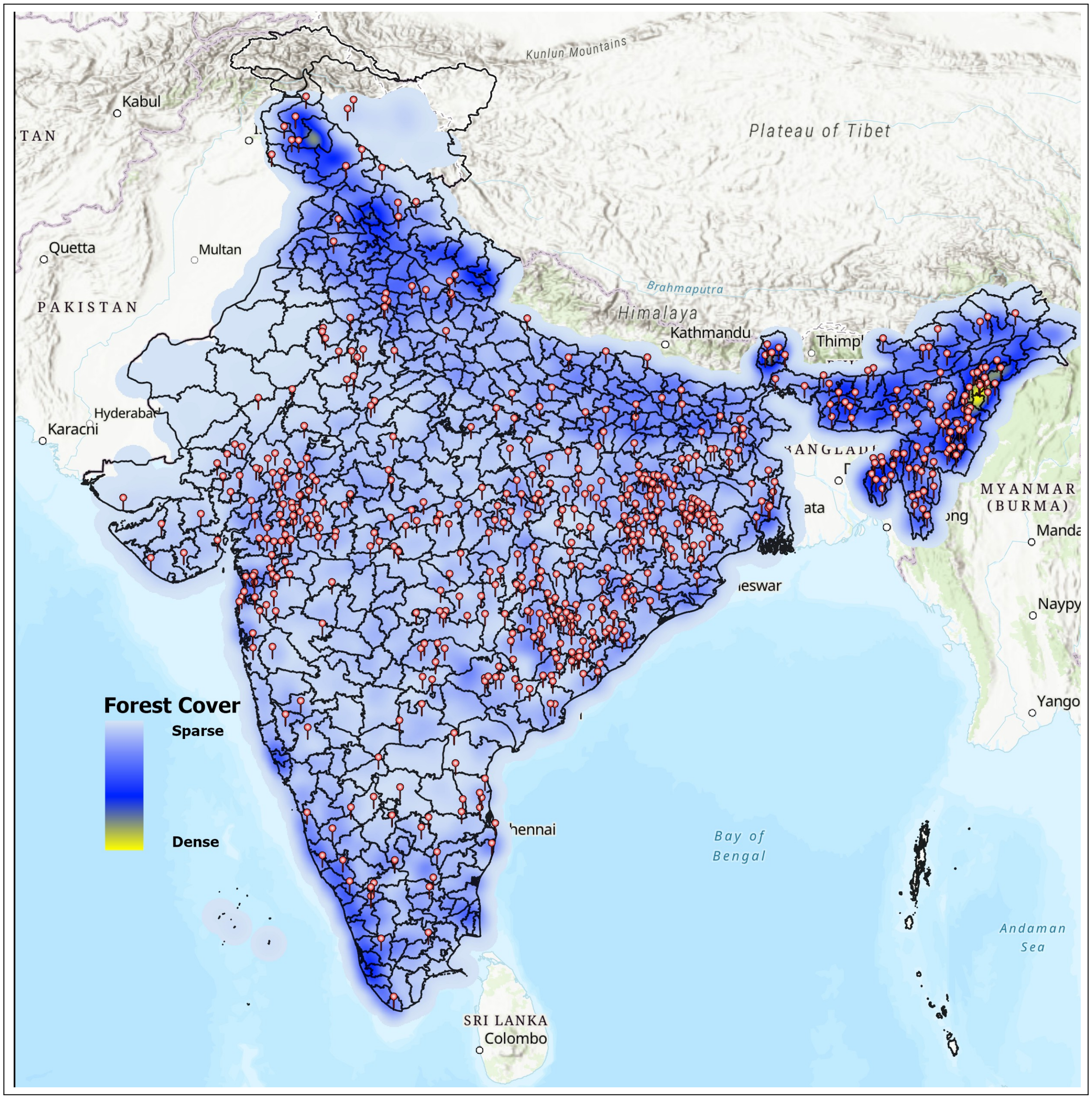

Figure 3: Overlap of EMRS on the Forest Cover of villages across India.

between eligible and ineligible cohorts should follow similar paths in low–forest areas, so that Z_{ivt} is not proxying for differential pre-trends. In the data, pre-treatment gaps between eligible and ineligible cohorts do not differentially trend by forest cover (Section 7). Given relevance, exclusion through interaction with the terms $ForestCover_v$ and $Post_{it}$, and parallel pre-trends, the IV design addresses reverse causality and selection on unobservables that bias OLS.

6 Results II: Instrumental-Variables

I first document instrument relevance. Tables 3(a) and 3(b) report the first stage using the interaction $ForestCover_v \times Post_{it}$ as an instrument for exposure. Across specifications, the coefficient on the interaction is positive and statistically significant. Column (6), which includes controls, state–year fixed effects, and district fixed effects, is the reference. The estimates suggest that cohorts of school-going age in higher-forest areas, where siting was more likely, are more likely to be exposed to an EMRS. Cragg–Donald Wald F-statistics exceed 10 throughout, indicating the instrument is not weak.

Tables 3(a) and 3(b) then turn to the second stage. For years of education, the IV coefficients are negative and stable across specifications, including controls and fixed effects, which leaves magnitudes essentially unchanged, indicating a robust negative impact of school exposure on education. In column (6), the point estimate of -4.145 implies that exposure lowers completed schooling by about four years for those who are eligible and live in higher-forest areas (compliers). For completing secondary education, the IV estimate is negative but statistically indistinguishable from zero at about -12 percentage points. In conjunction with the intensive-margin decline, the null effect on secondary completion indicates that adjustment occurs within the schooling trajectory, via earlier exit, slower grade progression, or increased grade repetition.

Table 3(a): Education: All Districts

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7550***	-0.0277	-0.0251	-0.0463	-0.0175	-0.0682**
	(0.0463)	(0.0499)	(0.0496)	(0.0325)	(0.0539)	(0.0325)
Forest Cover	0.7540***	-0.0312	-0.0289	0.0184	-0.0467	-0.0018
	(0.0537)	(0.0642)	(0.0634)	(0.0767)	(0.0719)	(0.0871)
Forest Cover × Post	-0.6977***	0.0694	0.0638	0.1488***	0.1062	0.1753***
	(0.0747)	(0.0727)	(0.0721)	(0.0548)	(0.0789)	(0.0565)
Cragg-Donald Wald F statistic	1651.402	2.717	2.196	29.160	6.239	36.038
IV Estimation		**Dep. Variable: Years of Education**				
	(1)	(2)	(3)	(4)	(5)	(6)
School	6.035***	-4.036	-2.434	-4.907	-5.051	**-4.145****
	(0.3561)	(9.499)	(9.598)	(3.264)	(4.235)	(2.084)
Observations	4,231	4,231	4,209	4,209	3,604	3,604
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, the gender of the household head, and ST status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 3(b): Completing Secondary Education: All Districts

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7550***	-0.0277	-0.0251	-0.0463	-0.0175	-0.0682**
	(0.0463)	(0.0499)	(0.0496)	(0.0325)	(0.0539)	(0.0325)
Forest Cover	0.7540***	-0.0312	-0.0289	0.0184	-0.0467	0.0018
	(0.0537)	(0.0642)	(0.0634)	(0.0767)	(0.0719)	(0.0871)
Forest Cover × Post	-0.6977***	0.0694	0.0638	0.1488***	0.1062	0.1753***
	(0.0747)	(0.0727)	(0.0721)	(0.0548)	(0.0789)	(0.0565)
Cragg-Donald Wald F statistic	1651.402	2.717	2.196	29.160	6.239	36.038
IV Estimation		**Dep. Variable: Completing Secondary Education**				
	(1)	(2)	(3)	(4)	(5)	(6)
School	0.0748***	0.6637	0.7807	0.1054	0.3045	0.1219
	(0.0085)	(0.7379)	(0.9178)	(0.1127)	(0.2137)	(0.0906)
Observations	4,231	4,231	4,209	4,209	3,604	3,604
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is an indicator which equals 1 if the individual completes secondary education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, the gender of the household head, and ST status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 4(a): Education: Districts with a Residential School

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.8059***	-0.0746	-0.0739*	-0.0478	-0.0911*	-0.0736**
	(0.0449)	(0.0464)	(0.0457)	(0.0339)	(0.0486)	(0.0342)
Forest Cover	0.8084***	-0.0747	-0.0738	0.0165	-0.1088*	-0.0068
	(0.0502)	(0.0599)	(0.0592)	(0.0767)	(0.0656)	(0.0874)
Forest Cover × Post	-0.7489***	0.1355**	0.1359**	0.1522***	0.1967***	0.1840***
	(0.0677)	(0.0662)	(0.0656)	(0.0564)	(0.0711)	(0.0586)
Cragg-Donald Wald F statistic	1751.413	12.760	12.286	26.949	20.492	33.580
IV Estimation	**Dep. Variable: Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
School	5.675***	11.924	12.945	-5.035	1.762	**-4.177****
	(0.3287)	(7.644)	(7.989)	(3.286)	(2.330)	(2.051)
Observations	3,860	3,862	3,844	3,844	3,238	3,238
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, the gender of the household head, and ST status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Tables 4(a) and 4(b) restrict the sample to districts with at least one EMRS. This restriction removes noise from districts that never host a school and reduces selection concerns arising from systematic differences between school and non-school districts. The first stage remains strong, with the interaction coefficients positive and significant and the Cragg-Donald Wald $F > 10$. Second-stage estimates mirror the full-sample results: in column (6) of Table 4(a), IV implies roughly four fewer years of education for compliers in heavily forested districts; secondary completion effects in Table 4(b) remain negative and imprecise at about -12 percentage points. The similarity to Tables 3(a)–3(b) indicates that the main IV results are not driven by differential composition across districts without schools.

Finally, I examine village-level outcomes to capture potential community-wide effects. Because schools were sanctioned at the village level, any impacts may operate both directly through those eligible to enroll and indirectly through peers, family labor allocation, or local attitudes toward schooling. Tables 5(a) and 5(b) report IV estimates using village-level

Table 4(b): Completing Secondary Education: Districts with a Residential School

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.8059***	-0.0772*	-0.0764*	-0.0487	-0.0950**	-0.0750
	(0.0451)	(0.0464)	(0.0455)	(0.0341)	(0.0477)	(0.0344)
Forest Cover	0.8074***	-0.0776	-0.0771	0.0115	-0.1130*	-0.0092
	(0.0505)	(0.0596)	(0.0590)	(0.0765)	(0.0653)	(0.0873)
Forest Cover × Post	-0.7458***	0.1418**	0.1426**	0.1560***	0.2063**	0.1871***
	(0.0678)	(0.0659)	(0.0654)	(0.0562)	(0.0706)	(0.0584)
Cragg-Donald Wald F statistic	1751.467	13.916	13.451	27.127	22.536	34.365
IV Estimation	**Dep. Variable: Completing Secondary Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
School	0.0714***	0.7158*	1.914	0.0986	0.2837**	0.1197
	(0.0084)	(0.3876)	(0.3966)	(0.1124)	(0.1411)	(0.0897)
Observations	3,858	3,858	3,840	3,840	3,238	3,238
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is an indicator which equals 1 if the individual completes secondary education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, the gender of the household head, and ST status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

averages. The first stage remains robust. For years of education, point estimates indicate a decline of about four years, albeit imprecisely estimated at the village average; for secondary completion, effects are small and statistically indistinguishable from zero. The alignment between respondent-level (Tables 3–4) and village-level (Table 5) estimates suggests that the intensive-margin loss in schooling is not purely idiosyncratic to treated individuals but is also reflected in local aggregates.

In sum, the IV evidence corroborates the triple difference findings: (i) the interaction instrument is relevant and not weak; (ii) exposure substantially reduces completed years of schooling for compliers; and (iii) there is no precise shift in the probability of secondary completion. The stability of magnitudes with controls and fixed effects, the persistence of the first stage, and the replication within the restricted school-district sample strengthen the conclusion that EMRS exposure lowers educational attainment along the intensive margin.

Table 5(a): Average Education: Village Level

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7550***	-0.0104	-0.0277	-0.0503	-0.0251	-0.0463
	(0.0463)	(0.0546)	(0.0499)	(0.0332)	(0.0496)	(0.0325)
Forest Cover	0.7540***	-0.0092	-0.0312	0.0182	-0.0289	0.0184
	(0.0537)	(0.0698)	(0.0642)	(0.0772)	(0.0634)	(0.0767)
Forest Cover × Post	-0.6977***	0.0660	0.0694	0.1504***	0.0638	0.1488***
	(0.0747)	(0.0800)	(0.0727)	(0.0554)	(0.0721)	(0.0648)
Cragg-Donald Wald F statistic	1651.402	2.793	2.717	31.651	2.196	29.160
IV Estimation	**Dep. Variable: Average Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
School	7.654***	-3.392	-2.994	-3.835	-3.519	-3.824
	(0.3895)	(7.430)	(7.300)	(2.450)	(8.096)	(2.438)
Observations	4,231	4,231	4,231	4,231	4,209	4,209
State FE	-	✓	✓	-	-	-
Year FE	-	-	✓	-	-	-
State-Year FE	-	-	-	-	✓	✓
District FE	-	-	-	✓	-	✓

Notes: The dependent variable is the average years of education. The interaction of forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 5(b): Average Completion of Secondary Education: Village Level

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7550***	-0.0104	-0.0277	-0.0503	-0.0251	-0.0463
	(0.0463)	(0.0546)	(0.0499)	(0.0332)	(0.0496)	(0.0325)
Forest Cover	0.7540***	-0.0092	-0.0312	0.0182	-0.0289	0.0184
	(0.0537)	(0.0698)	(0.0642)	(0.0772)	(0.0634)	(0.0767)
Forest Cover × Post	-0.6977***	0.0660	0.0694	0.1504***	0.0638	0.1488***
	(0.0747)	(0.0800)	(0.0727)	(0.0554)	(0.0721)	(0.0648)
Cragg-Donald Wald F statistic	1651.402	2.793	2.717	31.651	2.196	29.160
IV Estimation	**Dep. Variable: Average Completion of Secondary Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
School	0.0994***	0.2086	0.2258	-0.0136	0.2081	-0.0293
	(0.0089)	(0.3279)	(0.3395)	(0.0845)	(0.3542)	(0.0840)
Observations	4,231	4,231	4,231	4,231	4,209	4,209
State FE	-	✓	✓	-	-	-
Year FE	-	-	✓	-	-	-
State-Year FE	-	-	-	-	✓	✓
District FE	-	-	-	✓	-	✓

Notes: The dependent variable is the village average of an indicator, which equals 1 if the individual completes secondary education. The interaction of forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

7 Robustness

7.1. Difference-in-Difference

This subsection verifies that the triple-difference estimates are causal under the relevant parallel-trends restriction. Absent EMRS, the way the ST–non-ST gap changes when we move from cohorts too old/too young to attend (ineligible) to cohorts that were school-age at opening (eligible) would have been the same in villages with and without a school. I provide two complementary checks: (i) a triple difference event study that shows flat pre-period triple-interaction coefficients and the post-opening dynamics, and (ii) design-consistent placebos—ineligible cohorts, non-ST samples, and placebo school locations, as expected, produce null triple interactions.

First, I replace the binary eligibility indicator with year of birth and estimate a triple difference event-study with $ST \times School \times$ year of birth interactions. Figure 5 plots the sequence of triple-interaction coefficients. The ineligible cohort (before 1989) do not show any statistically significant coefficient, indicating no anticipatory divergence in the ST–non-ST gap prior to exposure. Eligible cohort coefficients trace the negative intensive-margin response for these cohorts.

Second, to assess whether the estimates are spuriously picking up unrelated shocks, I apply (1) to groups and exposures that should not be affected by EMRS:

1. *Ineligible cohorts within treated villages:* I re-estimate (1), restricting to women who were too old to attend EMRS at opening. Because eligibility is defined mechanically by age at the local opening date, these cohorts form a within-village placebo. The triple interaction is statistically nil, consistent with no effect outside the exposed age window.

2. *Non-ST placebo:* I re-estimate within the non-ST sample, specifically the minority of Scheduled Castes (SC). Because the triple difference estimand is the change in the

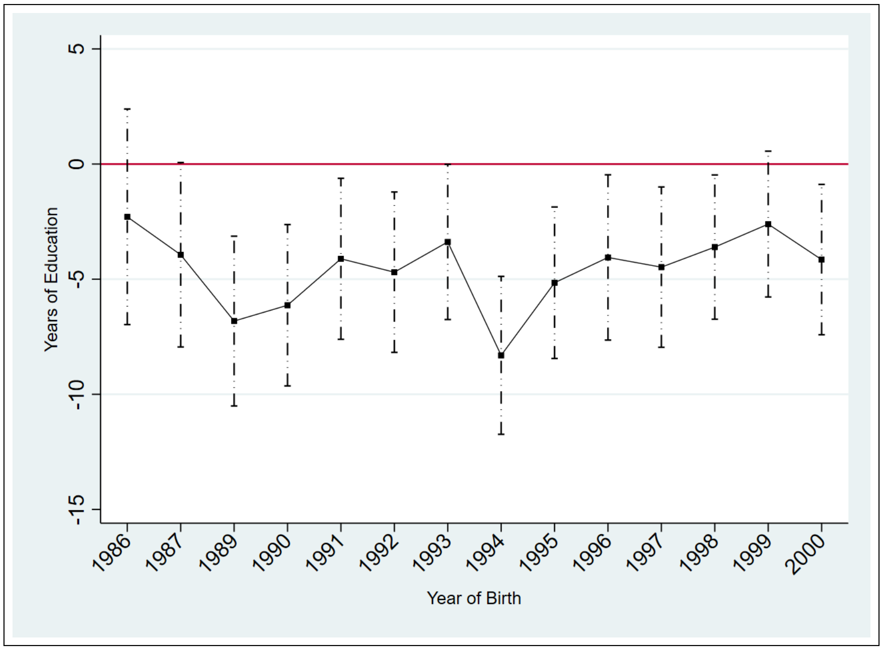

Figure 4: Education comparison across birth years.

ST–non-ST gap at eligibility, the coefficients are statistically insignificant from zero for SCs, as expected.

3. *Placebo siting ("fake schools").* Following Roth et al. (2023), I assign placebo EMRS locations to villages that never receive a school, recompute the school mapping and event time, and re-estimate. By construction, these villages are untreated; significant triple interactions would indicate that unrelated geography or time patterns are masquerading as treatment. Across random assignments, the distribution of the placebo triple interactions is statistically insignificant; Table 11 reports the baseline placebo specification.

4. *Lead-placebo (future openings):* I construct "leads" that code eligibility relative to a future opening year and re-estimate using these pseudo treatment timings. Coefficients on the lead triple interactions are indistinguishable from zero, ruling out spurious pre-activation.

Tables 6–9 report these results. Across all placebo designs, triple-interaction coefficients are statistically indistinguishable from zero. In particular, (i) ineligible cohorts show no effect, (ii) the non-ST placebo reveals no impact of EMRS, and (iii) placebo siting/future-lead tests produce nulls. Together with the flat pre-period in the DDD event study, these findings support the maintained parallel-trends restriction in (3) and indicate that the main triple-difference estimates are not artifacts of anticipatory behavior, sample composition, or unrelated place-level dynamics.

Table 6: Education: Triple Difference (Ineligible Cohorts)

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × Eligible × Placebo School	-0.0868	0.1380	0.2244	0.0055	-0.7737	-0.8730
	(0.7466)	(0.7110)	(0.7372)	(0.7589)	(0.7489)	(0.7435)
Observations	5,984	5,984	5,975	5,975	4,339	4,339
R-squared	0.0878	0.1680	0.2342	0.4278	0.4606	0.5288
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 7: Education: Triple Difference (Non-ST: Scheduled Castes)

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × Eligible × Placebo School	-0.0868	0.1380	0.2244	0.0055	-0.7737	-0.8730
	(0.7466)	(0.7110)	(0.7372)	(0.7589)	(0.7489)	(0.7435)
Observations	5,984	5,984	5,975	5,975	4,339	4,339
R-squared	0.0878	0.1680	0.2342	0.4278	0.4606	0.5288
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 8: Education: Triple Difference (Fake Schools)

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × Eligible × Placebo School	-0.0868	0.1380	0.2244	0.0055	-0.7737	-0.8730
	(0.7466)	(0.7110)	(0.7372)	(0.7589)	(0.7489)	(0.7435)
Observations	5,984	5,984	5,975	5,975	4,339	4,339
R-squared	0.0878	0.1680	0.2342	0.4278	0.4606	0.5288
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 9: Education: Triple Difference (Future Openings)

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × Eligible × Placebo School	-0.0868	0.1380	0.2244	0.0055	-0.7737	-0.8730
	(0.7466)	(0.7110)	(0.7372)	(0.7589)	(0.7489)	(0.7435)
Observations	5,984	5,984	5,975	5,975	4,339	4,339
R-squared	0.0878	0.1680	0.2342	0.4278	0.4606	0.5288
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

7.2. Instrumental Variable

This subsection assembles evidence that the exclusion restriction underlying the IV design is credible. Recall that the instrument is the interaction of local forest cover and cohort eligibility, $Z_{ivt} = ForestCover_v \times Post_{it}$. For $\widehat{\delta}_1$ to be interpreted causally, Z_{ivt} must shift years of education only through EMRS exposure. Any alternative pathway whereby Z_{ivt} directly alters schooling would bias the IV estimates. I therefore examine classes of violations that include geographic confounds, health channels, pre-policy associations, and a population that should be unaffected by EMRS.

Geographic confounds: If forested locations systematically differ along other geographic features that themselves lower schooling (e.g., elevation or terrain ruggedness), the interaction could proxy for hard-to-access places rather than EMRS exposure. Following Nunn and Qian (2014), I augment the baseline with these placebo geography measures and test whether they replicate the forest-cover pattern. Table 10 reports the results. Columns (1)–(3) compare the baseline to specifications that add elevation and ruggedness for all districts; columns (4)–(6) repeat the exercise within the subset of districts that have an EMRS. Across panels, placebo geography does not load in the way forest cover does: coefficients are small and statistically indistinguishable from zero, and their inclusion does not attenuate the instrumented effect. This pattern indicates that $ForestCover_v \times Post_{it}$ is not proxying for a generic "difficult terrain" channel.

Health channels: A second concern is that denser forests may be correlated with disease environments that raise absenteeism or reduce cognitive performance, and that these health burdens could differentially affect cohorts just as they reach school-going ages, independent of EMRS. To address this, I follow Magesan and Swee (2018) and include health covariates that capture conditions developing over the life course and plausibly rooted in early environments. Tables 11 and 12 report the augmented specifications. Columns (4) and (5) show that adding these health controls leaves the IV point estimates on years of education essentially unchanged and the first stage intact. This is consistent with the exclusion restriction: conditional on the saturated lower-order terms and observed health risks, the interaction affects schooling only through EMRS exposure.

Table 10: Other Geographical Characteristics

	All Districts			Districts with a School		
First Stage: School Exposure	Forest Cover	Terrain Ruggedness	Elevation	Forest Cover	Terrain Ruggedness	Elevation
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.0682	0.0003	-0.0765	-0.0736**	0.0027	-0.0856
	(0.0325)	(0.0504)	(0.0786)	(0.0342)	(0.0530)	(0.0879)
Geographic Characteristic	-0.0018	-0.0859	0.0737	-0.0068	-0.0821	0.0696
	(0.0871)	(0.0752)	(0.1051)	(0.0874)	(0.0760)	(0.1040)
Geographic Characteristic × Post	0.1753***	0.0150	0.0995	0.1840***	0.0116	0.1081
	(0.0565)	(0.0621)	(0.0851)	(0.0586)	(0.0655)	(0.0945)
Cragg-Donald Wald F statistic	36.038	6.742	15.960	33.580	5.795	14.674
IV Estimation	**Dep. Variable: Years of Education**					
School	**-4.145****	-6.130	-0.2917	**-4.177****	-6.308	-0.9918
	(2.084)	(7.341)	(2.042)	(2.051)	(1.653)	(2.081)
Observations	3,604	3,604	3,604	3,238	3,238	3,238
State-Year FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the geographic characteristic and cohort eligibility is the respective instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Pre-policy associations: The exclusion restriction also rules out the possibility that $ForestCover_v \times Post_{it}$ predicts education *before* schools are sanctioned, i.e., that the interaction captures cohort-local differences in schooling unrelated to EMRS. I therefore re-estimate in samples restricted to pre-sanction periods and to cohorts that are coded eligible relative to a future (not yet realized) opening date. Tables 13 and 14 show that the coefficients on the interaction are statistically insignificant across specifications. The absence of pre-treatment associations at the cohort margin supports the view that the interaction operates only once an EMRS is operational.

Table 11: Education with Health Controls: All Districts

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)
Post	-0.0251	-0.0503	-0.0463	-0.0682**	-0.0679
	(0.0496)	(0.0332)	(0.0325)	(0.0325)	(0.0321)
Forest Cover	-0.0289	0.0182	0.0184	0.0018	0.0010
	(0.0634)	(0.0772)	(0.0767)	(0.0871)	(0.0869)
Forest Cover × Post	0.0638	0.1504***	0.1488***	0.1753***	0.1742***
	(0.0721)	(0.0554)	(0.0548)	(0.0565)	(0.0564)
Cragg-Donald Wald F statistic	2.196	31.651	29.160	36.038	35.737
IV Estimation	**Dep. Variable: Years of Education**				
School	-2.434	-5.516	-4.907	**-4.145****	**-4.301****
	(9.598)	(3.387)	(3.264)	(2.084)	(2.105)
Abnormal Glucose Level	-	-	-	-	-1.500
	-	-	-	-	(1.298)
Stunting	-	-	-	-	-0.3786
	-	-	-	-	(0.2692)
Anemia	-	-	-	-	0.0052
	-	-	-	-	(0.2104)
Hemoglobin	-	-	-	-	-0.1415
	-	-	-	-	(0.1981)
Hypertension	-	-	-	-	-0.0212
	-	-	-	-	(0.0829)
Respiratory Disease	-	-	-	-	-0.1970
	-	-	-	-	(0.2141)
Heart Disease	-	-	-	-	0.0449
	-	-	-	-	(0.1852)
Observations	4,209	4,231	4,209	3,604	3,604
State-Year FE	✓	-	✓	✓	✓
District FE	-	✓	✓	✓	✓
Controls	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include ST status, income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 12: Education with Health Controls: Districts with a Residential School

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)
Post	-0.0751*	-0.0519	-0.0481	-0.0739	-0.0736**
	(0.0457)	(0.0350)	(0.0339)	(0.0344)	(0.0340)
Forest Cover	-0.0733	0.0156	-0.0157	-0.0077	-0.0069
	(0.0589)	(0.0774)	(0.0767)	(0.0874)	(0.0871)
Forest Cover × Post	0.1410**	0.1529***	0.1524***	0.1842***	0.1830***
	(0.0654)	(0.0569)	(0.0562)	(0.0585)	(0.0584)
Cragg-Donald Wald F statistic	13.007	28.622	26.780	33.558	33.226
IV Estimation	**Dep. Variable: Years of Education**				
School	11.497	-5.479	-4.870	**-4.082****	**-4.200****
	(7.225)	(3.404)	(3.256)	(2.041)	(2.054)
Abnormal Glucose Level	-	-	-	-	-1.399
	-	-	-	-	(1.313)
Stunting	-	-	-	-	-0.3050
	-	-	-	-	(0.2912)
Anemia	-	-	-	-	0.0643
	-	-	-	-	(0.2234)
Hemoglobin	-	-	-	-	0.0407
	-	-	-	-	(0.2118)
Hypertension	-	-	-	-	-0.0355
	-	-	-	-	(0.1096)
Respiratory Disease	-	-	-	-	-0.2146
	-	-	-	-	(0.2264)
Heart Disease	-	-	-	-	0.1816
	-	-	-	-	(0.1323)
Observations	3,839	3,858	3,839	3,234	3,234
State-Year FE	✓	-	✓	✓	✓
District FE	-	✓	✓	✓	✓
Controls	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include ST status, income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 13: Education Before Schools: All Districts

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7606***	-0.0780	-0.0885	0.0703	-0.0986	0.0651
	(0.0370)	(0.0642)	(0.0790)	(0.0560)	(0.0872)	(0.0573)
Forest Cover	0.7865***	0.0074	0.0021	0.0892	0.0083	0.0801
	(0.0409)	(0.0529)	(0.0528)	(0.0673)	(0.0564)	(0.0720)
Forest Cover × Post	-0.7648***	-0.0072	0.0014	0.0011	0.0007	0.0076
	(0.0404)	(0.0198)	(0.0208)	(0.0133)	(0.0292)	(0.0119)
Cragg-Donald Wald F statistic	2317.043	1.088	0.681	15.412	0.855	12.380
IV Estimation		**Dep. Variable: Years of Education**				
School	6.986***	-4.213	-1.719	-8.119	0.8116	-7.404
	(0.3535)	(8.581)	(9.938)	(6.982)	(8.582)	(6.464)
Observations	5,898	5,898	5,868	5,868	5,048	5,048
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include ST status, income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 14: Education Before Schools: Districts with a Residential School

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.8497***	-0.0420	-0.0665	0.0737	-0.0688	-0.0657
	(0.0321)	(0.0589)	(0.0748)	(0.0590)	(0.0782)	(0.0612)
Forest Cover	0.8425***	-0.0011	-0.0084	0.0870	-0.0018	0.0777
	(0.0376)	(0.0502)	(0.0497)	(0.0668)	(0.0521)	(0.0710)
Forest Cover × Post	-0.8489***	-0.0045	0.0086	0.0067	0.0057	0.0141
	(0.0359)	(0.0194)	(0.0217)	(0.0154)	(0.0224)	(0.0142)
Cragg-Donald Wald F statistic	2576.071	0.392	0.533	14.373	0.396	11.465
IV Estimation		**Dep. Variable: Years of Education**				
School	6.432***	26.078	24.634	-8.271	2.173	-7.229
	(0.3106)	(49.577)	(44.195)	(7.034)	(11.506)	(6.241)
Observations	5,371	5,371	5,345	5,345	4,528	4,526
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include ST status, income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Unaffected population: Finally, because the treatment margin in this study is EMRS exposure for Scheduled Tribes, non-ST women should not exhibit any relationship between the instrument and schooling. Tables 15 and 16 confirm this placebo: in the non-ST sample, the interaction does not impact years of education, and IV effects are indistinguishable from zero. The lack of a signal in an explicitly untreated population further supports the validity of the instrument.

Taken together, these checks point in the same direction. Placebo geographies do not replicate the instrument's predictive content; health channels do not mediate the estimated effect; the interaction fails to predict outcomes in pre-sanction periods; and untreated populations show no response. With relevance established in the first stage, the evidence is consistent with the interpretation that $ForestCover_v \times Post_{it}$ shifts years of education solely through EMRS exposure at the eligibility margin.

Table 15: Education For Non-ST: All Districts

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.7599***	-0.0501	-0.0889	0.0558	-0.1424	0.0291
	(0.0457)	(0.0685)	(0.0879)	(0.0407)	(0.1047)	(0.0323)
Forest Cover	0.7815***	-0.0008	0.0159	0.0528	-0.0019	-0.0027
	(0.0598)	(0.0703)	(0.0697)	(0.0782)	(0.0849)	(0.0956)
Forest Cover × Post	-0.7930***	-0.0297	-0.0467	-0.0297	-0.0512	-0.0254
	(0.0592)	(0.0435)	(0.0433)	(0.0261)	(0.0495)	(0.0225)
Cragg-Donald Wald F statistic	2065.250	1.338	1.628	1.361	3.176	0.802
IV Estimation	**Dep. Variable: Years of Education**					
School	9.441***	21.152	18.525	-5.747	0.9432	16.486
	(0.4844)	(39.252)	(24.083)	(12.842)	(4.472)	(42.014)
Observations	2,949	2,949	2,920	2,920	2,276	2,271
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table 16: Education For Non-ST: Districts with a Residential School

First Stage: School Exposure	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.8878***	0.0083	0.0094	0.0530	-0.0252	0.0327
	(0.0346)	(0.0570)	(0.0687)	(0.0448)	(0.0880)	(0.0378)
Forest Cover	0.8518***	-0.0490	-0.0300	0.0600	-0.0601	0.0095
	(0.0514)	(0.0599)	(0.0560)	(0.0767)	(0.0672)	(0.0933)
Forest Cover × Post	-0.9363***	-0.0240	-0.0472	-0.0370	-0.0685*	-0.0361
	(0.0484)	(0.0376)	(0.0374)	(0.0291)	(0.0415)	(0.0270)
Cragg-Donald Wald F statistic	2816.844	8.545	8.398	1.355	17.438	0.753
IV Estimation	**Dep. Variable: Years of Education**					
School	8.531***	11.371	11.846	-5.686	1.352	12.649
	(0.3903)	(11.806)	(11.160)	(11.763)	(2.605)	(27.234)
Observations	2,614	2,614	2,587	2,587	1,941	1,938
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is the total years of education. The interaction of the forest cover and cohort eligibility is the instrument. The estimation controls for state-year and district-fixed effects. Controls include income status, marital status, religion, and the gender of the household head. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

8 Mechanism: Domestic Work and the Cost of Residence

To investigate why EMRS lowers years of schooling, I switch the dependent variable to indicators for stated *reasons for leaving school* and estimate the same designs as in the main analysis. Outcomes are coded as intent–to–treat: $Y_i = 1$ if the respondent reports a given reason (e.g., "domestic work") and $Y_i = 0$ otherwise (including those still enrolled or who left for other reasons). This avoids conditioning on the post-treatment event of dropping out.

The triple-difference specification is unchanged except for the outcome, with state–year fixed effects $\gamma_{s(v),t(g(i))}$, district fixed effects $\theta_{d(v)}$, and the same individual controls as in the main tables. For the IV version, I instrument exposure with $VCF_v \times Eligible_{g(i)}$ and estimate 2SLS with the identical set of fixed effects and controls.

Results: Table 17 reports the coefficients for the leading reasons. For "domestic work," the triple-difference coefficient on $ST \times School \times Eligible$ is 0.0364 (s.e. 0.0183), implying a 3.64 percentage-point increase in the probability that an eligible ST woman cites domestic responsibilities as the reason for leaving school in villages with an operational EMRS. The IV estimates corroborate this mechanism on the complier margin: the coefficient on instrumented exposure is 0.0939 (s.e. 0.0550). By contrast, estimated effects for marriage, sibling care, "not interested," and "costly" are small and statistically indistinguishable from zero in both panels. Standard errors are clustered at the DHS cluster level; state–year and district fixed effects are included throughout; first-stage statistics exceed conventional thresholds.

Table 17: Dropout Reason

Dropout Reasons	Domestic Work	Marriage	Sibling Care	Not Interested	Costly
Triple Difference					
$ST \times School \times Eligible$	**0.0364****	0.0190	0.0017	-0.0265	-0.0112
	(0.0183)	(0.0199)	(0.0020)	(0.0209)	(0.0129)
Observations	7,042	7,042	7,042	7,042	7,042
R-squared	0.1533	0.1573	0.1208	0.1618	0.1487
IV Estimation					
School	**0.0939***	0.0298	-0.0105	0.0613	0.0629
	(0.0550)	(0.0430)	(0.0092)	(0.0785)	(0.0446)
Observations	3,604	3,604	3,604	3,604	3,604
Cragg-Donald Wald F statistic	36.362	36.362	36.362	36.362	36.362
State FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
State-Year FE	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓

Notes: Outcome equals 1 if the respondent reports the listed reason for leaving school, 0 otherwise. DDD coefficients are on $ST \times School \times Eligible$ with the full set of pairwise interactions; IV instruments exposure with $VCF \times Eligible$. All specifications include the same controls, state–year fixed effects, and district fixed effects as in the main analysis. Standard errors clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01

The evidence points to a time–use mechanism. The only stated reason that rises is "domestic work," while other reasons—including unaffordability and early marriage—do not change. EMRS requires full–year residence on campus, which removes daughters' labor from the household for most of the year. Many ST households depend on adolescent girls to support domestic production (cooking, water and fuel collection, and farming) and, in some cases, to contribute to small income–earning activities. When that labor is withdrawn, the day–to–day burden on the family increases. In this setting, the opportunity cost of keeping a child enrolled in a residential school becomes high, prompting parents to cancel enrollment or withdraw the child earlier than they otherwise would. The absence of effects on “costly” is consistent with EMRS covering direct fees and supplies, and the null for “marriage” indicates shifts in marriage timing do not drive the pattern. Taken together, the selective rise in “domestic work” and the nulls for competing explanations isolate a mechanism operating through household labor demands created by compulsory residence.

A second piece of evidence concerns the absence of day–school alternatives. If students could attend a local day school, they could return home after classes and continue to contribute to household tasks, easing the time pressure on families. Yet, as Figure 5 documents, districts that host an EMRS have few, if any, day schools. This scarcity effectively removes a nonresidential option and leaves families with a binary choice: enroll a daughter in a residential EMRS and forgo her household labor, or withdraw her from schooling. In the ST communities studied, the results are consistent with the latter response. The pattern points to a design problem, where the residential model crowds out feasible nonresidential schooling. It underscores the need for provision that accounts for household time constraints in socioeconomically disadvantaged settings.

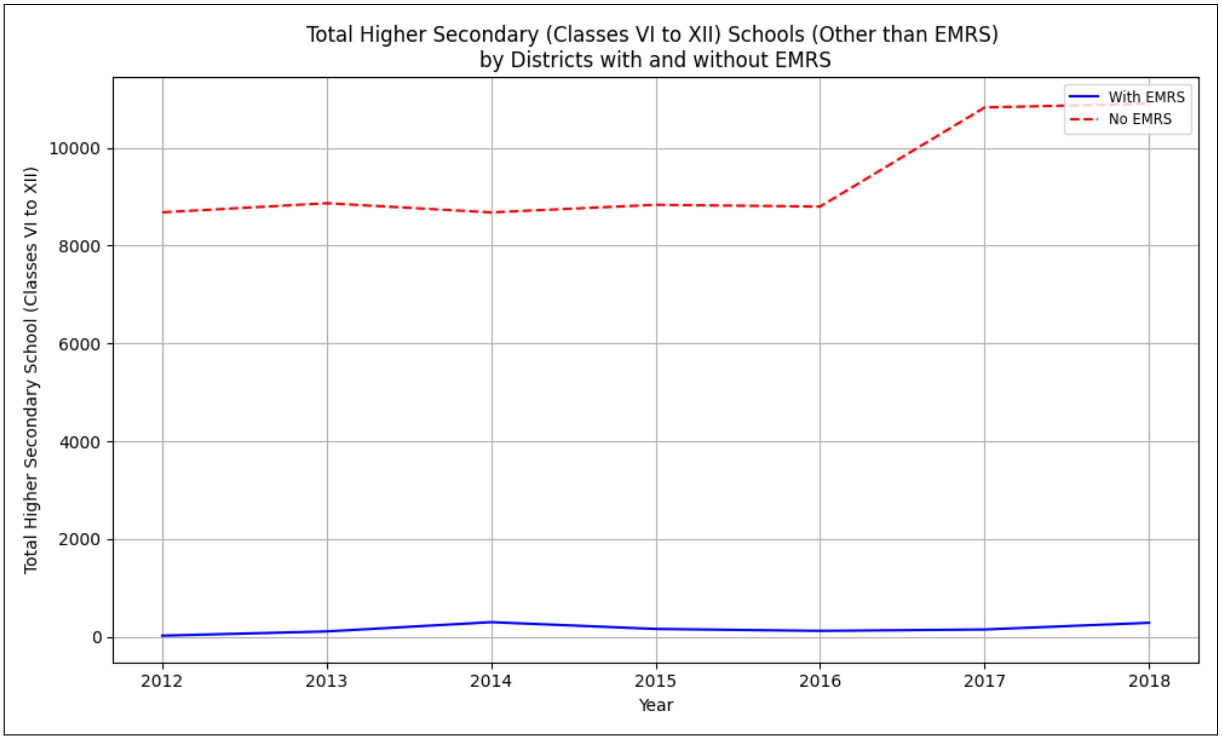

Figure 5

9 Conclusion

I provide the first comprehensive and causal evaluation of India's contemporary residential-school model for Indigenous[7] communities. The policy stakes are immediate and large. Against the backdrop of a global decline and phase-out of residential schools, India allocated INR 63.99 billion ($\approx$CAD 1.05 billion), sanctioned 38,800 teachers, and targets a network of 740 *Eklavya Model Residential Schools* (EMRS). To study the causal impact, I create a new micro–spatial dataset that links every EMRS to nationally representative NFHS village clusters and to village-level forest cover using deterministic nearest–neighbor matches based on exact geo–coding. This dataset delivers respondent–level exposure to the nearest operational school and a consistent environmental measure aligned with NFHS geography.

[7]Scheduled Tribes (ST) of India.

I use two identification strategies. First, a triple–difference approach contrasts (i) ST with non–ST, (ii) villages with versus without an EMRS, and (iii) adjacent cohorts split by age at opening. Second, an instrumental–variables approach uses the interaction of local forest cover and cohort eligibility as an instrument for school exposure. Forest cover captures siting propensity, and eligibility converts that propensity into potential access at the cohort margin. Both approaches point to the same conclusion: the residential model reduces the years of completed schooling for the intended beneficiaries.

The estimates are stark. In the triple–difference results, ST women who were school–age when an EMRS opened in their village complete 1.248 fewer years of education than the relevant counterfactual constructed from non–ST and ineligible cohorts in non–school villages. The instrumental variables estimates imply a larger decline of about four years for compliers, and I observe a similar decline at the village average. Event–study graphs show flat pre–period triple–interaction coefficients, and design–consistent placebos, including ineligible cohorts, the non–ST sample, future–opening leads, and placebo siting, produce null effects. IV exclusion checks based on geography placebos (elevation, ruggedness), added health controls, pre-policy associations, and the non–ST sample support the interpretation that the instrument affects schooling only through EMRS exposure.

I then examine why schooling falls. The evidence points to a time–use mechanism: only "domestic work" rises as a stated reason for leaving school, while unaffordability and early marriage do not change. EMRS requires full–year residence on campus, which removes daughters' labor from the household for most of the year. Many ST households depend on adolescent girls to support domestic production, including cooking, water and fuel collection, farming, and, in some cases, small income–earning activities. When that labor is withdrawn, the day–to–day burden on the family increases. In this setting, the opportunity cost of continued enrollment in a residential school becomes high, and parents cancel enrollment or withdraw earlier than they otherwise would. A second fact reinforces this interpretation:

EMRS districts have few, if any, day–school alternatives. Without a nonresidential option that allows children to contribute at home after classes, families face a binary choice: enroll in a residential EMRS and forgo essential household labor, or withdraw from schooling. The estimates are consistent with the latter response.

These findings carry direct policy implications for a program that now commands substantial public resources. First, covering tuition, uniforms, food, and boarding does not resolve the binding constraint when households depend on adolescents' time. Second, delivery design matters: a model that requires residence can depress attainment if it displaces feasible day–school options. Policymakers should consider expanding accessible day schools in EMRS catchments, introducing residency calendars that guarantee predictable periods at home during the academic year, and creating structured family–school engagement that reduces the household cost of enrollment. With the massive expansion planned, it is essential to align delivery with the lived constraints of ST households to achieve the program's stated objective.

I focus on short–run schooling outcomes in this paper. In future work, I will extend the analysis beyond education to political incorporation, exploiting the EMRS siting rule (blocks with high ST shares and populations) to generate quasi-experimental variation in exposure and linking geocoded school rollout to election returns. I will estimate effects on turnout and support for tribal versus non-tribal parties, and field pilot surveys[8] to measure political knowledge, attribution/patronage, and whether using school campuses as polling places lowers voting costs. I will also test whether the language of instruction shifts participation beyond the community level and whether impacts differ across plains and hill tribes. In parallel, I will trace downstream effects on health, earnings, and cultural participation; quantify heterogeneity by local demand for adolescent time and by the availability of day schools; and evaluate program variants that relax residence while maintaining instructional quality.

[8]Data collection contract in the final stages of discussion.

References

Adams, David Wallace. 1995. "Education for extinction." *Lawrence: University Press of Kansas* 23:164.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo and Leigh Linden. 2007. "Remedying education: Evidence from two randomized experiments in India." *The Quarterly Journal of Economics* 122(3):1235–1264.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L Linden and Francisco Perez-Calle. 2011. "Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia." *American Economic Journal: Applied Economics* 3(2):167–95.

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.

Bombay, Amy, Kimberly Matheson and Hymie Anisman. 2014. "The intergenerational effects of Indian Residential Schools: Implications for the concept of historical trauma." *Transcultural psychiatry* 51(3):320–338.

Bougie, Evelyne and Sacha Senécal. 2010. "Registered Indian children's school success and intergenerational effects of residential schooling in Canada." *International Indigenous Policy Journal* 1(1):1–41.

Breierova, Lucia and Esther Duflo. 2004. "The impact of education on fertility and child mortality: Do fathers really matter less than mothers?".

Census of India and Office of the Registrar General & Census Commissioner. 2013. "Scheduled Tribes in India: as revealed in Census 2011.". Census 2011 summary with ST share (8.6%) and group counts. Accessed 2025-10-11.

URL: *https://ruralindiaonline.org/en/library/resource/scheduled-tribes-in-india-as-revealed-in-census-2011/*

Constitution of India (Commentary). 2024*a*. "Article 342: Scheduled Tribes.". Explains Presidential notification defining STs (Article 342). Accessed 2025-10-11.
URL: *https://www.constitutionofindia.net/articles/article-342-scheduled-tribes-2/*

Constitution of India (Commentary). 2024*b*. "Article 46: Promotion of educational and economic interests of SCs, STs and other weaker sections.". Directive Principle on education/economic interests and protection from exploitation. Accessed 2025-10-11.
URL: *https://www.constitutionofindia.net/articles/article-46-promotion-of-educational-and-economic-interests-of-scheduled-castes-scheduled-tribes-and-other-weaker-sections/*

Deming, David and Susan Dynarski. 2009. Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. Technical report National Bureau of Economic Research.

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American economic review* 91(4):795–813.

Dynarski, Susan. 2004. The new merit aid. In *College choices: The economics of where to go, when to go, and how to pay for it*. University of Chicago Press pp. 63–100.

Dynarski, Susan and Judith Scott-Clayton. 2013. "Financial aid policy: Lessons from research.".

Feir, Donna L. 2016. "The long-term effects of forcible assimilation policy: The case of Indian boarding schools." *Canadian Journal of Economics/Revue canadienne d'économique* 49(2):433–480.

Feir, Donna L and M Christopher Auld. 2021. "Indian residential schools: Height and body mass post-1930." *Canadian Journal of Economics/Revue canadienne d'économique* 54(1):126–163.

Glenn, Charles. 2011. *American Indian/First Nations schooling: From the colonial period to the present*. Springer.

Government of India, Ministry of Tribal Affairs. 2010. "Revised Guidelines for Setting Up Eklavya Modern Residential School (EMRS).".

Gregg, Matthew T. 2018. "The long-term effects of American Indian boarding schools." *Journal of Development Economics* 130:17–32.

Jedwab, Remi, Felix Meier zu Selhausen and Alexander Moradi. 2022. "The economics of missionary expansion: Evidence from Africa and implications for development." *Journal of Economic Growth* 27(2):149–192.

Jones, Maggie EC. 2021. "The intergenerational legacy of indian residential schools." *University of Victoria, Victoria, British Columbia. At https://maggieecjones. files. wordpress. com/2021/02/intergenerationalrs. pdf* .

Jones, Maggie EC. 2023. "Post-secondary funding and the educational attainment of indigenous students." *Economics of Education Review* 97:102475.

Kazianga, Harounan, Dan Levy, Leigh L Linden and Matt Sloan. 2013. "The effects of" girl-friendly" schools: Evidence from the BRIGHT school construction program in Burkina Faso." *American Economic Journal: Applied Economics* 5(3):41–62.

Kremer, Michael, Edward Miguel and Rebecca Thornton. 2009. "Incentives to learn." *The Review of Economics and Statistics* 91(3):437–456.

Magesan, Arvind and Eik Leong Swee. 2018. "Out of the ashes, into the fire: The consequences of US weapons sales for political violence." *European economic review* 107:133–156.

Meriam, Lewis. 1971. *The problem of Indian administration*. Number 17 Johnson Reprint Corporation.

Milloy, John S. 2017. *A national crime: The Canadian government and the residential school system*. Vol. 11 Univ. of Manitoba Press.

Ministry of Tribal Affairs. 2025. "National Education Society for Tribal Students (NESTS)

— About.". Mandate and role of NESTS in establishing and operating EMRS. Accessed 2025-10-11.

URL: *https://nests.tribal.gov.in/*

National Family Health Survey (NFHS), India, 2019-21. 2021. "International Institute for Population Sciences (IIPS) and ICF.".

Nunn, Nathan and Nancy Qian. 2014. "US food aid and civil conflict." *American economic review* 104(6):1630–1666.

Oreopoulos, Philip, Marianne E Page and Ann Huff Stevens. 2006. "The intergenerational effects of compulsory schooling." *Journal of Labor Economics* 24(4):729–760.

Press Information Bureau, Government of India. 2020. "Government Schemes for Education of Scheduled Tribes; Literacy Statistics.". Census 2011 literacy: STs 59% vs overall 73%; PLFS updates noted. Accessed 2025-10-11.

URL: *https://www.pib.gov.in/PressReleasePage.aspx?PRID=1657743*

Press Information Bureau, Government of India. 2023. "Recurring Cost and Facilities for EMRS.". Up to INR 1.09 lakh per student per year; sanctioned/functional counts as of Mar 2023. Accessed 2025-10-11.

URL: *https://www.pib.gov.in/PressReleasePage.aspx?PRID=1906472*

Press Information Bureau, Government of India. 2025*a*. "EMRS Scheme — Establishment in Blocks with Majority ST Population.". Restates revamp criteria: >50% ST population and >= 20,000 ST persons (Census 2011). Accessed 2025-10-11.

URL: *https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=2117788*

Press Information Bureau, Government of India. 2025*b*. "Reforms in EMRS Operations — Sanctioned and Functional Schools.". Status as of 31-07-2025: 722 sanctioned, 485 functional. Accessed 2025-10-11.

URL: *https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=2159068*

Reyhner, Jon and Jeanne Eder. 2017. *American Indian education: A history.* University of Oklahoma Press.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235(2):2218–2244.

Subramanian, S. V. et al. 2023. "A comparative perspective, 2016–2021." *Proceedings of the National Academy of Sciences / PMC* . Cites Census 2011 figures: ST population >104 million across 705 groups. Accessed 2025-10-11.
URL: *https://pmc.ncbi.nlm.nih.gov/articles/PMC10794098/*

Szasz, Margaret Connell. 2006. "Through a wide-angle lens: Acquiring and maintaining power, position, and knowledge through boarding schools." *Boarding School Blues: Revisiting American Indian Educational Experiences* pp. 187–201.

Truth and Reconciliation Commission of Canada. 2015. *Honouring the Truth, Reconciling for the Future: Final Report of the Truth and Reconciliation Commission of Canada: Volume One: Summary.* James Lorimer Limited, Publishers.

Valencia Caicedo, Felipe. 2019. "The mission: Human capital transmission, economic persistence, and culture in South America." *The Quarterly Journal of Economics* 134(1):507–556.

Xaxa, Virginius, Usha Ramanathan, Joseph Bara et al. 2014. Report of the High Level Committee on Socio-Economic, Health and Educational Status of Tribal Communities of India. Technical report Ministry of Tribal Affairs, Government of India. Government-commissioned assessment of structural drivers of deprivation among STs.

Appendix

Table A1(a): Education: Eligible ST vs Eligible Non-ST

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × School	-1.241**	-1.134**	-1.253**	-1.280**	-1.120**	**-1.199*** **
	(0.5860)	(0.0623)	(0.5419)	(0.5526)	(0.4627)	(0.4489)
Observations	2,942	2,942	2,941	2,941	2,941	2,941
R-squared	0.0547	0.1154	0.1675	0.2891	0.3338	0.4175
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and years of education. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table A1(b): Education: Eligible ST vs Eligible Non-ST

Dependent Variable:	**Completing Secondary Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × School	-0.0322	-0.0321	-0.0448	-0.0711	-0.0563	**-0.0785***
	(0.0454)	(0.0433)	(0.0396)	(0.0486)	(0.0366)	(0.0448)
Observations	2,941	2,941	2,941	2,941	2,941	2,941
R-squared	0.0214	0.0435	0.0886	0.1764	0.1965	0.2670
Year FE	-	✓	-	-	-	-
State FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is an indicator which equals 1 if the individual completes secondary education. Estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A2(a): Education: ST vs Non-ST (Villages with a school)

Dependent Variable:	**Years of Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × School	0.6466	0.4017	-0.3521	-0.5625	-0.9675**	**-1.424*** **
	(0.4940)	(0.4885)	(0.4817)	(0.5382)	(0.4601)	(0.5181)
Observations	1,926	1,926	1,913	1,913	1,350	1,350
R-squared	0.1393	0.2124	0.3113	0.4767	0.4972	0.5947
State FE	-	✓	-	-	-	-
Year FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is years of education. The estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and years of education. Standard errors are clustered at the village level and shown in brackets. *p<.10; **p<.05; ***p<.01.

Table A2(b): Completing Secondary Education: Eligible ST vs Eligible Non-ST

Dependent Variable:	**Completing Secondary Education**					
	(1)	(2)	(3)	(4)	(5)	(6)
ST × Eligible	-0.0414	-0.0345	-0.0517	-0.0500	-0.0633	-0.0726
	(0.0225)	(0.0404)	(0.0420)	(0.0451)	(0.0410)	(0.0461)
Observations	1,826	1,826	1,913	1,913	1,350	1,350
R-squared	0.0404	0.0717	0.1657	0.3236	0.2600	0.3494
Year FE	-	✓	-	-	-	-
State FE	-	✓	-	-	-	-
State-Year FE	-	-	✓	✓	✓	✓
District FE	-	-	-	✓	-	✓
Controls	-	-	-	-	✓	✓

Notes: The dependent variable is an indicator which equals 1 if the individual completes secondary education. Estimation controls for state-year and district-fixed effects. Controls include religion, marital status, gender of the household head, and income status. Standard errors are clustered at the village level and shown in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.