ISSN No. 2454 - 1427

CDE
December 2025

# The persistent effects of compulsory education in Baroda, 1901-2011

**Hemanshu Kumar**
Email: hemanshu.econ@gmail.com

**Meeta Kumar**
Email: meetakumar@econdse.org
Delhi School of Economics,
University of Delhi

**Rohini Somanathan**
Email: rohini@econdse.org
Delhi School of Economics,
University of Delhi

**Working Paper No. 356**

Centre for Development Economics
Delhi School of Economics
Delhi- 110007

# The persistent effects of compulsory education in Baroda, 1901-2011

Hemanshu Kumar* Meeta Kumar† Rohini Somanathan‡

December 19, 2025

**Abstract**

Literacy was extremely low in colonial India - by 1931, average gross literacy was about 8%. In comparison, the princely state of Baroda stood out by achieving an average literacy rate close to 18% in the same year. The ruler of Baroda introduced a set of policies in 1906 that included compulsory education and public provision of free, primary schools. We examine the short and long-run effects of this set of policies. We do this through a comparison of areas within Baroda with regions bordering them, using a difference-in-difference framework. Since administrative boundaries changed dramatically over this period, our long-run comparisons rely on a careful mapping of boundaries. We find large effects through the colonial period and in the decades immediately following independence. These differences eventually narrowed as public good provision expanded. In 2011, sixty-four years after independence, there still remained a gap in literacy rates in areas that were historically in Baroda, and those that were outside it.

**Keywords**: Literacy, persistence, education policy, compulsory education, colonial India, princely states, Baroda.

**JEL Classification**: I21, I28, N35

We thank Harigovind S for invaluable research assistance.

*Independent researcher, hemanshu.econ@gmail.com.

†Delhi School of Economics,meetakumar@econdse.org, corresponding author.

‡Delhi School of Economics, rohini@econdse.org.

# 1 Introduction

Can leadership alter the course of history? Colonial India provides scope to study this because nearly 40% of the country's territory was under 'princely states.'[1] These states had substantial autonomy over internal policy. They were also all autocratic, their rulers had minimal accountability to their subjects.[2] They operated within a common macro-economic setting determined by British Indian policies. On the other hand, they varied widely in their size, resource endowments, history and geographic location on the subcontinent. (M. Kumar and Somanathan, n.d.).

Almost 350 of these 550 odd princely states were located in western India. Multiple jurisdictions not just co-existed, they were interspersed with each other. This allows us to use policy variation *across* jurisdictions to establish causality between policies and development outcomes to over the course of the twentieth century.

Baroda was among the largest, most prestigious, princely states in western India. The ruler of Baroda, Sayaji Rao Gaekwad III (1881-1939), introduced a unique set of education policies in his state early in the twentieth century, far more aggressive than anything adopted in British India, or in other princely states of the region. These include a universally applicable Compulsory Education Act (CEA) promulgated in 1906, public provision of free primary education, and a set of special schools (Antyaj schools) set up for children of the so called 'depressed' castes. (ibid.) These policies are all the more remarkable because Baroda remained an autocratic state until its merger with the Indian Union in 1949. Baroda was also dispersed in several non-contiguous tracts across western India, for completely random, historical reasons. It shared boundaries with six districts of British India, and seventeen princely states, that were a part of the Bombay Presidency. The history and geography of Baroda thus provide us with a near-ideal natural experiment that permits us to evaluate the effect of its remarkable education policies through a comparison with contiguous areas that were administered by other

[1] The princely states were erstwhile kingdoms, dispersed across the Indian subcontinent, that had not been annexed into British India, but existed as indirectly administered 'feudatory states'. They occupied over 40% of colonial India's geographical area, and were home to more than a quarter of its population.

[2] The political arrangement that followed from the Crown takeover of the East India Company's empire in 1858 assured the *states'* continuance – however, *rulers* of the states were subject to Imperial approval. Rulers could be, and were deposed. 21 such cases occurred between 1857 and 1947, many more were threatened.(Iyer, 2010)

princely rulers or by British India.

We proceed as follows: we situate our paper in the literature on persistence, and on leadership in development. We also explain why we use literacy as a proxy for development in princely states (section 2). Section 3 describes the data we use. Baroda and its neighbours were dispersed over the present day Indian state of Gujarat. The very interspersing of jurisdictions that permits the case of Baroda as a natural experiment, generates a challenge. All administrative boundaries in this region underwent extensive changes post-Independence. We use multiple sources to track these changes over the long period 1901-2011 to arrive at sub-district level administrative units with consistent boundaries for each year. Following H. Kumar and Somanathan (2017), we define a 'consistent unit' as one where all boundary changes occurred within the unit for every year between 1901 and 2011, and not across units in any year. We then use sub-district level data from the Census of India to generate a new data-set for literacy rates for these units for the entire century. Next, we apportion these units to three pre-Independence regions: the princely state of Baroda, British India, and other princely states (OPS). In section 4, we outline a difference-in-difference model to estimate DiD margins on literacy rates between Baroda, British India and OPS. We discuss why these margins may be ascribed to Baroda's remarkable education initiative in the early twentieth century. The results from our estimation are presented and analyzed in section 5. Section 6 concludes.

# 2 Motivation and literature review

A large literature addresses the impact of history on development – Acemoglu and Robinson (2013) and Nunn (2014, 2020) contain exhaustive surveys of this literature. Early studies on persistence explored long term effects of specific institutions on economic outcomes (Acemoglu, Johnson, et al., 2001; Banerjee, Iyer, et al., 2005; Dell, 2010; Iyer, 2010). They also examined the impact of history on culture as both, a coordination device, and as a source of social capital (Greif, 1994; Nunn and Wantchekon, 2011). But neither institutions nor culture are immutable. A big question is: do they change according to pre-determined laws – for example, those of historical materialism – or is it possible to sway the course of history? Banerjee and Duflo (2014) emphasize the contingent nature of history

itself, and the consequent scope for policy-induced changes. In a world where multiple equilibria are possible, persistence may result simply from a coordination failure. This gives rise to the possibility that the course of history might be changed *via* multiple channels: through leadership (discussed below), through a change in formal rules (Beaman et al., 2009; Chattopadhyay and Duflo, 2004), or simply a cascading effect of a small, random event (Dell, 2012; Hornbeck and Naidu, 2014). Regardless of whether we consider history deterministic or contingent, it helps understand change, and gives insights into the genesis, as well as impact, of policies that are transformative. (Nunn, 2020, p. 367)

Jones and Olken (2005) use leadership transitions to show that it matters in economic growth. Besley et al. (2011) find a positive correlation between leader education and economic performance of nations. On the face of it, precisely what leadership means is the subject of some debate: clearly, it needs to be separated from '*office*', from a titular position atop an hierarchy and from popularity. Ahlquist and Levi (2011) identify a set of characteristics that define a leader, based on implicit consensus in the literature. A leader is defined relative to a set of followers, among whom the leader's identity is common knowledge. An essential asymmetry marks the relation between the leader and her followers. Most importantly to us, the leader is instrumental in affecting change. In the presence of multiple equilibria, a leader could influence the final outcome by effecting her followers' beliefs or preferences *via* example, persuasion or coercion. A game-theoretic formulation that explores this idea is Acemoglu and Jackson (2015). In the model they develop, leaders can trigger a change in social norms: the leader's strategy choice is clearly visible to followers, and credible. Followers take their cue from leaders in choosing their own strategy, and thus leadership has the potential to overcome historic lock-in and cause a 'switch' from paths of low growth. Ahlquist and Levi also point out that, typically, principal-agent or social exchange models underlie discussions on leadership – but this tends to downplay the coercive power that leaders acquire.

An important concern in political economy is whether leader preferences shape public policy or whether leaders simply respond to their constituents. As mentioned earlier, the diversity of princely states, and the fact that they were geographically closely interspersed may be leveraged to gain insight into the importance of leadership for development. The colonial structure sanctioned the autocracy of these princes. However, in the context of Princely states, we need to

be careful to not conflate leadership with the ruler. Leadership could vest with a powerful Diwan (minister); or a Resident.[3] An example of rulers taking a backseat in administration is from Hyderabad, the largest and richest of Princely states. Since the early nineteenth century, affairs of state in Hyderabad were controlled by "powerful diwans... through alliances with domineering residents" Ramusack (2004, pp. 62–63)[4] In Travancore, (a large state in south India that comprised more than half of present day Kerala), T. Madhava Rao, the Diwan from 1858 to 1872, was acknowledged as the architect of a 'model state' rather than the ruler.

The leadership roles of Princes, Diwans and Residents may be collated from a variety of official and biographical sources - much gets recorded about men in power. Development in the princely states is far harder to track: data is sparse and scattered. Sivasubramonian (2000, pp. 307–312) discusses the challenge in estimating the contribution of the government sector to the national income for the princely states. Other sectors are even more poorly documented.

Caruana-Galizia (2013) uses the data from *Prices and Wages in India*, published annually between 1884 and 1924, to decompose national level estimates from Sivasubramonian (2000) and Heston (1983) to sub-national levels. These estimates allow very crude comparisons across some states, and for a short span of time.[5] The Census of India, however, documents reasonably consistent literacy data virtually since its inception in 1872. We know that literacy is strongly positively correlated

[3]This was the representative of the British Government in-charge of overseeing the state administration's compliance with Imperial rule

[4]See also Jeffrey (1975, 1977) on Travancore.

[5]Growth rates of sub-national GDP in Caruana-Galizia (2013) are also very sensitive to the choice of starting date. The rate of growth in per capita income (PCI) for 1875-1911 is negative for two large princely states, Baroda and Hyderabad, as also for Bombay and Bengal in British India. This is contrary to most qualitative evidence. A closer look at the numbers reveals that this is because of the unusually high value of the PCI recorded for 1875 across regions. In particular, Baroda shows more than 40% decline in PCI between 1875 and 1881. 1876-78 were years of severe famine in south and south-western India, so the fall in PCI in most places is expected. However, the sharp fall in Baroda seems anomalous. Official sources describe Baroda's experience as a 'scarcity' rather than famine. The Baroda Administration Report, 1878-79 states that 'Happily...the ordeal...proved much less severe than anticipated' Row (1879, p. 208). We calculate the compound rate of growth for the period 1881-1931 to find that Baroda was among the fastest growing regions in the country at the turn of the century - it grew twice as fast as Bombay. Mysore and Madras grew at rates comparable to Baroda. PCI growth in Travancore, a large princely state in present day Kerala, trailed Madras and Mysore – but this was probably because it experienced very high rates of growth of population. Hyderabad had negative growth rates, regardless of the starting date we use (notwithstanding its ruler's fabled personal wealth!).

with development, and can serve as a good proxy. Moreover, literacy appears to be the only indicator of development for which we have officially recorded, long term data for the princely states. Therefore, we focus on literacy attainments.

Literacy gains post-1870 were universally driven by increases in public spending on education, especially primary education. Public expenditure on education, in turn, increased as franchise was widened. (Chaudhary and Lindert, 2021). On the other hand, literacy is known to contribute to improved social and political participation, generating substantial positive externalities. (J. Cameron and S. Cameron, 2006) At first glance, colonial India appears to fit this typology, because the government was quite obviously *not* representative and literacy gains were very low. Between 1870 and 1931, census data suggests that gross literacy rates doubled from 4% on average to about 8%. However, even during this period, three regions in the country were clear outliers in literacy attainments: Cochin and Travancore, which are part of present-day Kerala; and Baroda which is part of present-day Gujarat. All three were princely states. The colonial antecedents of high levels of human development in Kerala have been discussed in Desai (2005, 2007). We focus on Baroda and its neighbouring areas in western India.

Fortuitously for us, Baroda is far better documented than most other states in the region. Sayaji Rao Gaekwad III ruled Baroda from 1881 to 1939. His leadership role, is historically documented. (See for example, Bhagavan (2001) and Mehta (1993)). Education of the masses, especially female education, as well as eradication of caste discrimination had several advocates in early twentieth century India, but few had the power associated with the Maharaja of Baroda. We investigate if the policies introduced by him at the beginning of the 20th century continued to yield dividend a hundred years later.

The existing literature on colonial policy in India is ambivalent about its persistent effects. On the one hand, Banerjee, Iyer, et al. (2005) show that property rights regimes introduced by the British in India did have a long-term effect. Areas where proprietary rights were granted to landlords (*zamindari* areas), had lower investments in land and agriculture, lower agricultural productivity, as well as lower public investment in health and education post-independence. They ascribe these differences the fact that zamindari areas were unable to take advantage of opportunities that emerged post the 1960s for political economy reasons. When read together with Banerjee and Somanathan (2007) (who find convergence in public

good provision between 1971 and 1991 across the country), it would appear that the persistence is indirect – land tenure systems generated an antagonistic class structure (historic elite vs masses). This class structure had long term political implications – it ultimately generated variations in public spending and public good allocation. Iyer (2010), on the other hand, finds that although princely states spent more on public good provision than the British did in 'directly ruled' areas, differences in public good availability reduced post-Independence, as specific policies were adopted to equalize access. In the same vein, Chaudhary and Garg (2015) study the long term impact of variations in education expenditure in British India on literacy rates. They find the impact of these variations persisted till the 1960s; literacy rates began to converge after the introduction of the New Education Policy, 1968.

Based on this, we expect literacy to be higher in the 'Baroda areas' of present-day Gujarat than in the other parts of the state. However, we also expect convergence by the end of our period. Our analysis reveals that while the first is true, convergence in literacy rates has been slow to come about. We do, however find a convergence in availability of schools across the Baroda and non-Baroda areas of the state. We describe the data we use to establish this in section 3.

# 3 Data

Present day Gujarat corresponds to the Gujarat division of the Bombay Presidency (including the princely states in this region), Baroda state, (and for 1931, the Western Indian States Agency).

As mentioned earlier, approximately 350 of the 550-odd princely states of India were located in the Bombay Presidency, most of them in the Gujarat region. These states were merged into the newly created Indian state of Bombay, Saurashtra and Kutch. between 1948 and 1950. Baroda, dispersed in multiple tracts across Gujarat, was incorporated into Bombay, Saurashtra and Kutch state in 1949. Consequently, administrative boundaries were extensively redrawn.[6]

[6]According to the 1951 Census of India, only one district in the Bombay, Saurashtra & Kutch state – Kanara – remained unchanged from 1941 after this merger (Bowman, 1953, p. 8).

H. Kumar and Somanathan (2009, 2017) develop a methodology to track changes in district boundaries over the long term. Applying this methodology to Gujarat at the district level, as a first pass, reveals that no district in 2011 can be treated as 'pure Baroda' or 'pure British'. A sensible tracking of administrative units, particularly from the point of view of apportioning them between what was historically Baroda state and what was not, can only be carried out at the sub-district (*taluka*) level[7]. With the merger of so many princely states into the Indian union, many new talukas were created, and several talukas were moved across districts. A second round of very extensive boundary changes occurred in 1960, when the state of Bombay, Saurashtra and Kutch was reorganized to create Gujarat and Maharashtra. Administrative boundaries in Gujarat remained stable thereafter, till 1991. Between 1991 and 2011, further reorganization was undertaken.

The colonial census records consistent taluka-level literacy data from 1901, which we draw on. However, in order to use this we need to ensure that the unit for which we are tracking literacy remained unaltered from 1901 to 2011. In 1901, the Gujarat division of the Bombay presidency had five 'directly ruled' districts, and approximately 300 princely states[8]. In addition, the four 'divisions' (equivalent to district) of the Baroda state were dispersed in multiple tracts over this area. In 2011, the state of Gujarat had 225 talukas distributed over 26 districts. We extend the methodology in H. Kumar and Somanathan (2017) to the taluka level to construct the smallest administrative unit (a combination of talukas) whose boundaries remained unchanged between 1901 and 2011. This is detailed in Appendix A.

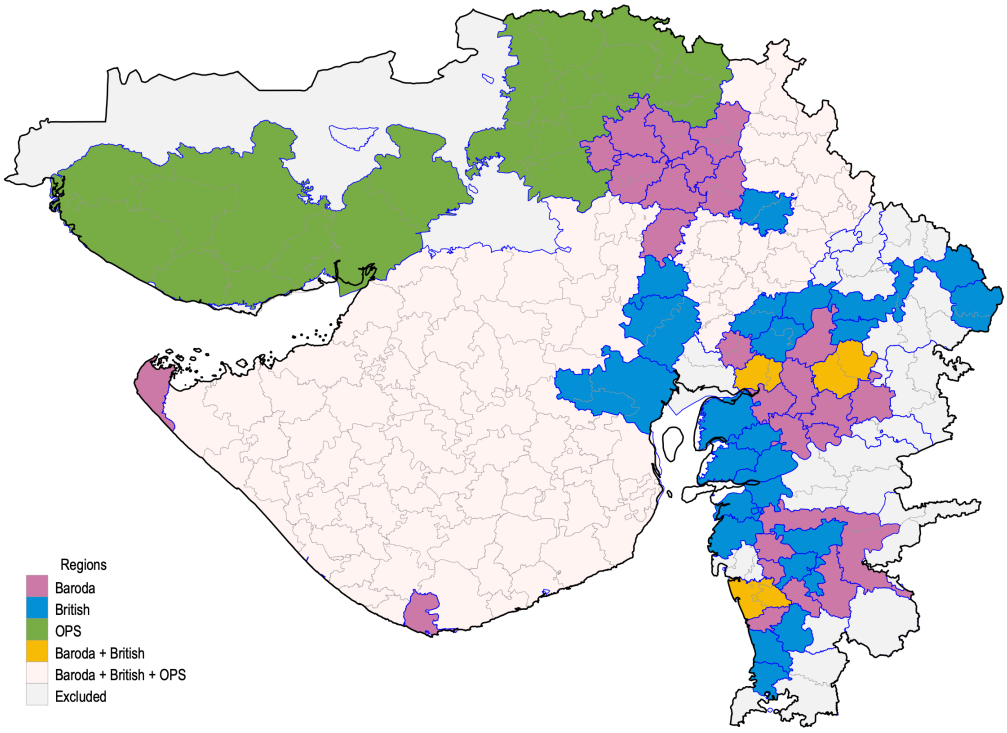
The number of consistent units we identify varies with the size of jurisdictional changes we decide to tolerate as insignificant. At a 10% laxity level,[9] we identify 58 administrative units, consistent between 1901 and 2011, that lie completely within

[7]The colonial administration followed a three-level hierarchy below the Imperial government – province, district, sub-district. Larger princely states, including Baroda, adopted this administrative structure. Most princely states, however, were too small to require this subdivision. Post-Independence Indian states continued the state–district–taluka structure inherited from the colonial period. The regional label for sub-district was 'taluka'.

[8]In 1901, the Gujarat division comprised British districts of Ahmedabad, Broach, Kaira, Panchmahals and Surat. The princely states that were part of this division were Cambay, Cutch and all those in the agencies of Kathiawar, Mahi Kantha, Palanpur, Reva Kantha and Surat Agency.

[9]This implies that we accept as negligible an interchange of areas with upto 10% of the population of the talukas involved.

Figure 1: Gujarat: Administrative Boundaries, 1901 and 2011




present day Gujarat. Four talukas interchanged areas with Maharashtra. We exclude these from our frame for reasons we explain below. In 1901, unfortunately, literacy data is available only for the three *agencies* of Kathiawar, Mahi Kantha and Palanpur, and not for individual princely states. We combine all talukas that contain any part of a princely state belonging to these agencies into a single 'consistent unit'. We are also constrained to exclude from our frame of analysis those talukas that contain parts of states for which we do not have data in 1901[10].

We apportion the consistent units thus created to their historic region - Baroda, British, OPS or mixed. Excluding the seven 'mixed' units from our panel leaves us with a final count of 51 units for each year. Figure 1 depicts these.

Post Independence, the Primary Census Abstract (PCA) lists the number of literates down to the level of the village[11]. Digital versions of the Primary Census

[10]These include Cambay and states belonging to the Reva Kantha and Surat Agencies

[11]The 1951 census records a discrepancy between the number of literates reported for districts

Abstract are available only for the years 1991, 2001 and 2011. We have digitized data from the pre-1991 censuses, going back to 1901, thus creating a new data set on gross literacy data in Gujarat for the long period 1901 to 2011 at the taluka level.

Literacy was defined as 'the ability to read and write a simple letter' in 1901, and this definition remained unchanged thereafter.[12] From 1931, the census began to exclude the age-group 0-5 years from the number of literates in British India. Baroda, in keeping with its CEA, excluded the age-group 0-6 years from the number of literates. Unfortunately, age composition of the population is reported only at the district level in the colonial census, not at the taluka level. Therefore, we use 'gross literacy rates'- we deflate the number of literates for each consistent unit by the total population of that unit. This implicitly assumes that literacy among children aged 0-5 years was negligible - not an unreasonable assumption to make.

Gross literacy rates for each region (comprising these consistent units) for the period 1901-2011 are in figures 2 to 4 on page 12. We observe that gross literacy in Baroda catches up with British India by 1911, and then remains consistently above, for the long period 1921-2011. Other princely states have consistently lower gross literacy than both Baroda and British areas, but the gap appears to narrow post 1981. This suggests persistence. The question is whether these differences are statistically significant. We discuss a methodology to test this in section 4.

in table D-VII of the census, and that reported for talukas in the PCA. The PCA numbers are based on village level information in 'National Registers' prepared post-census. The flyleaf to the PCA in each District Census Handbook (DCHB) suggests that the table D-VII numbers are more reliable. Accordingly, we scale the number of literates in each taluka as reported in the PCA, so the district total matches that in table D-VII.

[12]Earlier censuses counted as literate all those 'under instruction' and those not under instruction but able to read and write. This definition included as literate anyone who was enrolled in an educational institution, regardless of whether they could actually read and write or not. The change in definition in 1901 effectively excluded those that were enrolled, but not able to read and write. Enrollment rates were far too low in 1901 for this to have made a significant difference to literacy rates.

Figure 2: Gross Literacy, 1901-2011, Persons

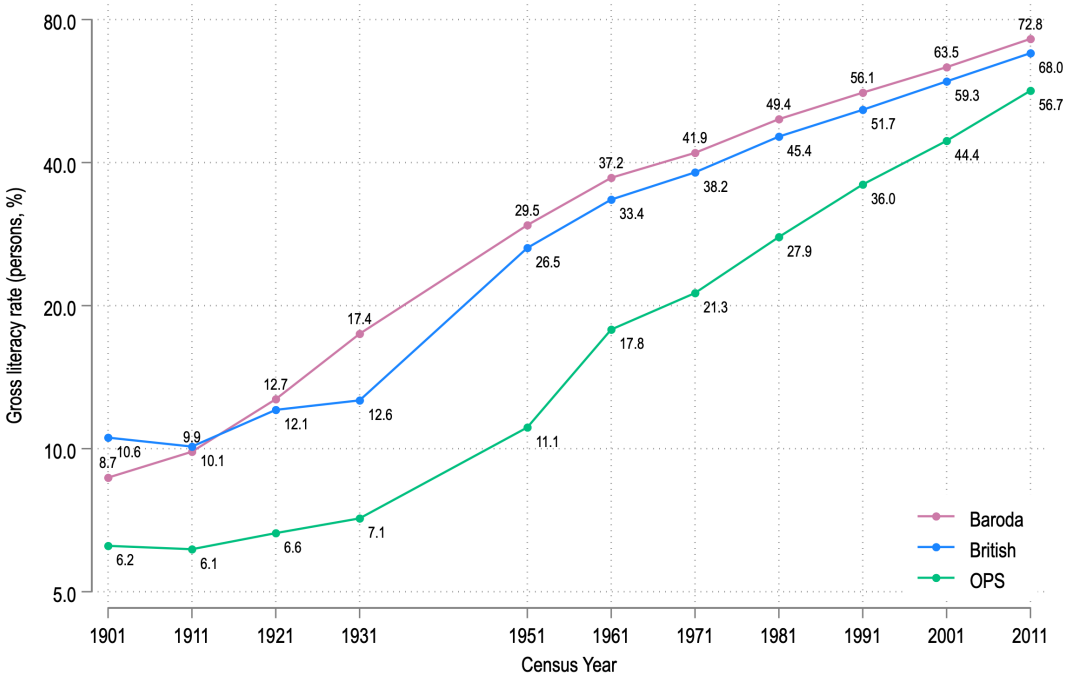



Figure 3: Gross Literacy, 1901-2011, Males

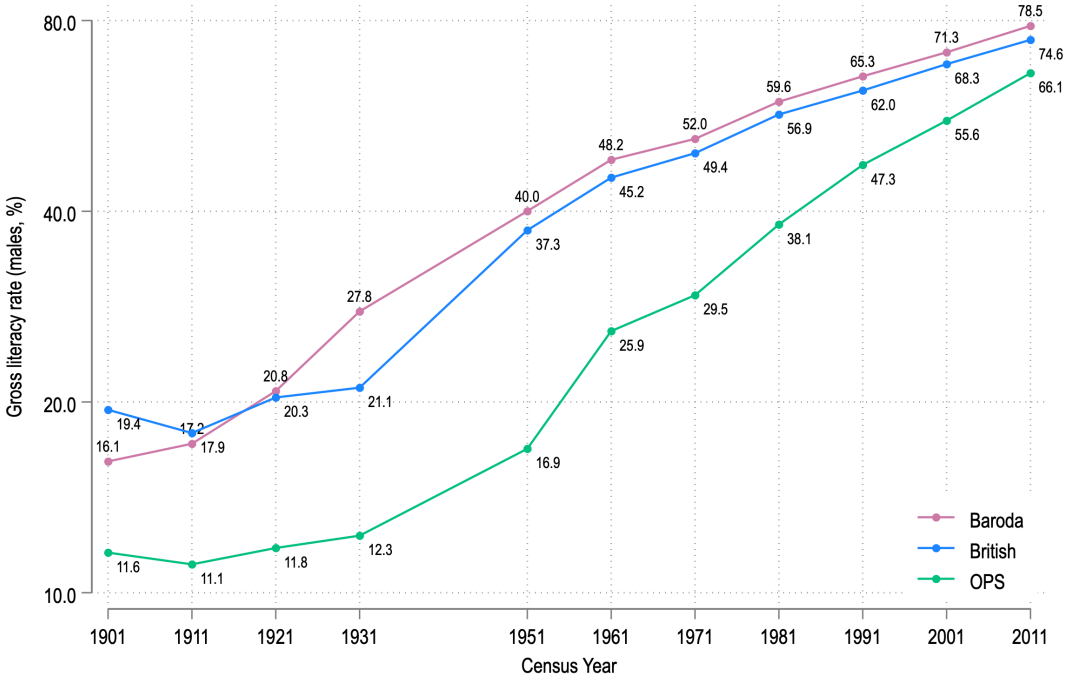
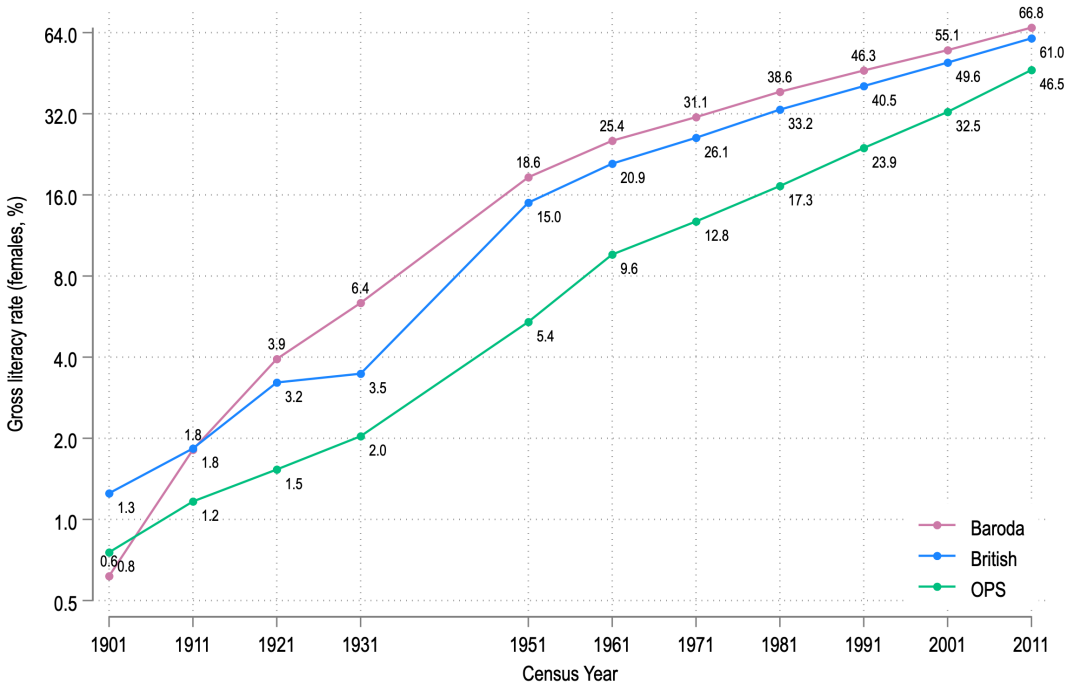


Figure 4: Gross Literacy, 1901-2011, Females




# 4 Methodology

In the colonial system, education was a provincial subject – funded mainly from a cess on land revenue.(Chaudhary, 2010). Princely states had the autonomy to organize their own school systems. It is this very heterogeneity in policy that we seek to exploit. Post-Independence, education became a 'state' subject[13]. Therefore, we restrict our analysis to present-day Gujarat. We do this because policy toward school education was very likely to vary across post-Independence state boundaries. Within a state, however, it is reasonable to assume that policy remained uniform. Our estimation strategy depends on calculating the difference-

[13]The Indian federation divided administrative heads into three lists: Union, State and Concurrent. At Independence, education was placed in the 'State' list of subjects, with the Centre retaining limited responsibility for higher education. Education was brought under the concurrent list in 1986, but school education remained largely the responsibility of States (Tilak, 1989, 2003)

in-difference in literacy rates for areas within Baroda (where progressive education policies were implemented), and the rest of Gujarat (where they were not). We include in our panel only those units that are 'pure' – that is, units that did not experience any (significant) intermixing with units in other historic regions over the period 1951-2011. We exclude units that contain areas from more than one region. As already mentioned, the exact number of units varies with the extent of inter-mixing of talukas we decide to tolerate: we choose a 10% laxity level. Our treatment group, comprises the units that are 'pure Baroda'; our control is the set of units that are 'pure BI' and 'pure OPS'. Table 4 on page 27 lists the number of units in each region.

The period we examine is 1901-2011. We exclude 1941, which was a war-time census and does not record literacy at sub-district levels for British India, or OPS. The 'treatment' occurred in 1906. Our 'pre-period' is therefore 1901. 1911-2011 comprises the post period. We divide the long 'post-period' 1911-2011 into three sub-periods: 1911-1931, 1951-1981 and 1991-2011, to better capture time fixed effects. 1911-1931, of course, is the colonial sub-period. A large expansion in primary school provision began nation-wide after 1971, following the New Education Policy (1968). Banerjee and Somanathan (2007) show that provision of public goods was beginning to converge across parliamentary constituencies by 1991. We expect that policies to equalize access to amenities had weaker impact till 1981. Therefore, we divide the post-independence period into two sub periods: 1951-1981 and 1991-2011.

We estimate the following equation:

$$L_{it} = \beta_0 + \sum_{t=1}^{T} \tau_t \mathbb{1}_t + \gamma_1 \mathbb{1}_{bar} + \gamma_2 \mathbb{1}_{ops} + \sum_{t=2}^{T} \delta_t \mathbb{1}_{bar} \mathbb{1}_t + \sum_{t=2}^{T} \omega_t \mathbb{1}_{ops} \mathbb{1}_t + \epsilon_{it} \quad (1)$$

$L_{it}$ is the gross literacy rate - the number of literates divided by the total population. $\beta_0$ is a constant. $\mathbb{1}_{bar}$ is an indicator variable which takes the value 1 if a unit is in erstwhile Baroda, 0 if not. Similarly, $\mathbb{1}_{ops}$ is an indicator variable for whether or not a unit was a princely state other than Baroda (OPS). $\mathbb{1}_{bar}$ captures baseline differences in gross literacy rates between Baroda and British India (BI), as $\mathbb{1}_{ops}$ does between OPS and BI. The size of our data set does not permit us to have a dummy specific to each of our 51 units.

$\mathbb{1}_t$ is an indicator for the $t$th sub-period. The coefficients of interest are $\delta_t$ which indicate the difference-in-difference in gross literacy rates between Baroda and

BI for the period $t$. Analogously, $\omega_t$ captures the difference-in-difference between OPS and BI for the period $t$.

How robust is this methodology? The non-availability of taluka-level literacy data before 1901 is a problem - parallel pre-trends are impossible to demonstrate. However, we do have district level data, which we can use to track literacy changes in each of our three regions. For British districts and Baroda, we can tack these changes all the way back to 1872.

For the OPS in 1872, literacy data is either missing (Mahi Kantha and Palanpur) or obviously anomalous (Kathiawar). A large state in Kathiawar (Junagarh) reports an improbably high male literacy of 47%, while female literacy is not atypical. We have found no qualitative evidence regarding any extra-ordinary literacy attainments for Junagarh. Literacy dips sharply for Kathiawar between 1872 and 1881. Excluding Junagarh from Kathiawar in 1872 yields a literacy rate more in line with the rest of the region, and with subsequent years. Unfortunately, subsequent censuses do not contain state-level literacy rates, so we cannot correct this anomaly by excluding Junagarh from our frame for the entire period 1872-1901. Therefore, for the OPS we look at literacy over 1881-1901. As shown in figures 5 to 7 on page 16, the pre-trends are remarkably parallel.

More importantly, the district is the basic unit for implementation of state policy in India, and we have discussed how each district in Gujarat comprised areas from all three historic regions. We have no reason to expect that post-Independence policy implementation would vary systematically with historic region. This gives us confidence that our exercise is valid.

Figure 5: Gross Literacy, 1872-1901, Persons

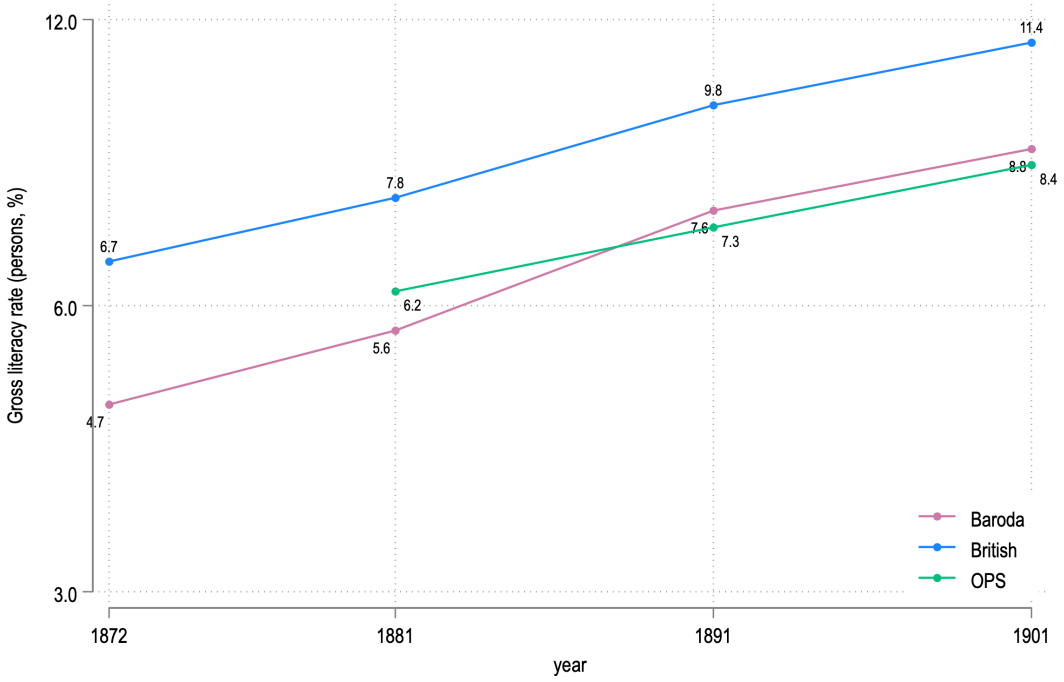



Figure 6: Gross Literacy, 1872-1901, Males

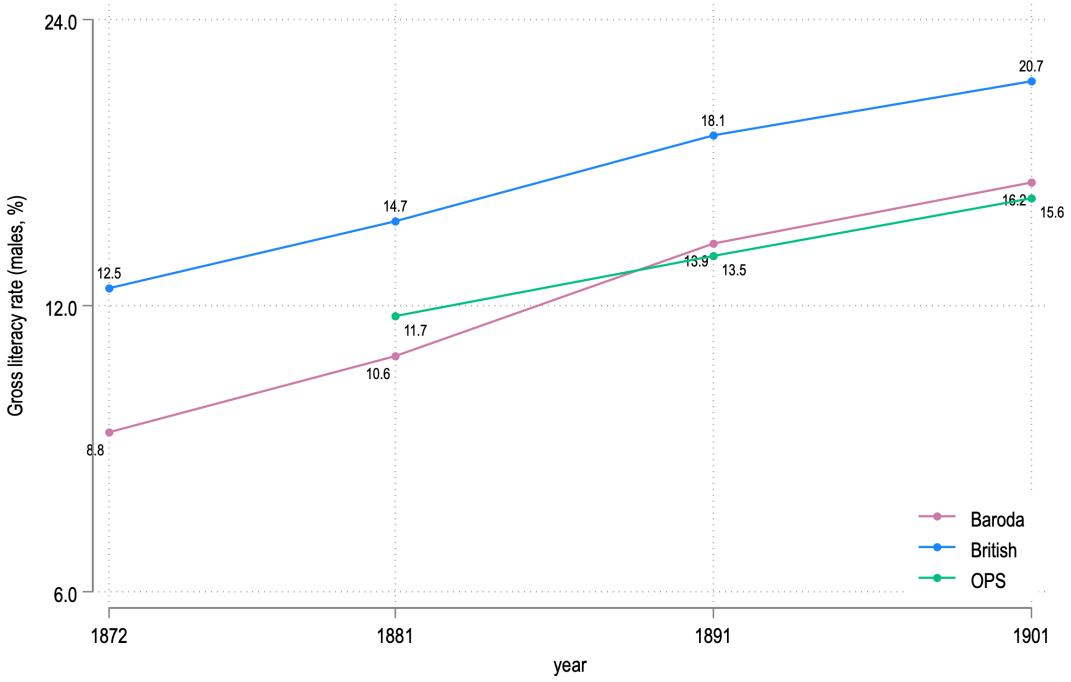
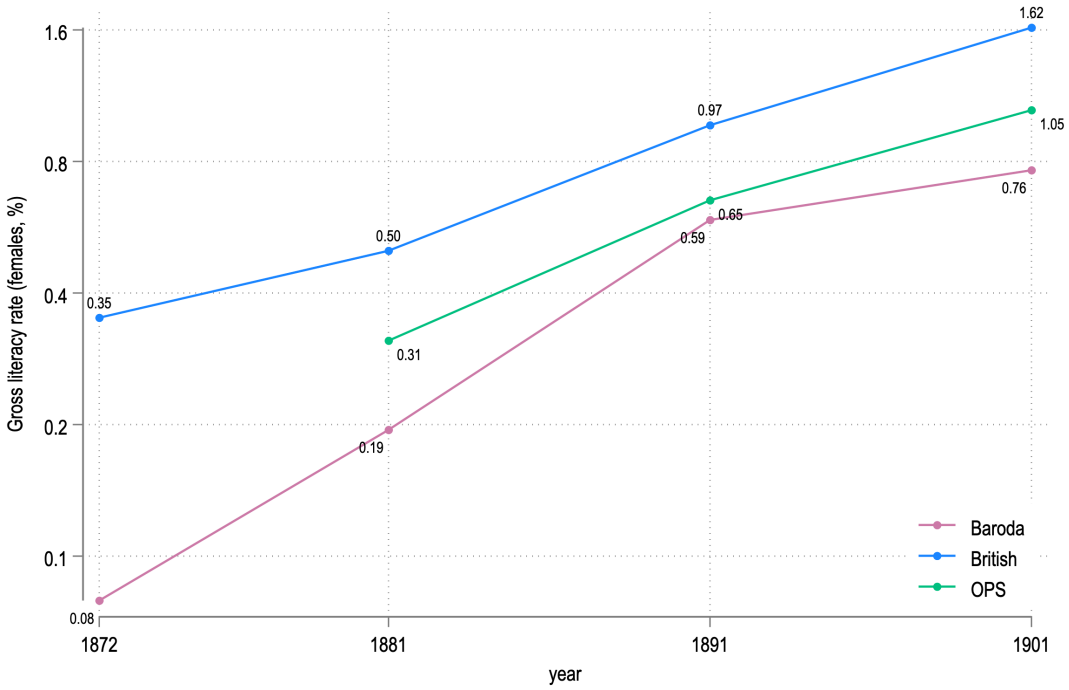


Figure 7: Gross Literacy, 1872-1901, Females




# 5 Results

Table 1 on the next page shows the results from our estimation. Our results show strong and persistent treatment effects for Baroda. The DiD estimate for all persons in Baroda for the period 1911-1931 is significant at the 5% level, suggesting that literacy was higher in Baroda by 3.4% as a result of the CEA and its associated policies. This is two-thirds of the literacy differential between British districts and Baroda in 1931. The effect on female literacy was somewhat smaller. The DiD coefficient, significant at the 1% level, suggests that more than half the observed difference in female literacy between Baroda and British districts may be attributed to this set of policies.

The DiD coefficients for the period 1951-1981 continue to be positive and significant at the 5% level for all persons (and for females as well). What is more, these coefficients are not significantly different from those estimated for the previous period, suggesting the persistence of treatment effects into this period as well. We observe a weakening of these effects in 1991-2011. The DiD coefficient is significant only at the 10% level for all persons, and not significant for females.

How do our results square with the evidence that provision of public goods has expanded substantially since the 1990s? We examine the availability of various amenities across the three regions in our frame, based on data available in the Village Directory, Census of India, 2011.

We find that availability of most village amenities, and schools in particular, had indeed converged across the three historic regions by that year. Clearly, Gujarat was no exception to the national trend discussed in H. Kumar and Somanathan (n.d.). Our exercise suggests that literacy levels are slower to converge than school availability, though we do see a tendency in that direction by 2011.

Table 1: Difference-in-differences estimation for gross literacy

| | persons | males | females |
|---|---|---|---|
| Baroda | -2.089* | -3.850* | -0.596*** |
| | (-1.87) | (-1.88) | (-3.00) |
| Other Princely States | -4.185*** | -7.847*** | -0.431 |
| | (-2.68) | (-2.71) | (-1.30) |
| 1911-1931 | 1.117 | 0.504 | 1.499*** |
| | (1.16) | (0.29) | (7.49) |
| 1951-1981 | 25.04*** | 27.38*** | 22.35*** |
| | (18.34) | (14.19) | (21.17) |
| 1991-2011 | 49.17*** | 48.68*** | 49.31*** |
| | (31.20) | (25.44) | (30.96) |
| Baroda × 1911-1931 | 3.388** | 5.351** | 1.613*** |
| | (2.40) | (2.17) | (3.73) |
| Baroda × 1951-1981 | 4.217** | 5.131* | 3.648** |
| | (2.22) | (1.94) | (2.48) |
| Baroda × 1991-2011 | 3.551* | 5.055** | 2.331 |
| | (1.69) | (1.97) | (1.11) |
| Other Princely States × 1911-1931 | -0.641 | -0.211 | -0.649 |
| | (-0.34) | (-0.06) | (-1.18) |
| Other Princely States × 1951-1981 | -10.53*** | -10.10** | -10.56*** |
| | (-2.88) | (-2.15) | (-3.44) |
| Other Princely States × 1991-2011 | -8.573** | -3.148 | -14.17*** |
| | (-1.97) | (-0.70) | (-2.96) |
| Constant | 10.39*** | 19.34*** | 1.176*** |
| | (12.97) | (12.88) | (9.86) |
| Observations | 561 | 561 | 561 |

$t$-statistics in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 2: Availability of educational and health facilities, 2011

| | Region | | |
|---|---|---|---|
| | Baroda | British | OPS |
| Primary schools | 7.79 (2.82) | 7.90 (2.00) | 8.87 (2.87) |
| High schools | 1.25 (0.49) | 1.22 (0.31) | 0.98 (0.44) |
| Health centers | 2.46 (0.52) | 2.27 (0.72) | 1.99 (0.69) |
| Primary health sub-centres | 2.07 (0.48) | 1.92 (0.65) | 1.62 (0.61) |
| Primary health centres | 0.33 (0.07) | 0.32 (0.09) | 0.31 (0.12) |
| Community health centres | 0.06 (0.05) | 0.04 (0.04) | 0.07 (0.06) |
| Dispensaries | 0.09 (0.12) | 0.12 (0.28) | 0.13 (0.23) |
| Hospitals | 0.02 (0.07) | 0.03 (0.08) | 0.02 (0.08) |
| Maternity and child welfare centers | 0.29 (0.47) | 0.06 (0.13) | 0.25 (0.31) |
| Family planning centers | 0.03 (0.09) | 0.03 (0.06) | 0.04 (0.08) |
| No. of talukas | 33 (37.1%) | 32 (36.0%) | 24 (27.0%) |

*Source:* Census 2011 village directory. Amenities are reported as the number of facilities per 10,000 population. Standard deviations, in brackets, are computed over component talukas. Schools include both private and public. 'Baroda' refers to areas historically entirely in Baroda State, 'British' to areas that were entirely part of British territories, and 'OPS' to areas that were entirely in other princely states.

Table 3: Availability of village-level amenities, 2011

| | Region | | |
|---|---|---|---|
| | Baroda | British | OPS |
| Piped water | 95.9 (11.8) | 96.4 (10.3) | 98.8 (5.1) |
| Water tanks | 40.0 (20.7) | 56.3 (25.8) | 22.2 (22.2) |
| Wells | 52.6 (26.5) | 69.1 (26.2) | 37.2 (28.7) |
| Electricity | 100.0 (0.0) | 100.0 (0.0) | 100.0 (0.0) |
| Post offices | 13.7 (9.8) | 11.9 (10.1) | 6.9 (6.4) |
| Phone connections | 93.6 (13.8) | 94.1 (11.1) | 91.8 (9.6) |
| Paved roads | 98.0 (3.1) | 98.5 (2.3) | 99.0 (1.7) |
| No. of talukas | 33 (37.1%) | 32 (36.0%) | 24 (27.0%) |

*Source:* Census 2011 village directory. Availability is reported as the percentage of the rural population that lives in villages that have the facility. Standard deviations, in brackets, are computed over component talukas. 'Baroda' refers to areas historically entirely in Baroda State, 'British' to areas that were entirely part of British territories, and 'OPS' to areas that were entirely in other princely states.

# 6 Conclusion

We have shown that literacy rates in Baroda were below those observed in British India in 1901; in fact, for females it was even lower than that observed in other princely states. Literacy rose everywhere over the period 1901-2011, but it rose much faster in Baroda than in other parts of what is today Gujarat. By 1921, literacy in Baroda was higher than British India, and this difference appears to have persisted over the period 1951-2011.

We also find a convergence in public good availability across the three historic regions - Baroda, British districts and other princely states. In part, the persistent gap in literacy rates may be because we have examined average literacy over the long period. rather than literacy among a younger cohort. It may well be that literacy rates will converge with the passage of time. The experience of other princely states post 1981 appears to suggest this.

Our paper highlights the importance of compulsory education, and more generally, the long-run effects of progressive leadership.

# References


Acemoglu, Daron and Matthew O. Jackson (2015). "History, expectations, and leadership in the evolution of social norms". In: *The Review of Economic Studies* 82 (2) (2015), pp. 423–456.

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American Economic Review* 91 (5) (2001), pp. 1369–1401.

Acemoglu, Daron and James A. Robinson (2013). "Economics versus politics: Pitfalls of policy advice". In: *Journal of Economic Perspectives* 27 (2) (2013), pp. 173–192.

Ahlquist, John S. and Margarest Levi (2011). "Leadership: What it means, what it does, and what we want to know about it". In: *Annual Review of Political Science* 14 (1) (2011), pp. 1–24.

Banerjee, Abhijit V. and Esther Duflo (2014). "Under the thumb of history? Political institutions and the scope for action". In: *Annual Review of Economics* 6 (1) (2014), pp. 951–971.

Banerjee, Abhijit V., Lakshmi Iyer, and Rohini Somanathan (2005). "History, social divisions, and public goods in rural India". In: *Journal of the European Economic Association* 3 (2-3) (2005), pp. 639–647.

Banerjee, Abhijit V. and Rohini Somanathan (2007). "The political economy of public goods: Some evidence from India". In: *Journal of Development Economics* 82 (2) (2007), pp. 287–314.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova (2009). "Powerful women: does exposure reduce bias?" In: *The Quarterly journal of economics* 124 (4) (2009), pp. 1497–1540.

Besley, Timothy, Jose G. Montalvo, and Marta Reynal-Querol (2011). "Do educated leaders matter?" In: *The Economic Journal* 121 (554) (2011), F205–227.

Bhagavan, Manu (2001). "Demystifying the 'ideal progressive': Resistance through mimicked modernity in princely Baroda, 1900-1913". In: *Modern Asian Studies* 35 (2) (2001), pp. 385–409.

Bowman, J. B. (1953). *Census of India 1951, Vol. IV, Bombay, Saurashtra and Kutch, Part II-A*. 1953.

Cameron, John and Stuart Cameron (2006). "The economic benefits of increased literacy". In: *Background paper prepared for the Education for All Global Monitoring Report 2006 - Literacy for Life* (2006).

Caruana-Galizia, Paul (2013). "Indian regional income inequality: estimates of provincial GDP, 1875-1911". In: *Economic History of Developing Regions* 28 (1) (2013), pp. 1–27.

Chattopadhyay, Raghabendra and Esther Duflo (2004). "Women as policy makers: Evidence from a randomized policy experiment in India". In: *Econometrica* 72 (5) (2004), pp. 1409–1443.

Chaudhary, Latika (2010). "Land revenues, schools and literacy: A historical examination of public and private funding of education". In: *The Indian Economic & Social History Review* 47 (2) (2010), pp. 179–204.

Chaudhary, Latika and Manuj Garg (2015). "Does history matter? Colonial education investments in India". In: *The Economic History Review* 68 (3) (2015), pp. 937–961.

Chaudhary, Latika and Peter Lindert (2021). "Healthy, literate and smart: The global increase in human capital". In: *The Cambridge Economic History of the Modern World, Volume 2, 1870 to the Present*. Ed. by Stephen Broadberry and Kyoji Fukao. Cambridge University Press, Cambridge, 2021, pp. 329–355.

Cotton, James Sutherland, Sir Richard Burn, Sir William Stevenson Meyer, and Sir William Wilson Hunter, eds. (1908). *Imperial Gazetteer of India*. Clarendon Press, 1908.

Dell, Melissa (2010). "The persistent effects of Peru's mining mita". In: *Econometrica* 78 (6) (2010), pp. 1863–1903.

— (2012). "Path dependence in development: Evidence from the Mexican Revolution". In: *Harvard University, mimeograph* (2012).

Desai, Manali (2005). "Indirect British rule, state formation, and welfarism in Kerala, India, 1860–1957". In: *Social Science History* 29 (3) (2005), pp. 457–488.

— (2007). *State formation and radical democracy in India*. Taylor & Francis, 2007.

Greif, Avner (1994). "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies". In: *Journal of political economy* 102 (5) (1994), pp. 912–950.

Heston, A. (1983). "National Income". In: *The Cambridge Economic History of India, Volume II*. Ed. by Dharma Kumar and Meghnad Desai. Cambridge University Press, Cambridge, 1983, pp. 453–532.

Hornbeck, Richard and Suresh Naidu (2014). "When the levee breaks: black migration and economic development in the American South". In: *American Economic Review* 104 (3) (2014), pp. 963–990.

Iyer, Lakshmi (2010). "Direct versus indirect colonial rule in India: Long-term consequences". In: *The Review of Economics and Statistics* 92 (4) (2010), pp. 693–713.

Jeffrey, Robin (1975). "The politics of 'indirect rule': Types of relationship among rulers, ministers and residents in a 'native state'". In: *Journal of Commonwealth & Comparative Politics* 13 (3) (1975), pp. 261–281.

— (1977). "Bureaucratic politics in an imperial system: The rivalry between two Indian Dewans, 1936–1941". In: *Journal of Commonwealth & Comparative Politics* 15 (3) (1977), pp. 266–285.

Jones, Benjamin F. and Benjamin A. Olken (2005). "Do leaders matter? National leadership and growth since World War II". In: *The Quarterly Journal of Economics* 120 (3) (2005), pp. 835–864.

Kumar, Hemanshu and Rohini Somanathan (2009). "Mapping Indian districts across census years, 1971-2001". In: *Economic and Political Weekly* 44 (41) (October 2009), pp. 69–73.

— (2017). "Creating long panels using census data (1961-2001)". In: *Economic and Political Weekly* 52 (29) (July 2017), pp. 105–109.

— (n.d.). "Public goods". In: *Cambridge history of modern South Asia, Part II*. Ed. by Latika Chaudhary, Tirthankar Roy, and Anand Swamy. Cambridge University Press. Forthcoming.

Kumar, Meeta and Rohini Somanathan (n.d.). "Princely states in colonial India". In: *Cambridge history of modern South Asia, Part II*. Ed. by Latika Chaudhary, Tirthankar Roy, and Anand Swamy. Cambridge University Press. Forthcoming.

Mehta, Makrand (1993). "Princely ruler as an agent of change: A study of Maharaja Sayajirao Gaekwad's policies towards untouchability, 1882-1915". In: *Proceedings of the Indian History Congress*. Vol. 54. Indian History Congress. 1993, pp. 403–408.

Nunn, Nathan (2014). "Historical development". In: *Handbook of economic growth* 2 (2014), pp. 347–402.

— (2020). "The historical roots of economic development". In: *Science* 367 (6485) (2020), eaaz9986.

Nunn, Nathan and Leonard Wantchekon (2011). "The slave trade and the origins of mistrust in Africa". In: *American Economic Review* 101 (7) (2011), pp. 3221–3252.

Ramusack, Barbara N. (2004). *The Indian princes and their states*. Cambridge University Press, 2004.

Row, Raja Sir T. Madava (1879). *Report on the administration of the Baroda State, 1878-79*. Calcutta, 1879.

Sivasubramonian, S. (2000). *The national income of India in the twentieth century.* Oxford University Press, New Delhi, 2000.

Tilak, Jandhyala B. G. (1989). "Center-state relations in financing education in India". In: *Comparative Education Review* 33 (4) (1989), pp. 450–480.

— (2003). "Public expenditure on education in India: A review of trends and emerging issues". In: *Financing in India: Current Issues and Changing Perspectives* (2003), pp. 3–54.

# Appendix

## A Mapping sub-districts, Gujarat, 1901-2011

As mentioned in Section 3, administrative boundaries changed extensively between 1901 and 2011. We extend the methodology in H. Kumar and Somanathan (2017) to the taluka level to track these changes. We combine as a 'consistent unit' those talukas where transfer of population/area from a source taluka (parent) to receiving (child) taluka, between two consecutive census years, is significant. Whether or not a transfer is significant depends on the 'laxity level' we are willing to tolerate. We define a 'laxity level' as the level of individual transfers that we choose to ignore. For example, at a laxity of 10%, if the population transferred from a parent taluka to a child exceeds 10% of *either*, we merge the two into a single 'consitent unit'. We ignore any transfer from a parent to a child taluka that is less than 10% of the population of both. A consistent unit is one where all changes in area and population between 1901 and 2011 occur within the unit, and not across units. In order to construct these units, we need details on populations and/or areas interchanged between talukas for each census year.

Between 1961 and 1981, an appendix to the General Population Tables lists all transfers across talukas at the level of the village, reporting the transferred area, as well as the source and destination talukas. for this period, our laxity norms are difficult to apply in terms of area. Variations in areas of talukas took place for two reasons: either there was a transfer across taluka boundaries, or improved survey methods allowed a more precise reporting of areas. Therefore we apply the norm in terms of the size of population transferred. For each of these years, we combine information from this appendix with village level population data sourced from Village Directories in individual District Census Handbooks. This tells us very precisely, the size of the transfeered population. As explained above, if this transfer exceeds 10% of the population of either the 'parent' or the 'child' taluka, we combine the talukas into a single consistent unit, year by year. If the interchange is less than 10% for *both* talukas, we ignore it.

Unfortunately, the publication of this appendix is discontinued from 1991. For

computing the extent of interchange between talukas for the period 1991-2011, we use information on the area of each taluka, available in the various District Census Handbooks. By 1991, survey methodology had stabilized. We can reasonably attribute area changes to a transfer, 1991 onward. Once again, we ignore any interchange that was less than 10% of the area of both talukas involved. We combine talukas where the area exceeded 10% of the area of either the parent or child taluka. Our computation of the size of the transfer is still quite precise.

Tracking boundary changes between 1931 and 1951 is far more complex than the procedure described above. The 1951 flyleaf to Table II, 'Variations in Population for the last fifty Years' Census of India, volume IV, part II-A (Bombay, Saurashtra and Kutch) lists the princely states that were merged into each district, but does not detail these changes at the taluka level. We have used multiple documents to track the 1931 'parents' of each 1951 taluka. Talukas belonging to British districts in 1931 typically appear in one or the other 1951 district, and these have been the simplest to track.

Many talukas in the newly created/reconfigured district were princely states, or parts thereof, in 1931. Sometimes, they appear in 1951 with a name different form the parent state in 1931[14]. Saurashtra and Kutch regions comprise largely, but not exclusively, of these princely states. The colonial census does not list populations for most princely states individually - the standard practice being to list two or three larger states in each agency, and clubbing the rest as 'rest of the Agency'. Locating the hundreds of small states in 1931 and their destinations in 1951 has involved some sleuthing.

*The Imperial Gazetteer of India*, (Cotton et al., 1908), published in twenty six volumes, is an invaluable source of information regarding the location of many small princely states. We use this to locate 1931 parents of several 1951 talukas. Fortunately, two large states in Kathiawar - Bhavnagar and Nawanagar - have independent census volumes for 1931. We use these to trace divisions of these states that became talukas in 1951. In some cases, the District Census Handbooks (DCHB), 1951 provide information on states merged with specific talukas. Finally, Census Administrative Atlas of 2011 provides a map of Gujarat with 2011 taluka boundaries, which we overlay on the 1931 map of the region, for a rough-cut

[14]For example,Sankheda Mehwas from the Mahi Kantha agency in 1931, for example, becomes Naswadi taluka in Baroda district in 1951.

impression of boundary changes. Talukas are usually named after a prominent town within them, and we use the DCHBs, 1951, to confirm our finding from the map overlay. We use the same sources to identify the destination of talukas (and states) that existed in 1931, but not in 1951. We assume that any change in a taluka boundary that we are able to detect between 1931 and 1951 is significant, because we have neither population nor areas for most of the smaller princely states, or for talukas in most larger ones.

We return to the Census as a source of information on taluka changes before 1931. Boundaries of talukas in British areas mostly remained unchanged between 1901 and 1931[15] We assume, given the political set-up of the Empire, that boundaries of princely states remained unchanged as well, although this is not recorded. On the other hand, we have not come across accounts of changes in the boundaries of these states for the period 1901-1931[16]. Taluka level changes did occur in Baroda between 1901 and 1931. These are recorded in each census, although areas and populations of the tracts involved are not available. As earlier, we assume that any change we are able to detect are significant.

Figure 1 on page 9 shows the consistent units we construct using a 10% laxity level, based on taluka boundary changes culled from all the sources cited above. Naturally, changing the laxity level will change the composition of the consistent units. Table 4 shows the impact of varying laxity levels on the number of units in each region.

[15]The exception is the bifurcation of the Daskroi taluka in Ahmedabad district into North and South Daskroi in 1921.

[16]Baroda Administration Reports and the Census of Baroda report settlements of boundary disputes in earlier volumes.

Table 4: Region Typology by Laxity

| Laxity level | 0% | 1% | 5% | 10% | 15% | 20% |
|---|---|---|---|---|---|---|
| Baroda | 15 | 16 | 22 | 25 | 25 | 27 |
| British | 16 | 18 | 20 | 24 | 25 | 27 |
| OPS | 02 | 02 | 02 | 02 | 02 | 02 |
| Baroda+British | 03 | 03 | 03 | 03 | 03 | 02 |
| Baroda+OPS | 00 | 00 | 00 | 00 | 00 | 01 |
| British+OPS | 00 | 00 | 00 | 00 | 00 | 00 |
| All | 04 | 04 | 04 | 04 | 04 | 03 |
| Total | 40 | 43 | 51 | 58 | 59 | 62 |