LONG-RUN PROSPERITY
WORKING PAPER SERIES
HOOVER INSTITUTION

Did Railways Affect Literacy? Evidence from India

Latika Chaudhary & James Fenske

Working Paper 22006

March 2, 2022

DID RAILWAYS AFFECT LITERACY? EVIDENCE FROM INDIA

LATIKA CHAUDHARY* AND JAMES FENSKE†

ABSTRACT. We study the effect of railroads, the single largest public investment in colonial India, on human capital. Using district-level data on literacy, we find railroads had positive effects on literacy, in particular on male and English literacy. We employ two identification strategies. First, we exploit synthetic panel variation contained in cohort-specific literacy rates due to differences in the timing of railroad exposure of different cohorts within the same district and census year. We find a one standard deviation increase in railroad exposure raises literacy by 0.29 standard deviations. Second, we use distance from military cantonments and an early railway plan as instruments for district railway exposure in the cross section and find similar results. We show that railroads increased literacy by raising secondary and elite primary, rather than vernacular primary, schooling. Our mediation analysis suggests that non-agricultural income, urbanisation, and opportunities for skilled employment are important mechanisms, while agricultural income is not.
Keywords: Colonialism, Railways, Literacy.
JEL Codes: N75, N35, R40

*Associate Professor, Graduate School of Defense Management, Naval Postgraduate School
†Professor, Department of Economics, University of Warwick
E-mail addresses: lhartman@nps.edu, J.Fenske@warwick.ac.uk.
Date: March 2, 2022.
We are grateful to audiences at Ashoka University, the Economic History Association Meetings, Heriot-Watt University, the Indian Statistical Institute Delhi, New York University, the Royal Economic Society Meetings, Stanford University, the University of Geneva, the University of Rome Tor Vergata, and the Virtual Economic History Seminar for comments, to Jinlin Wei for research assistance, and to the Centre for Competitive Advantage in the Global Economy and the Leverhulme Trust for funding.

1. Introduction

By 1900, the rail network in colonial India was the fourth largest in the world, covering more than 40,000 kilometers across more than 200 districts (Bogart and Chaudhary, 2016). In striking contrast, public education was poorly funded and saw marginal progress under British rule. Education was an insignificant line item in the government budget, a mere 1.7% compared to 21% for railroads in 1881 (East India, 1887). Education levels were low; in 1891 only 9.6% of primary school-age children were in school (Chaudhary, 2016). According to official opinion, demand for basic education was low in India, where children helped parents in the field (Chaudhary, 2016). By increasing trade, agricultural income, and other labor market opportunities, railways in India may have increased demand for schooling, even in the absence of supply-side, government-led interventions. In this paper, we ask whether there was a demand-driven increase in education in colonial India in response to the extension of the rail network.

Using decennial census data on Indian literacy from 1881 to 1921, we estimate the effect of railroads on total, male, female, and English literacy at the district level. Railroad construction began in the 1850s, and 52% of British Indian districts were connected to a railroad by 1881. This increased to 87% in 1901 and then 96% in 1921. From 1881, data on literacy became available in the decennial censuses. The early censuses (1881 to 1901) cannot be compared to each other, or to later censuses, due to different enumeration standards and to changing age bins used to define cohorts. We therefore use two principal strategies to identify the effect of railroads. The first exploits panel-like variation across birth cohorts within a given district in a given census year. The second exploits cross-sectional variation across districts in a given census year.

In our first approach, we estimate the differential effect of exposure to railroads across cohorts within districts using the censuses of 1911 and 1921, years with comparable literacy data by cohort. Since 93% of districts are connected to the railroad by 1911, we construct upper and lower bound estimates of the cumulative number of years a cohort is exposed to railroads. Our first measure is the cumulative number of years a railroad was present in a district before the youngest member of the cohort of interest reached age 6. Our second measure is the cumulative number of years a railroad was present in a district before the youngest member of the cohort of interest reached age 12. These two ages capture the regular start and end of primary school.

We create a synthetic panel by including district fixed effects that absorb any unobservable, time-invariant factors that correlate with the timing of connection to railroads and that may also correlate with differences in literacy across districts. We also include cohort $\times$ province and census year $\times$ province fixed effects. These control for both national and provincial factors that affect cohort literacy flexibly over time. These fixed effects mitigate common endogeneity concerns relating to railroad exposure.

Variation in railroad exposure within this fixed effects strategy comes from differences across cohorts in exposure within a district, and how this differs from the same cross-cohort differences in other districts. Consider, for example, two hypothetical districts. The first is connected to a railroad in 1901 and the second is connected in 1906. In the 1911 census, the youngest member of the cohort aged 10-15 would have turned 6 in 1907, while the youngest member of the cohort aged 15-20 would have turned 6 in 1902. The youngest member of the 10-15 cohort would have had 6 years of exposure prior to reaching age 6 in the first district and 1 year of exposure in the second. The youngest member of the 15-20 cohort, by contrast, would have 1 year of exposure in the first district and no exposure in the second.

In our second, cross-sectional approach, we use two instruments, which exploit cross-sectional variation in the years of railroad exposure in each census between 1881 and 1921. Building on recent techniques in the transportation literature (Redding and Turner, 2015), we construct one instrument using a 1852 plan proposed by Major Kennedy, consulting engineer to the Government of India (GOI). This plan predated railway construction (Davidson, 1868). Kennedy proposed low-cost routes favouring gentle terrain over more direct routes. Our instrument measures the distance between a district and the lines proposed in this plan. Our second instrument exploits military reasons for building railroads. It measures the distance between a district and a tree connecting 54 military cantonments circa 1864 before major railroad expansion began. Military cantonments were located in places at moderate elevation and away from ravines where the enemy could hide. Distance from military cantonments and lines in the Kennedy plan strongly predict when a district received a railroad and hence years of exposure to the railroad. Our exclusion restriction assumes distance to military cantonments and the lines in the Kennedy plan only affects literacy via railroads and is uncorrelated with unobserved determinants of literacy once we control for observable differences in geography, crop suitability, pre-railroad urbanisation, and religion across districts. In our cross-sectional analyses, we complement our IV strategy with ordinary least squares (OLS), fixed effects for grid cells, and matching estimates.

We find positive and significant effects of railroads on literacy, in particular male and English literacy in the synthetic panel regressions. A standard deviation increase in railroad exposure (17 years) increases total literacy by 0.29 standard deviations for total, 0.31 for male, and 0.25 for male English literacy. We find small and insignificant effects on female literacy. We find similar effects using the second measure of exposure at 0.26 standard deviations for total and 0.29 standard deviations for male literacy. In our cross-sectional regressions, we find positive and significant effects of railroad exposure. Standardised coefficients suggest effect sizes ranging from 0.1 to 0.54 standard deviations depending on the measure of literacy, the census year used, and the specific statistical model. These results support the synthetic panel findings, with the exception of the positive impact on female literacy. There are several possible reasons for this difference, including the greater variation in female literacy that exists across districts than across cohorts within the same district, or the fact that the cross-sectional comparisons allow individuals to be exposed to railroads over their entire lives, rather than simply up to the usual schooling age.

Why did railroads increase education? The proximate mechanism is greater school enrolment. We complement our literacy results using a novel panel data set we have created on primary and secondary enrolment at the district level. These data cover 179 British Indian districts in 1901 and 1911.[1] Using these data, we find a one standard deviation increase in railroad exposure increases secondary enrolment by 0.44 to 0.54 standard deviations, with larger standardised effects in panel models compared to the cross-sectional regressions. We find small and insignificant effects on primary enrolment. One interpretation consistent with these results is that railroads changed not whether children initially enrolled in school, but rather how long they remained in school. Another interpretation arises from a limitation of the colonial Indian data. Many "secondary" schools were in fact elite institutions with attached primary schools, and so our results may be driven by greater enrolment in elite primary classes.

What deeper mechanisms link railroads with greater secondary schooling and literacy? We offer tentative answers using cross-sectional OLS mediation techniques. Since these may suffer from endogeneity problems, we caution the reader to interpret them accordingly. Past work has shown that railroads increased agricultural income (Donaldson, 2018), which in theory can increase literacy if schooling is a normal good. But, employing mediation approaches from Imai et al. (2010, 2011), we find agricultural income is not a significant mediator. Rather, our analysis shows that income taxes, urbanisation, and service sector employment are key mediators through which railroads increase literacy and enrolment. These measures proxy for non-agricultural income and the returns to skill. Because we cannot disaggregate the possible mediating effects of rising non-agricultural income, the relaxation of income constraints for families, and increasing returns to literacy, we view these results as suggestive evidence that railroads increased the demand for education via non-agricultural channels.

Colonial railroads had positive and significant effects on human capital, in particular upper-tail human capital such as English literacy and secondary enrolment. Are these effects large? Our effect sizes on railroads are modest compared to comparable estimates from the nineteenth century United States (Atack et al., 2012). The effects are also modest if we compare them to the impacts of colonial supply-side investments in education. We make

[1]We also have data on a smaller set of districts for 1894, 1897 and 1905.

a back of the envelope calculation that suggests the cost of making one additional person literate via railroad investment would be 200 times as expensive as using public education funding to achieve the same goal. While railroads had many benefits other than increasing education, they could not make a significant dent in Indian illiteracy given the effect sizes we estimate.

We conduct several robustness checks. First, in the synthetic panel estimation, we drop the cohort aged 20 and above because our approach will lead to the most mis-measurement of railway exposure for the older ages in this cohort. Second, we exploit variation across cohorts in the 1901 census. There are more districts unconnected to the railroad in 1901 compared to later. By comparing across cohorts in 1901, measurement error in literacy is less of a concern. Third, we estimate spatially adjusted Conley (1999) standard errors for our pseudo-panel and cross-sectional regressions. Fourth, we test for outliers by dropping the four main cities of Bombay, Calcutta, Delhi and Madras that were among the first to be connected. In another test, we drop one province at a time. We find similar results.

1.1. **Contribution.** Our paper contributes to three literatures. First, our paper contributes to the rich economic history literature on Indian railroads. Much has been written about the effects of colonial railroads on trade, with studies showing large effects on price convergence and income (e.g. Studer (2008), and Donaldson (2018)), small positive effects on city growth (Fenske et al., 2021), ambiguous effects on cropping patterns, and null effects on wage convergence (Collins, 1999).[2] Indian railroads have also featured heavily in nationalist debates on colonisation. Critics argue that the financing of Indian railways delivered excessive returns to British investors, that the network benefitted colonial interests by emphasising port to interior connections over interior to interior connections, and worsened the negative consequences of famines (Dubey, 1965; Satya, 2008). In this view, railways did not industrialise India because they were built to benefit the Empire. An alternative view argues that, although railroads helped colonial interests, they had positive effects on income and returns to British investors were not excessive (Bogart and Chaudhary, 2019; Hurd, 1983). We add to this literature, showing that there were positive effects on schooling, though these favoured men, English literacy, and secondary enrolment.

Second, we contribute to the literature on the effects of transportation infrastructure. For example, building on classic work by Fogel (1964), Donaldson and Hornbeck (2016) find large effects of railroads on market integration and income. Donaldson (2018) finds colonial Indian railroads reduced transport costs and increased agricultural income, which in turn reduced real income volatility and mitigated the effects of famine (Burgess and Donaldson, 2017). Much of this work focuses on trade and market integration. In the case of US railroads,

[2]See Andrabi and Kuehlwein (2010), Hurd II (1975), McAlpin (1974), and Mukherjee (1980) for other work on prices and income.

scholars have looked at the effects on urbanisation, banking, and schooling, among other outcomes (Atack et al., 2010, 2011, 2014). Although there exist papers that look at how current infrastructure expansions affect human capital outcomes such as schooling via labor markets and income, such work is uncommon in the context of railroads. A notable exception is Atack et al. (2012), who study US school attendance in the nineteenth century. Tang (2017), similarly, looks at mortality effects of railroads in Meiji Japan, while Zimran (2020) examines impacts on stature in the United States. Sztern (2018) finds the share of children learning to read in late imperial Russia correlates negatively with proximity to a railroad at the commune level. Our paper looks at the effects of historical railroads on literacy and enrolment, outcomes more commonly examined in work on recent transportation projects (roads and highways) rather than older projects (railroads). We show that the impacts of transportation infrastructure on human capital have not been limited to modern economies.[3]

Third, our paper contributes to the literature on the effects of demand and supply in explaining schooling (Glewwe and Muralidharan, 2016). Many papers estimate the effect of labor demand shifts on education in India with positive effects due to outsourcing facilities (Jensen, 2012) and negative effects related to NREGA (Li and Sekhri, 2020). These relate to larger debates on the relative efficacy of demand versus supply interventions in schooling (Banerjee and Duflo, 2011). On one side, scholars argue that increasing demand for education is sufficient to increase schooling, while other scholars argue public investments are necessary to increase mass education in developing countries. Our paper shows that one of the biggest infrastructure expansions, railroads, had positive effects on literacy and enrolment in India. Yet, these effects are modest and hence not cost-effective if we consider them against increased public funding of education.

The rest of the paper is organised as follows: Section 2 provides historical background on literacy and railways in colonial India, and outlines our conceptual framework. Section 3 describes our data, and is followed by an outline of our estimation strategies in Section 4. We report the results in Section 5, and discuss potential mechanisms in Section 6. And, we conclude in Section 7.

2. Historical Background

2.1. Literacy in British India. As British rule spread in India, former indigenous schools were either replaced or incorporated into the new colonial system of schools. This was a slow and uneven process, which was largely complete by the end of the 19th century (Nurullah and Naik, 1951). Public education funding, which was meagre in the 1850s when the British

[3] Aggarwal (2018) and Adukia et al. (2020) use recent contemporary road and highway construction projects in India to examine their effects on schooling. Similar to our historical findings, these studies find the effects of contemporary infrastructure expansions are driven more by increasing returns to education and higher non-agricultural income rather than by income or substitution effects of rising agricultural income.

Crown took control from the East India Company, increased from 1.5% of the colonial budget in 1882 to 5.2% in 1921. Even with this increase, public spending was under 1% of national income in 1921. Nonetheless funding from public sources increased from 56% of total education spending in 1881 to 63% in 1921 (Chaudhary, 2016).

Did the transition to colonial schooling increase literacy? Unfortunately, we cannot answer that question because there are few comparable estimates of literacy before the 1880s.[4] We know indigenous villages schools were common in the early 19$^{\text{th}}$ century. They attracted boys from different backgrounds, but few girls (Rao, 2020). Yet, these accounts offer few specifics on literacy. One notable exception is a Scottish missionary named William Adam. He estimates literacy was around 4% (ability to read and write) to 6% (ability to sign) across a handful of districts in eastern India in the 1830s (Adam and Long, 1868). It is hard to extrapolate from these estimates because we do not know if Adam's districts were positively or negatively selected compared to the Indian average.

The British began collecting basic education statistics on schools and enrolment in the 1860s and 1870s. And then literacy was added as a standard question in the census of 1881. Although subsequent censuses changed the measurement of literacy, the raw data suggest literacy was low but increased from 1881 to 1921. Male literacy increased from 6% in 1881 to 12% in 1921, while female literacy increased from under 1% to 2% over the same decades (Census of India, 1931).

Literacy in this period increased because more children began attending school. Indeed, enrolment rates increased faster than literacy, from one in ten children attending school in 1891 to just over 1 in 5 in 1921 (Chaudhary, 2016). This is not to say people did not learn to read and write outside of formal schools. Rather, schools offered a natural venue to acquire functional literacy in a society where the vast majority of adults were not literate.[5] Moreover, service sector jobs in colonial bureaucracies often required school completion certificates, which offered an economic incentive to become educated via schools (Sharp, 1914). Indeed, productivity per worker in services grew more rapidly in colonial India than in the economy as a whole (Broadberry and Gupta, 2010).

The increase in enrolment mirrors the increase in public and private aided schools, i.e., privately managed with public subsidies. Between 1881 and 1921, schools per 100,000 people almost doubled from 44 to 70 with publicly managed and funded schools tripling from 9 to 30, while privately managed aided schools almost doubled from 26 to 41 per 100,000

[4]This also precludes comparing literacy in districts before and after they were connected to a railroad in the 1860s and 1870s because we do not have data for the pre-railroad period.

[5]Indian leaders and colonial administrators recognised adult illiteracy was a big problem. Yet, there was no policy effort or funding to address the problem before the 1920s when newly elected Indian ministers took charge of education under Dyarchy. Before the 1920s, adult learning was limited to a few night schools in Bengal, Bombay and Madras, but in the words of Nurullah and Naik (1951), these were "*primary schools conducted at night* rather than adult classes proper (p. 741)."

people (Chaudhary, 2010a). Public funding was used to both expand the number of schools and reduce fees. In 1900 primary school fees accounted for 12% of expenditures in public schools compared to 38% in aided schools, with fees declining to 2% and 16% of expenditures respectively by 1931. Although primary education was not free in public schools, fees were not a significant barrier for skilled workers. For example, annual primary school fees in 1900 would have represented less than 0.5% of the annual wages of a skilled labourer (Chaudhary, 2016). Towards the end of our study period, several provinces introduced legislation to make education compulsory at the primary level in a few municipalities. Such acts were weak by design, with no enforcement resulting in no progress on the compulsory education front by 1921.

On the expenditure side, public funding of education was decentralised to provinces in the 1870s with further decentralisation of primary education to districts and municipalities in the 1880s. The decentralisation led to stark differences in public spending across provinces, for example between the coastal provinces of Bombay and Bengal. Such differences were due to historical circumstances (Chaudhary, 2010a).[6] Bombay received more public funds and built a network of publicly funded and managed schools charging low fees. Unlike Bombay, Bengal received fewer public funds and subsidised privately managed aided schools charging higher fees to build their network.[7] These subsidies went to paying teachers, rather than facilities.[8] A majority of privately managed schools were run by Indians, although Christian missionaries played a larger role at the collegiate level in the 19th century.[9] Other provincial systems fell in between those of Bombay and Bengal.

Yet, these differences in public spending across provinces did not translate into differences in school outputs, namely enrolment or literacy. The coastal provinces of Bengal, Bombay and Madras had higher enrolment and literacy in each decade, with male literacy averaging 20% compared to 11% in the interior for Central Provinces and United Provinces. Apart from regional differences, there were large differences in schooling by caste and religion.[10] Castes ranked higher in the Hindu caste hierarchy were more literate and better educated. Literacy among Brahmans, the traditional priestly caste, averaged 33%, higher than among

[6]The type of Land Settlement, Temporary or Permanent, and the ad-hoc distribution of funds to the provinces between 1833 and 1871 when funding was centralised were big factors driving these patterns (Chaudhary, 2010a).

[7]Annual primary school fees ranged from 1.65 rupees in Bengal, 0.75 in Bombay to 0.23 in the United Provinces in 1901, which was 1%, 0.3% and 0.2% of the annual skilled wage in these provinces (Chaudhary, 2016).

[8]Most elementary schools were rudimentary one room school houses, if that, with multi-grade teaching. There were some new buildings housing secondary schools and colleges in major cities such as Bombay and Calcutta. Yet, they were the exception rather than the norm. See Sharp (1914) for more details on schools.

[9]In 1881, 97% of primary schools and 64% of secondary schools were managed by Indians. Non-Indians, i.e., Christian missionaries managed a larger share of colleges at 70% in 1881 but this declined to 47% by 1901 (Nurullah and Naik, 1951), p.260.

[10]This discussion draws on 1931 literacy for individuals aged 10 and over from Chaudhary (2016).

Christians (28%), another group with high literacy. Regional differences were also visible within groups, with Brahman literacy at 54% in Madras compared to 16% in the United Provinces.[11] In comparison, literacy among lower castes, also known as depressed castes or Scheduled Castes in modern India, averaged 1.6%. Tribal groups had even lower literacy at under 1%.

Against this backdrop of low but varying literacy, few scholars have looked at the effects of demand shifters in explaining levels of schooling. Much of the scholarship argues that poor public funding led to low literacy, which is a reasonable conclusion given the national patterns (Chaudhary, 2016). Yet, differences in public spending *alone* cannot explain the differences in outcomes across and within provinces. To that end, we study if and how railroads affected the demand for basic literacy exploiting temporal and spatial variation within British India.

2.2. **Railroads in colonial India.** Unlike schooling, the British were early promoters of railroads in India, building an extensive rail network.[12] The first passenger line opened in 1853, connecting Bombay to Thane, a distance of 20 miles. Prompted by mercantile interests in Britain, the early lines connected the ports of Bombay, Calcutta and Madras to the interior. Given the few good roads and navigable rivers, British firms hoped railways would lower the costs of exporting raw cotton from India and of importing British manufactured goods to new Indian markets (Thorner, 1951, 1955). Indeed, the British believed goods traffic would significantly exceed any passenger traffic. They proved to be wrong, with passenger traffic accounting for 60% of revenues.

Indian railways were built by British firms with British financing, albeit subsidised by a guaranteed dividend backed by the GOI. Such firms were the main players up to the 1870s, when the GOI began to build lines. This was followed by mixed public-private partnerships in the 1880s. These partnerships were the norm until the 1920s (Sanyal, 1930). Route mileage expanded quickly in the early decades, especially from 1881 to 1901. Total route miles increased from 9,893 in 1881 to 17,283 to 1891, 25,365 in 1901, 32,839 in 1911 and then 37,266 in 1921 (Bogart and Chaudhary, 2016).

Figure 1 maps the spread of the network from 1881 to 1921. The important ports were connected to the interior before 1881. Many lines crossed the densely populated Indo-Gangetic plain with fewer interior lines in the Deccan plateau. Early proposals such as the Kennedy plan in 1852 called for lines parallel to the coast in order to economise on costs. Some were never built because subsequent officials opted for more expensive routes cutting

[11]Brahmans were more literate than the average person in most provinces, but other upper castes such as Kayasthas that traditionally worked as accountants or scribes were more literate than Brahmans in the United Provinces with 45% literacy.

[12]There is a large literature on Indian railways. Edited volumes by Kerr (2001, 2007) offer an excellent introduction to the main issues, while Sanyal (1930) offers a detailed history of railway development.

through mountains (Davidson, 1868). We return to the Kennedy plan below in order to construct an instrument for route placement.

Although British firms built the railways, the GOI dictated route placement. What guided their decisions? Military, commercial, and famine concerns were cited as the main drivers in official correspondence (Hurd, 1983). Following the Sikh Wars in the 1840s and the Indian Mutiny in 1857, the British were cognisant of the need to transport troops and supplies across the country at low cost. Existing transport routes were of poor quality and slow, which made it necessary to station troops at multiple locations in the event of an uprising (Papers, 1854). On the commercial side, British merchants lobbied for Indian railways that would connect the ports to cotton-growing regions in the interior, and from the eastern and western ports to Delhi in the north. Another consideration was famines. Following devastating crop failures and famines in the 1870s, the GOI built "protective lines" in famine-prone regions of the South. Finally, a few small lines were built connecting hill stations such as Simla where British officials liked to spend their summers. While not random, the network of railroads across districts was not uniformly indicative of positive or negative selection. Rather, a mix of factors affected where and when railroads were built. Coastal districts with important ports were connected early as were those in the Ganga valley. Yet, a few cotton-growing interior districts were connected before 1881, as were districts closer to Afghanistan. Neither group would be considered positively selected for rail access. To address the endogeneity of railroads, we compare cohorts within districts in panel models and use an instrumental variables strategy among other cross-sectional models.

2.3. **Conceptual Framework.** There are many channels through which railways may have increased schooling and literacy in colonial India. First, there is evidence that greater access to transportation infrastructure raises agricultural incomes and farm values (Redding and Turner, 2015) in several contexts, including modern Nepal (Jacoby, 2000), historical England (Bogart, 2009), and the historical United States (Atack et al., 2012; Donaldson and Hornbeck, 2016). In colonial India, railways increased agricultural incomes (Donaldson, 2018), though this was limited to some degree by a lack of peasant price responsiveness (McAlpin, 1974). If schooling is a normal good, we would expect greater agricultural income to increase schooling. Indeed, there is evidence of negative income shocks reducing human capital investments in the Ivory Coast (Cogneau and Jedwab, 2012) and Uganda (Björkman-Nyqvist, 2013). Yet, the effects of road infrastructure programs in contemporary India on schooling do not appear to be mediated by rising agricultural income (Adukia et al., 2020).

Second, railroads may also increase the opportunity cost of education, either in agricultural or non-agricultural work. This channel is consistent with modern evidence from fracking in the United States (Cascio and Narayan, 2022), manufacturing in Mexico (Atkin, 2016), agricultural wages in India (Shah and Steinberg, 2017), and workfare in India (Li and Sekhri, 2020; Shah and Steinberg, 2021). Colonial officials argued that the importance of child labor in agriculture reduced demand for schooling in India (Chaudhary, 2016). More generally, greater access to trade may increase the returns to both adult and child labor, leading to ambiguous predictions for the response of schooling (Edmonds and Pavcnik, 2005). This could be due to opportunities for child labor in both the non-agricultural and the agricultural sector, both of which existed in colonial India (Balagopalan, 2014; Hindman, 2009; Sen et al., 1999).

Third, railroads may also increase human capital via access to more non-agricultural income. Transportation infrastructure has increased income in manufacturing, services, or both in many places (Atack et al., 2011; Banerjee et al., 2020; Faber, 2014; Gibson and Olivia, 2010; Jaworski and Kitchens, 2019). Adukia et al. (2020) identify three channels by which access to non-agricultural labor markets may increase education: raising the skill premium, increasing household earnings, and easing liquidity constraints. By raising the returns to skill, access to non-agricultural income can increase schooling. This may occur if, for example, literacy increases productivity more in industrial work than in agricultural work, or if industries change their composition of output towards more skill-intensive goods. This would be consistent with modern evidence from business process outsourcing in India (Jensen, 2012), information technology in India (Oster and Steinberg, 2013), and Bangladeshi garment factories (Heath and Mobarak, 2015). Relatedly, by facilitating information flows (as in Kaukiainen (2001)), railways may have increased knowledge of the returns to education. That said, there is mixed evidence on the effectiveness of interventions that provide information to parents and students on the returns to education (Glewwe and Muralidharan, 2016) in developing countries today.

Another possible channel linking railways to education is urbanisation. Several studies have shown that connection to transportation infrastructure increases city growth and urbanisation (Redding and Turner, 2015), including in Argentina (Pérez, 2018), and the United States (Atack et al., 2010). In colonial India, Fenske et al. (2021) find that, while modest, there was a positive effect of railroads on city size. Their results do, however, consider urbanization at the level of cities rather than districts. Rural-urban gaps in literacy have long been apparent in developing countries, including India (Das and Pathak, 2012; Gollin et al., 2021).

Railways may also have improved health outcomes, which in several modern developing countries have been shown to improve schooling (Currie, 2009; Miguel and Kremer, 2004). In colonial India, Burgess and Donaldson (2017) have shown that railways reduced famines. In the short run, however, railways may increase exposure to infectious disease, worsening mortality as in the case of Meiji Japan (Tang, 2017). Finally, railroads may also have increased the supply of education, either by reducing the cost of providing schooling, or by increasing access to existing schools. Adukia et al. (2020) consider but find no evidence that India's national rural road construction program affected the supply of teachers and schools or access to existing schools. Land taxes in colonial India affected the degree of public spending on rural education. Although railroads increased agricultural income, land taxes were fixed in the east of British India, and revised infrequently in other parts of the country (Kumar, 1983). Yet, to the extent they responded to changing agricultural income, railroads could affect enrolment and literacy by increasing public spending on education.

As this discussion highlights, railroads could have increased, decreased or had no impact on education. It may have been that incomes in colonial India were too low for the positive channels above to operate. Indeed, railways in colonial India did not lead to convergence in wages in the same way that they drove price convergence (Collins, 1999). Further, Indian households faced high costs of borrowing that may have dampened the impacts of increases in the returns to education (Chaudhary and Swamy, 2017; Cheesman, 1997). We explore the mediating role of such channels in Section 6. While we are unable to examine all mechanisms considered in the literature, the availability of historical data allows us to evaluate some of these. We evaluate the role of rising incomes and returns to labor in agriculture using data on agricultural income, land taxes, and unskilled wages. And, we proxy for measures of the returns to skill and rising non-agricultural incomes using income tax revenues, urbanisation and the share of workers in industry and services. We consider, and rule out, that the results are driven by railway schools, migration, the presence of Europeans, missions, or railroad workers.

3. Data

We construct a new district-level dataset for British India from 1881 to 1921 in order to test the relationship between railroads and education. Our data pulls information from four primary sources:

(1) the decennial censuses of 1881 to 1921, which we use to measure literacy and several other control variables;
(2) the 1934 edition of *History of Indian Railways Constructed and in Progress*, which we use to code the opening dates of the railroad;
(3) the District Gazetteers of India, which we use to code primary and secondary enrolment rates, and;
(4) multiple sources of Geographic Information System (GIS) data, which we use to construct geographic controls.

We begin with the 1881 census rather than the first 1872 census because of incomplete coverage and inconsistencies with the 1872 census (Walby and Haan, 2012).[13] We also focus on British Indian districts because the Princely State data are inaccurate in the early censuses.[14] We describe our data sources in the remainder of this section.

3.1. **Measuring Literacy.** The colonial census reports literacy by gender and age-cohort. From 1901, the census also reports English literacy. Despite its richness, enumerating literacy over time is difficult because of changes in definition and measurement. In the 1881 and 1891 censuses, individuals were classified into three categories: literate, learning, and illiterate. Yet, enumerators were given no guidance on measuring literacy or accounting for learners apart from an age threshold (Gait, 1913).[15] Age bins in these early censuses were also different across provinces.

Beginning with the 1901 census, the "learning" category was dropped and literacy was reported for standard age bins: those under age 10, aged 10 to 15, aged 15 to 20 and those over age 20. A uniform definition was adopted, namely "the ability to read and write." But again, census administrators were not given official guidance on measuring literacy. This led to differences across provinces. Some used a rigorous standard while others enumerated individuals as literate if they could sign their name (Gait, 1913).[16] It was only in 1911 that a uniform standard, the ability to read and write a short letter, was introduced. This makes literacy in the 1901 and later censuses difficult to compare. For example, many children under age 10 were counted as learners in the 1891 census, then some children under 10 were recorded as literate in the 1901 census, but not in subsequent censuses. Indeed, we observe literacy decline in some districts for those under age 10 between 1901 and 1911 because of these problems.

[13]Walby and Haan (2012) summarise the many issues with the first 1872 census including incomplete territorial coverage and inconsistent enumeration across provinces. Walby and Haan (2012) aptly quote official opinion: "Later commentators said that the only consistency in the 1871-72 Indian census was the "uniform absence of uniformity"" (p. 309).

[14]Colonial India encompassed British India with territories that were under direct British rule and Princely States that were governed by Indian rulers. Territorial coverage of the Princely States is incomplete and inconsistent up to 1911 (Census of India, 1901; 1911).

[15]See the education chapter in the Census of 1911, Part 1 - Report written by E.A. Gait. The final enumeration of the census in each village or town was typically done at night on a date that was announced in advance. Enumerators did not assess the literacy of each household member. Rather, they used the provincial standard for literacy decided on before the census to record individuals as literate. As noted by Parulelar (1939), a former head master and the secretary of the Municipal Schools Committee in Bombay, "On the whole, therefore, according to responsible authorities competent to pronounce views on the accuracy of the Census literary figures, it may be safely assumed that the figures of literacy as revealed in the Census reports are fairly reliable" (p. 93).

[16]Gait, writing in the 1911 census, attests to the inconsistency: "In some parts criteria similar to those mentioned above appear to have been taken, while in others persons were entered as literate who could do little more than write their own name and spell out a few simple printed words" (p. 291).

To get around these issues, we focus on cohort literacy in the 1911 and 1921 censuses, when literacy was uniformly measured. The cohorts reported in these censuses are under 10, 10-15, 15-20, 20 and above. Using cohort literacy allows us to draw on the more accurate 1911 and 1921 census while retaining the ability to exploit the panel-like changes across cohorts over time within a given district. These censuses exploit more intensive margin variation in exposure to railroads. To exploit the more extensive margin growth of railroads before 1901, we perform two additional analyses. The first analysis uses the panel-like variation across cohorts within the 1901 census. The second relies on cross-sectional regressions estimated separately for each census year from 1881 to 1921. We address the endogeneity of railroad exposure in these cross-sectional exercises using instrumental variables, matching, and grid cell fixed effects. We describe these methods in the next section.

Table 1 shows literacy by cohort, gender, and language from 1901 to 1921. These are crude literacy rates equal to the number of literates in each group divided by the population of that group. Focusing on the cohort 20 and above, total literacy increased from 6.45% in 1901 to 8.41% in 1921. Men were more literate than women, though this gender gap narrowed over time. Men were 17 times more likely to be literate than women in 1901 compared to 9 times more likely to be literate in 1921. English literacy was low in absolute terms, but was a sizeable share of all literacy. Almost 15% of literate individuals in 1921 were, for example, also literate in English. Most children typically learned to read and write in a vernacular language before learning English (Sharp, 1914). So: English literacy was, in particular, a measure of upper-tail human capital (Basu, 1974).

Table 2 shows total, male, and female literacy for each cross-section from 1881 to 1921. As noted above, we do not have comparable data on cohort or English literacy earlier in the 19$^{\text{th}}$ century. Total literacy doubled from an average of 3% in 1881 to 6% in 1921. Differences by gender are again visible. Moreover, the standard deviation and range highlight the large differences across districts. Figures 2 through 4 show the distribution of total, male and female literacy across districts in each census year. While the distribution of literacy was highly skewed in 1881, it became more dispersed by 1921. Less than 1 person in 10 could read and write in over 95% of districts in 1881. By 1921 more than 1 person in 10 was literate in roughly 10% of districts, with a maximum literacy of 32% in Madras city. Male literacy was more dispersed than total literacy, as shown in Figure 3. It increased on average and became more dispersed from 1881 to 1921. Female literacy increased from its low base in 1881, yet the distribution remained highly skewed as late as 1921, as shown in Figure 4. Figure 5 shows literacy maps by quintile for 1881, 1901 and 1921.

3.2. **Measuring School Enrolment.** Unlike literacy, which measures the stock of human capital, enrolments capture the flow of human capital. As we expect railroads to affect the stock of literacy by increasing the flow of children in school, we complement our analysis of literacy with an analysis of school enrolment.

District enrolments are not reported in the colonial census, or national reports. Rather, they are reported in many district gazetteers. These sources are less uniform than the decennial census. Nonetheless, we construct a series on primary and secondary enrolment rates between 1894 and 1911, years with the most uniform data.[17] This is an unbalanced panel, as most provinces report enrolment for a subset of years with only a few provinces reporting more years.

Primary school enrolment is recorded as the number of children enrolled in primary schools divided by the cohort under age 10. Secondary enrolment is children in schools other than primary schools, divided by the cohort aged 10 to 15.[18] Apart from a handful of districts, these data are not reported by gender. Another shortcoming is many secondary schools had attached primary classes, so some primary aged children will, then, be included in secondary enrolment. For example, 47% of children in English secondary schools in 1912 were in primary classes (Sharp, 1914) increasing to more than 60% in Assam and Eastern Bengal. Such primary classes were of higher quality than regular vernacular primary schools. As stated in Richey (1923),

> "The fact is that a very large percentage of the boys receiving elementary education in towns are not attending primary schools but the preparatory departments of secondary schools. It is only parents of the poorest class who send their boys to municipal primary schools" (p. 109).

While this introduces measurement error in both primary and secondary enrolment, we are unaware of district-level enrolment data of all primary school children, regardless of school type. Our measure of primary enrolment averages 4%, compared to 3% for secondary enrolment in 1901 and 1911. These enrolment rates are highly correlated with literacy, as we would expect. For instance, the correlation between 1901 literacy and total enrolment is 0.88, primary enrolment is 0.81 and secondary enrolment is 0.6.

3.3. **Measuring Access to Railroads.** To estimate the effect of railroads, we follow Fenske et al. (2021) to code the opening dates of railway access in each district. Fenske et al. (2021), following a procedure similar to Donaldson (2018), construct a polyline shapefile of the Indian railway network with an opening date for each segment. These dates are based on the 1934 edition of *History of Indian railroads Constructed and In Progress*. For each listed railway

[17]In the case of Madras, we use the data reported for the nearest year to 1901 and 1911 in the analysis, namely 1903 and 1913 respectively. We adjust the years of railroad exposure for Madras districts accordingly.

[18]By definition, secondary enrolment includes students in colleges and other schools. In Bengal where we have detailed enrolment information, high school and middle school enrolment accounts for 77% of secondary enrolment. In less advanced provinces, this percentage is likely to be higher because there were fewer colleges, training and other schools as there were in Bengal.

line, they record the opening dates along with the beginning and end points of each line. We intersect this shapefile of railway lines with a map of modern sub-districts. Using a GIS mapping of colonial districts to these modern sub-districts, we compute the earliest year that each colonial district is connected to the railroad.

Two common methods of measuring railroad access are (1) simple indicators for whether a location is connected to a railroad or not (Andrabi and Kuehlwein, 2010; Atack et al., 2010) and (2) market access (Donaldson and Hornbeck, 2016). Neither of these measures is well suited to our question. Both measures fail to capture whether exposure occurs over the typical ages of elementary schooling for members of the population. Nor do they capture the duration of this exposure. Market access is derived from structural trade models and summarises well the economic forces acting on outcomes such as city growth and agricultural incomes. It is less suited to other potential mechanisms linking railroads to literacy, such as the diffusion of cultural practices and of information. We thus, in our synthetic panel analysis, use two measures of the cumulative number of years that a given cohort in a given district was exposed to railroads up to a cohort's elementary school ages. We also show robustness checks using an indicator for railroads in a matching exercise.

For the panel analysis, we define two measures of railroad years assuming railroads affect literacy only up to the beginning or end of elementary school. This is a reasonable assumption because most children are literate or not by the end of elementary school (age 12).[19] If, for example, a railroad arrived in a district in 1893, it would not affect literacy for the cohort 20 and above in 1901 because they would be age 12 and above in 1893, and so would have finished primary school. If railroads arrive after the youngest age in a cohort is out of elementary school, we assume this cohort has no exposure to railroads (coded as 0). Unlike the 20 and above cohort, the arrival of railroads in 1893 would affect cohorts under age 10 and 10-15 in 1901 because they would presumably be in elementary school as railroads arrive. Our railroad measure captures such differences across cohorts.

Since the age bins do not perfectly correspond to elementary school years, we use the youngest age in the bin to measure cumulative exposure up to elementary school and construct two measures of railroad exposure. Our first measure is the number of years a railroad has been operating in a district minus the number of years since the youngest member of a cohort would have regularly begun elementary school, i.e. at age 6.[20] Denote this number of years since schooling began as $y(c)$. For cohorts aged 20 and above, $y(c)$ is 14. For cohorts

[19]See Sharp (1914) and other official Quinquennial Reviews of Education for discussion on colonial primary schools, enrolment and literacy. Wastage was a common concern in the primary school system because more than a third of students would drop out before completing 3 or 4 years of primary schooling, which was considered necessary to attain functional literacy (Parulelar, 1939).

[20]We use age 6 as the beginning of elementary school because the Indian compulsory school schemes in the 1910s used age 6 as an entry point. Sharp (1914) notes primary school lasted for six years and could begin as early as age 5.

aged 15-20, it is 9. For cohorts aged 10-15, it is 4. For cohorts aged below 10, it is 0. For cohort c, $y(c)$ years since schooling began, in district d, with a railroad that opened in year r, measured in census year t, we define our treatment measure $RailroadYears_{cdt}$ as:

$$(1) \qquad RailroadYears_{cdt} = \begin{cases} \max\{t - r - y(c), 0\} & \text{if } r \leq t, \\ 0 & \text{if } r > t. \end{cases}$$

Our second measure is the number of years a railroad has operated in a district minus the number of years since the youngest member of a cohort would have regularly finished elementary school.[21] This measure assumes railroads affect literacy up to age 12 for the index age in a cohort. In equation (1), this is equivalent to replacing $y(c)$ with 8 for the cohort age 20 and above, 3 for cohorts aged 15 to 20, and zero for the cohorts aged 10-15 and 0-10.

As constructed, the two measures bound the duration of exposure of railroads. To make these measures more concrete, consider the following hypothetical. Suppose a railroad opens in a district in 1901. In the 1911 census, the railroad has been active for 10 years. In the cohort aged 10-15, the youngest member of that cohort was 6 in 1907 and has not yet turned 12. By our first measure, the railroad exposure for that cohort is $\max\{1911-1901-4, 0\} = 6$ years. By our second measure, the railroad exposure for that cohort is $\max\{1911 - 1901 - 0, 0\} = 10$ years.

To give a real example, Dehra Dun district was connected to a railroad in 1900. In the 1911 census, an individual aged 20 would have been 9 when railroads arrived in 1900. Our first measure codes the cohort 20 and above as untreated (0) in 1911, while our second measure codes this cohort's exposure as 3 years in 1911. Our first measure assumes parents decide whether to enrol their children in school based on cumulative exposure to the railroad up to the beginning of elementary school (age 6). Our second measure assumes parents decide whether to keep their children in elementary school based on cumulative exposure to the railroad up to the end of elementary school.

One may be concerned that age was incorrectly reported to census enumerators, which could introduce measurement error in cohort literacy. Indeed, age heaping at even numbers and multiples of five was common in colonial India (Census of India, 1911). But, census enumerators estimated an individual's age if it was at odds with their appearance. Census officials believed the age enumeration by cohort was reasonably accurate although the number of people at a specific age say 2 years old may be incorrect.[22] Using two measures of railroad

[21]The length of primary school varied across Indian provinces from 5 to 6 years. We chose age 12 as an upper bound on a child completing elementary school.

[22]According to PJ Mead and G Laird Macgregor in the 1911 Census of India for Bombay, "the census is taken on each occasion by the same class of individual dealing with much the same sort of material, and with the vast numbers that form our population the errors tend to counteract each other and age returns

exposure further alleviates concerns of measurement error as does the cross-sectional analysis on total literacy.

For the cross-sectional analysis, we count the number of years a district has been connected to a railroad in each census. As seen in Table 2, 50% of districts were connected to a railroad by 1881, increasing to 96% by 1921. Indeed, most of the increase happened before 1901, with 87% of districts already connected. The railroad years measure better illustrates the variation across districts. For example, the number of railroad years averaged 7.4 years across districts in 1881, increasing to 22 years in 1901, and 41 years by 1921. Comparing the railway and literacy maps suggests a small positive correlation between railroads and literacy. Figures 6 and 7 confirm the positive correlation for 1881 and 1921.

3.4. **Geographic and Socioeconomic Controls.** India has a wide range of terrain with the Himalayas in the north, mountain ranges along the east and west coasts, the Thar desert in Rajasthan, alluvial plains along the Indus and Ganga river valleys, and the Deccan plateau. Such differences in topography affected the railroad network because of the inherent difficulty in building railroads crossing mountains and deserts. These differences may also be relevant to explaining literacy gaps between districts. To this end, we construct a rich set of geographic variables in order to control for the selection into railroad exposure driven by geography.

In particular, we collect data on the latitude and longitude coordinates of the centroid of the district, which we compute ourselves. We control for ruggedness from Nunn and Puga (2012). We control for altitude, precipitation, temperature, slope, and suitability for growing specific crops such as cotton, dryland rice, wetland rice, and wheat, averaged over raster cells within a district. These are taken from the Food and Agriculture Organisation of the United Nations' Global Agro-Ecological Zones (FAO-GAEZ) data portal. While the FAO data are based on current conditions, they are exogenous to human action and represent expected yields under low-input rainfed agriculture, and so have become widely used, including in economic history (Dimico et al., 2017; Kung and Ma, 2014). Since proximity to the coast and rivers likely influenced railroad access, we include indicators for rivers and coastal districts as captured in Natural Earth Data's shapefile maps of rivers and coasts. We also control for medieval ports recorded by Jha (2013). We control for the seasonality of rainfall. In particular, using data on historic rainfall from Matsuura and Willmott (2018), we compute the Feng et al. (2013) entropy-based measure of seasonality. Finally, we control for the Kiszewski et al. (2004) index of the stability of malaria transmission.

Apart from geography, we control for the scale of urbanisation before the advent of railroads using the population of cities enumerated in Chandler and Fox (1974) circa 1850.

en masse are probably much nearer the mark that they appear to be, though the precise number at any particular age period is probably quite inaccurate" (Census of India, p. 75).

These cities range in population from 6,000 to 580,000 across 59 districts. This effectively controls for more urban districts that were likely to be connected with railroads before less urban districts.[23] We also control for the religious and caste composition of a district. In particular, we include the share of Brahmans, traditional Hindu elites in the caste hierarchy that enjoyed higher literacy on account of their traditional occupations as priests and teachers. We also include the shares of Muslims, Christians, and tribal groups. Such shares are intended to capture historical differences in education among these groups that may be correlated with railroad access.[24] These data are taken from the colonial censuses.

4. Estimation Strategy

Our main results exploit variation within districts and across cohorts to identify the effects of railroads on literacy. We complement this synthetic panel exercise with cross-sectional results using instrumental variables, grid cell fixed effects, and matching techniques.

4.1. **Synthetic Panel.** We estimate the following model using the log of the literacy rate by year, district, and cohort as the outcome.

$$\ln(LiteracyRate_{cdt}) = \beta RailroadYears_{cdt} + \theta_d + \delta_p \times \eta_t + \delta_p \times \gamma_c + \epsilon_{cdt} \tag{2}$$

In this model, $LiteracyRate_{cdt}$ is literacy for cohort c in district d and census year t. We transform literacy into logs because it is a highly skewed variable, as shown in Figures 2 - 4. We estimate the model for $t \in \{1911, 1921\}$ and cohort $c \in \{0-10, 10-15, 15-20, 20+\}$. $RailroadYears_{cdt}$ measures the cumulative years of railroad exposure for cohort c in district d in year t.

We control for several fixed effects. First, district fixed effects, θ_d, capture unobservable time-invariant district features that lead some districts to get railroads before others and that may correlate with literacy. Second, we include interactions of province $\times$ year and province $\times$ cohort fixed effects captured by $\delta_p \times \eta_t$ and $\delta_p \times \gamma_c$ to control for provincial changes in census enumeration methods by year and cohort. We do not include separate fixed effects for year (η_t) and cohort (γ_c) because they are collinear with $\delta_p \times \eta_t$ and $\delta_p \times \gamma_c$. Since official guidance on census enumeration was set by provinces, administrative units larger than the district (Gait 1913), such flexible controls address most measurement concerns related to literacy as well as accounting flexibly for omitted variables at the province and cohort level

[23]Our results are robust to an indicator for districts with a city enumerated in Chandler and Fox (1974). We believe population is a better control because it captures the intensive dimension of urbanisation, which the extensive measure does not.

[24]Some may view these as bad controls because these groups did not settle randomly across India. Our results are similar, if anything showing larger effects for railroads, when we do not control for them.

that may change over time.[25] We cluster standard errors by district to account for serial correlation over time. As a robustness check, we estimate Conley (1999) standard errors that account for spatial correlation with cutoffs ranging from 200km to 500km.[26]

In this setup, we identify the effects of railroads using variation in cumulative exposure to railroad years across cohorts within districts over time. The key identifying assumption is that such exposure in railroad years is uncorrelated with the error term ϵ_{cdt}. We believe this is a reasonable assumption given the flexible fixed effects included in the model. As a robustness check, we run the same analysis using the 1901 census and controlling for district fixed effects and province × cohort fixed effects.[27] Since we use only the 1901 census for this exercise, changes in the standards used to measure literacy in different censuses are not an issue. We report other robustness checks in the results section.

4.2. **Cross-Section.** We complement the panel methods with several cross-sectional models as follows.

4.2.1. *Ordinary Least Squares.* We exploit the complete data from 1881 to 1921 using repeated cross-sections. For each census year, we estimate a separate OLS regression of the following form:

$$\ln(LiteracyRate_{dt}) = \beta RailroadYears_{dt} + \gamma' x_{dt} + \delta_p + \epsilon_{dt} \tag{3}$$

We estimate this regression separately for $t \in \{1881, 1891, 1901, 1911, 1921\}$. In this equation, $\ln(LiteracyRate_{dt})$ is the log literacy rate in district d in year t. Unlike cohort literacy in the synthetic panel, this measure picks up adult literacy because everyone who is literate is included in total literacy regardless of age. $RailroadYears_{dt}$ is the number of years district d in year t has had a railroad. This is 0 if the district is unconnected in t. We do not adjust $RailroadYears$ because the outcomes are total, male, female, and English literacy. The vector x_{dt} includes the GIS controls, pre-rail urbanisation and social controls described in section 3. We also include province fixed effects captured by δ_p. Finally, ϵ_{dt} is the error term. We estimate robust standard errors and report results with Conley (1999) standard errors with distance cutoffs of 200km as a robustness check.

Such a regression may generate biased estimates of the causal effect of railroads. For example, if more developed districts were the first to receive railroads, our estimate of railroad years would be positively biased because it would conflate the effect of railroads with those of prior development. On the other hand, if famine-prone areas received access early on,

[25] With four cohorts and two years, we do not have sufficient power to include district-cohort and district-year interactions, akin to a triple difference specification.

[26] We only report results for 200km. Any result that is significant at the 5% or 1% level with a cutoff of 200km remains so at the same level of significance using cutoffs from 300km to 500km.

[27] We are unable to do the same exercise for 1881 and 1891 because they do not report uniform age bins.

then our estimates would likely have a negative bias. Indeed, military strategy also does not provide clear guidance on the potential selection problem. Railroads from Calcutta to Delhi facilitated quick movement of troops *and* were of commercial value in transporting goods. This would suggest positive selection. Railroads from Delhi heading northwest towards Afghanistan were of less commercial value and would, alternatively, suggest negative selection. To address such endogeneity concerns, we employ matching, fixed effects, and instrumental variables solutions.

4.2.2. *Grid Cell Fixed Effects and Matching.* We complement the OLS results with grid cell fixed effects and matching models. For the grid cell fixed effects, we construct $2° \times 2°$ grid cells based on the latitude and longitude coordinates of district centroids. 1° of latitude is roughly 111 kilometers. 1° of longitude ranges from roughly 111 kilometers at the equator to 85 kilometers at $40°N$ latitude, just south of New York, Madrid, or Tashkent. These grid cells will range, then, between 10,000 and 12,000 square kilometers in area in our data. For comparison, modern-day Tripura state has an area of 10,486 square kilometers, while Punjab has an area of 50,362 square kilometers. We include fixed effects for these grid cells in the OLS regressions. This ensures we are comparing neighbouring districts with different durations of railroad exposure for identification. We include our standard controls in these specifications.

We also estimate nearest neighbour matching models using the rich GIS controls to match districts on their time-invariant characteristics. Since 87% of districts are connected by 1901, we focus on the 1881 and 1891 cross-sections, in which there is more balance between connected and unconnected districts. In these nearest neighbour matching models we measure railroad connection as an indicator variable.

4.2.3. *Instrumental Variables: Military Cantonment and 1852 Kennedy Plan.* Finally, we construct two instrument for $RailroadYears_{dt}$ that exploit different sources of variation. First, we build a tree spanning the 54 British military cantonments that existed as of 1864 before major expansion of railroads. If military concerns drove the placement of railroads, we expect cantonments where army troops were stationed to get early access. Using Prim's algorithm, we construct the shortest tree that spans these 54 military cantonments. Figure 8 shows a map of this tree superimposed on the 1881 railway network. After constructing the tree, we compute the distance of each district from the spanning tree. We then use the log of (one plus) distance to this tree as an instrument for $RailroadYears_{dt}$. According to Kulkarni (1979), two factors determined the location of cantonments. First, these places had to be "suitable for outdoor training round the year" (p. 214). This favoured areas at moderate elevation. Second, cantonments could not be located near ravines where an enemy could hide. Both factors favour a particular aspect of climate and geography that is plausibly exogenous to other factors affecting human capital conditional on the geographic and socio-economic controls. Apart from geography, more cantonments were in northern India because of its plains that were vulnerable to attacks compared to hill areas (Kulkarni, 1979). Our analysis includes province fixed effects that capture this dimension of location.

Our second instrument uses Major J. P. Kennedy's 1852 proposal for building railroads in India. Major Kennedy was the Consulting Engineer for the GOI and played a key role in planning India's railway network. He pushed for building low-cost railroads that, in his view, would confer innumerable benefits. As stated in his own words:

> It is not sufficient to be convinced as I am, that the establishment of Railways in India is an essential preliminary to the attainment of the highest degree of efficiency of which our military and civil administrations are capable; to the prevention of local famine, and to the uniform dispersion of food; to any vigour and activity in manufacture or commerce; to the increased consumption of English goods: to the power of competing with America in furnishing to England raw cotton and other important articles: in short, to the growth of everything connected with the extension of British interests in India as well as with the industry, the wealth, and the comfort of its vast population (Parliamentary Papers 1854, p.3).

Yet, Major Kennedy was aware of the costs of building railroads. So he emphasised lower-cost routes connecting the ports with the interior. In particular, his plan called for a network in "strict harmony with the natural advantages" of the country. Unlike routes that would cut through the Eastern and Western coastal ranges of India, his plan called for routes that favoured softer gradients, following the coast and natural topography.

Donaldson (2018) used portions of the Kennedy plan that were not implemented to construct placebo lines. In many cases, however, Kennedy's routes were adopted, as seen by comparing the Kennedy plan in Figure 9 to the actual network in Figure 1. In other cases, however, more expensive routes were selected. Exploiting geographical features in favour of low cost routes was Kennedy's focus. Hence, we are assuming here that, conditional on controls, the 1852 Kennedy plan is uncorrelated with factors that would affect literacy other than through access to railroads. To construct the instrument, we convert the map of Kennedy's proposal into a polyline shapefile. We then calculate the shortest distance of each district from this route. We use the log of (one plus) distance to the lines in the Kennedy plan as an instrument for $RailroadYears_{dt}$.

In our view, the two instruments exploit different sources of variation. The cantonment instrument favours districts close to areas at moderate elevation and away from ravines where railroads arrived earlier due to strategic military reasons. In comparison, the Kennedy plan instrument relies on geographic differences favouring easier terrain for building railroads. Indeed, the two instruments are uncorrelated with each other with a correlation coefficient of 0.03. We report results using the two instruments in the same regression with an over-identification test and results of each instrument used individually.

5. Results

5.1. **Synthetic Panel.** Table 3 shows our main results on railroad exposure, which exploit variation across cohorts within districts in 1911 and 1921. Column (1) focuses on the log of total literacy, column (2) on male literacy, and column (3) on female literacy. In the second panel we show results for English literacy. We report results for non-English literacy in the bottom panel. We calculate non-English literacy by subtracting English literates from total literates and dividing by the relevant population.[28]

As seen in Table 3, the coefficient on railroad exposure is positive and significant for total and male literacy, but not female literacy. In terms of magnitude, the standardised β coefficients (multiplying the β coefficient in Table 3 by the standard deviation of cohort railroad years, 17.6 years, and dividing by the relevant standard deviation of log literacy) range from $\frac{0.0202\times17.63}{1.250} = 0.29$ standard deviations for total and $\frac{0.0224\times17.63}{1.267} = 0.31$ standard deviations for male literacy in the top panel. We find smaller effects on male English compared to male non-English literacy with standardised coefficients at $\frac{0.0266\times17.58}{1.888} = 0.25$ for English and $\frac{0.0235\times17.62}{1.226} = 0.34$ for non-English literacy. Unlike males, we find small and insignificant effects of railroads on female literacy, female English and female non-English literacy.[29]

This exercise includes the cohort aged 20 and above. Individuals aged 20 in this cohort began elementary school fourteen years prior at age 6, but others at age 30 in the cohort were past elementary school fourteen years prior when they were age 16. To ensure our results are not driven by such mismeasurement in cohort railroad years, we estimate the same regressions as above for the cohorts under 10, 10-15 and 15-20, removing those aged 20 and above. Any measurement error in cohort railroad exposure is smaller for these tighter age bins. As seen in Table A4, the results are similar, albeit with stronger results for non-English literacy compared to English literacy.

As outlined in section 3, our first measure of cohort railroad exposure uses an index age in a cohort based on the beginning of elementary school at age 6. Our second method of constructing cohort exposure uses an index age for a cohort based on the completion of elementary school at age 12. Table A5 shows the results using these different exposure measures. We find similar results with positive and significant effects only for male literacy. In terms of magnitude, they are marginally smaller at $\frac{0.0186\times17.76}{1.250} = 0.26$ standard deviations for total literacy and $\frac{0.0206\times17.76}{1.267} = 0.29$ standard deviations for male literacy.

[28]Appendix Table A1 shows regressions using spatially adjusted Conley (1999) standard errors. In particular, we use the implementation for panel data developed by Hsiang (2010) and Fetzer (2020), and report results with a maximum of five lags and distance cutoffs of 200 kilometers. Any result that is significant at the 5% or 1% level with a cutoff of 200km remains so at the same level of significance using cutoffs from 300km to 500km.

[29]Our results are robust to dropping the four cities of Bombay, Calcutta, Delhi and Madras as seen in Table A2. They were among the first cities to be connected to a railroad. The results are also not driven by a specific province. Table A3 shows they are robust to dropping one province at a time.

An advantage to using the 1911 and 1921 census is the consistent enumeration of literacy across the two years. A disadvantage is that 94% of districts are connected by 1911. Unlike Table 3, we exploit across cohort within district variation using the 1901 census in Table A6. We find positive effects of railroads for male and English literacy, although the estimates on total male and non-English male literacy are smaller in magnitude and less precisely estimated than for 1911 and 1921. Increasing railroad exposure by 14 years (the standard deviation on cohort railroad years) increases male literacy by $\frac{0.0081 \times 14.10}{1.214} = 0.09$ standard deviations and male English literacy by $\frac{0.0293 \times 14.10}{1.758} = 0.24$ standard deviations. We again find small and insignificant effects on female literacy.

Taken together, these results suggest railroads had large and positive effects on male English and non-English literacy. These results are not driven by mismeasurement of exposure in the cohort age 20 and above. They are similar across different measures of cohort exposure, across more intensely and less intensely treated cohorts in 1911-1921, and across cohorts within districts in 1901.

5.2. **Cross Section.**

5.2.1. *Ordinary Least Squares.* We turn next to cross-sectional results. Table 4 reports OLS estimates for each census year. While we report robust standard errors in this table, we show in Table A7 that our results are similar when we use Conley (1999) standard errors to adjust for spatial correlation in the error term. Columns (1) to (3) show results for log literacy with no controls in (1), including province fixed effects in (2), and including province fixed effects with the full set of controls in (3). In columns (4), (5) and (6) we report results for male, female and English literacy. Two patterns stand out. First, the estimates are positive and significant across specifications. Second, the effects are larger for female and English literacy compared to male literacy in the later years.[30]

In specifications (3) to (6) that include the controls and province fixed effects, standardised β coefficients range from 0.1 to 0.22 standard deviations, with those for English and female literacy being on the higher end of the range. For example, in column (3) for 1881, a one standard deviation increase in railroad years translates into a $\frac{0.0073 \times 9.182}{0.590} = 0.11$ standard deviation increase in literacy. By 1921 the standardised magnitude decreases to $\frac{0.0027 \times 17.62}{0.558} = 0.084$. We also find differences in effect sizes by gender. For example, the 1921 standardised coefficient for female literacy at $\frac{0.0073 \times 17.62}{0.863} = 0.15$ is larger than for male literacy at $\frac{0.0021 \times 17.62}{0.533} = 0.07$. The effect sizes for English literacy are also larger than for male at $\frac{0.0088 \times 17.62}{0.967} = 0.15$ standard deviations of English literacy.[31] Finally, the consistent estimates between the five cross-sectional years from 1881 to 1921 are reassuring in that they

[30]While we do not report results for non-English literacy in the cross section, coefficient estimates on railroad years are also positive and significant for this outcome.

[31]We computed the p-values of these coefficient comparisons and they are significantly different.

suggest that one time mortality shocks alone such as the 1917 influenza epidemic are not driving the results.

5.2.2. *Grid Cell Fixed Effects, Matching and Market Access.* Table A8 presents results comparing neighbouring districts using grid cell fixed effects. We first construct a $2^\circ \times 2^\circ$ grid using the latitude and longitude coordinates of each district's centroid. Then we include grid cell fixed effects along with the geographic and socio-economic controls. The coefficients on railroad years are positive and significant with similar sizes to the OLS estimates in the 1881 and 1891 cross-sections. Beginning in 1901, the effects sizes are larger for female and English literacy. In case of male literacy, they are smaller and less precisely estimated than the OLS estimates in 1901 and 1911.[32]

Because we have rich controls, we also undertake a matching exercise for the 1881 and 1891 cross-sections. In these years we observe both connected and unconnected districts. Table A9 reports average treatment effect on the treated (ATET) estimates using a nearest neighbour matching exercise. We measure railroad exposure as an indicator variable and match districts with railroads to those without railroads using the geographic and socio-economic controls, plus province fixed effects. We drop a few small provinces in this exercise with no variation in railroad years in 1881 and 1891. Access to railroads increases 1881 male literacy by 23% and female literacy by 44%. The coefficient on railroads for female literacy becomes smaller in size and insignificant in 1891. In results not reported here, we find no significant relationship between market access and literacy, which suggests that duration of exposure to railroads matters more that market access for schooling.

5.2.3. *Instrumental Variables.* Table 5 shows second stage instrumental variables results using the military cantonments and the 1852 Kennedy plan instruments. We show the first stage results for each year in Table A10. Columns (1)-(6) correspond to the same outcomes and controls as in Table 4. The two instruments strongly predict railroad years in each census year as seen by the large Kleibergen-Paap F-statistic (KPF). Using a Hansen test, we fail to reject the over-identification restriction in a majority of the specifications.

Our IV results confirm our previous findings: railroads positively predict literacy. In terms of magnitude, the IV estimates are largest for English literacy, female literacy next and then male literacy. For example, in standardised terms, the effects of railroad years on 1901 English literacy are $\frac{0.0376 \times 15.18}{1.056} = 0.54$ standard deviations, female literacy are $\frac{0.028 \times 15.18}{0.986} = 0.43$ standard deviations, and male literacy are $\frac{0.0128 \times 15.18}{0.538} = 0.36$ standard deviations. We find similar patterns on effect sizes in the other years. These estimates are larger than the OLS estimates reported in Table 4. These IV estimates are local average treatment effects (LATE), namely the effect of increasing railroad years for those districts that gained access

[32]We also ran robustness checks with grid cell fixed effects in our synthetic panel. We find similar effects with positive and significant effects of railroads on total, male, female and English literacy.

to railroads earlier because of their proximity to military cantonments and to the lines in the 1852 Kennedy plan. This translates into more isolated districts incidentally being connected to a railroad because they are on a direct line between major centres. It may well be such isolated places benefited more from railroads, which would account for their larger effect sizes.

Tables A11 and A12 show the results using the two instruments separately. We find similar results for the military cantonment instrument. With the Kennedy Plan instrument, we find similar results in the 1881 and 1891 cross-sections, albeit less precisely estimated for male literacy. Beginning in 1901, the estimates across the two instruments differ with positive and significant estimates using the military cantonment instrument but positive and insignificant results using the Kennedy Plan instrument. This may be related to differences in the predictive power of the instruments in those years with the KPF statistic for the military cantonment instrument (20) being double that for the Kennedy Plan (10) in those years.

5.3. **Discussion.** Are these effects big or small? To answer this question, we first benchmark our results against those in Atack et al. (2012). They estimate the effect of railroads on individual school enrolment in the United States. Their estimates suggest that increasing rail access across US counties in the 1850s predicts 56% of the increase in mean school enrolment between 1850 and 1860 (p. 16). We find smaller effects for India. In our case, increasing exposure to railroads between 1881 and 1891 predicts 16% of the actual increase in literacy.[33] It may well be infrastructure expansions have larger spillovers in more developed countries where schools were more widespread than in India.

Another way to consider the size of these estimates is in comparison to supply interventions. Chaudhary (2010b) finds it would have cost the colonial government roughly 3 rupees to make an additional person literate using causal estimates of public education spending on literacy. To construct a similar estimate for railroads, we have to monetise the increase in railway years. One crude approach is to use the change in capital outlay and working expenses between the relevant years, which we obtain from Bogart and Chaudhary (2016). This suggests an increase in railroad years between 1881 and 1891 of 6.28 years translates into 844,889,000 rupees. This increase predicts 16% of the increase in literacy between 1881 and 1891, translating into 1,394,904 additional literates. Converting this into per capita terms suggests a cost of 606 rupees to make one additional person literate. This is a simple,

[33]In this calculation, we multiply the increase in railroad years of 6.28 between 1881 and 1891 with the 1881 OLS estimate on railroad years in Table 4, column (3), to predict the increase in literacy by 1891. We then compare this predicted increase to the actual increase in literacy. The equivalent coefficient from Table 5 is larger than the OLS estimate, and increases the share of the actual literacy increase explained by railroad exposure to 23%.

illustrative back of the envelope exercise. Railroads conferred many benefits on Indian society that are not captured here. What this exercise merely shows is that railroad effects on schooling would have had to be implausibly large to be a cost-effective strategy to increase mass education.

Both the cross-sectional and synthetic panel methods, then, point in the same direction of positive and significant effects of railroads on male and English literacy. Why do we find significant results for female literacy in our cross-sectional regressions, and insignificant results in the synthetic panel? First, the local average treatment effects estimated by the two approaches may differ. For instance, the variation used in the cross-sectional estimation means that districts that were connected early to a railroad have the highest values of railway exposure. These include southern districts such as Malabar and North Arcot, where outcomes for women have traditionally been more relatively equal to those of men. The effect of the railway in these districts may have been relatively large. By contrast, in the panel, some of the cohorts that receive the most residual exposure to railroads net of district fixed effects are in regions such as Dera Ghazi Khan and Chittagong, where outcomes for women have traditionally lagged those of men. In these districts, the power of railways to increase female literacy may have been more muted.

Second, in the cross-section we allow literacy of the entire population to respond to the duration of railroad exposure, regardless of the age at which this exposure occurred; it may be the case that the effect of railroads on literacy gained in later life was greater for women. Home schooling at older ages was more common for Indian women compared to men because of norms surrounding *purda* that made it difficult for girls to attend school at younger ages. As quoted in Sharp (1914), "*zenana* teaching is carried on either by missionary agencies or by associations of Indians or by both" (p. 221). Moreover, such instruction "writes a missionary lady of experience, give an opportunity to the married and elderly people and to the widows of being able to read" (p. 221). Although it is unclear if home schooling was extensive or effective, it was used more by women than men. Unlike the cross-section, our panel regressions exploit variation in cohort exposure before the beginning of elementary school and are unable to capture this aspect of exposure to railroads, which differed by gender.

Third, statistically, fixed effects approaches like ours can exacerbate attenuation bias due to measurement error. Dyson (1989, p. 165) identifies "female age shifting into the reproductive span" in the colonial censuses; women's ages were sometimes misstated towards their main reproductive years. The number of unmarried girls aged 9-15 was understated in the 1901 census, while the number of married women aged 15-20 was also understated (Census of 1901, p. 115). Since the cross-sectional regressions study total male and female literacy, they circumvent measurement error in age enumeration. Further, the standard deviation in female literacy across cohorts within districts is only 0.6%, compared to a between standard deviation of 2.9%. This may also attenuate the synthetic panel estimates.

6. Mechanisms

In this section, we begin by documenting the proximate mechanism through which railroads increased literacy – greater school enrolment. We then provide suggestive evidence on the deeper mechanisms linking railroads to schooling. We use formal mediation models to consider the mediating effects of different channels in the cross-sectional OLS and IV frameworks. Because we cannot construct cohort-specific measures of these possible mediators, we do not use the synthetic panel framework for this exercise.

6.1. **Enrolment.** Table 6 shows the results on enrolment for the panel and cross-sectional methods. As seen in the top panel, where we include district and year fixed effects, increasing exposure to railroads has a positive and significant effect only on secondary enrolment. Indeed, the coefficient on primary enrolment is negative albeit insignificant. It would be possible for railways to increase secondary enrolment without similarly increasing primary enrolment if, for example, they had no effect on the extensive margin, but raised the continuation rate into secondary education. Given that many of the secondary schools in the data combined secondary schooling with elite primary education, these results may also reflect greater enrolment in elite primary schools. What these results rule out is the interpretation that railroads led to an increase in children attending basic vernacular primary school.

In terms of magnitude, the effects of railroad exposure increase secondary enrolment by $\frac{0.0305 \times 15.57}{0.87} = 0.55$ standard deviations in specification (6). We also find large and positive effects of railroads in the cross-sectional OLS and instrumental variable models where we use both the military cantonment and Kennedy plan instruments. The effect sizes are again smaller in the cross-sectional regressions at $\frac{0.0253 \times 15.83}{0.92} = 0.44$ standard deviations of secondary enrolment for the 1911 IV specification.[34] In comparison to literacy, these standardised β coefficients are larger for both the panel and cross-sectional models. This is unsurprising. We would expect bigger effects of railroads on the flow of children into school compared to the stock of literates because of the high drop out and wastage in the education system with many children leaving before completing 3 to 4 years of schooling, which educationists in this period argued were necessary to become literate (Parulelar, 1939).

Both the enrolment and literacy findings also point to larger effects on what, in the context of colonial India, constituted upper tail human capital. We find positive effects on secondary enrolment, which in our data sources includes higher quality primary classes and English instruction. With literacy, we find positive effects on English literacy, which was

[34] We find similar results using grid cell fixed effects.

more common among Indian elites than the rest of the population (Basu, 1974). The benefits of railroads were thus concentrated, rather than shared by the general population.

6.2. **Agricultural Income and Land Taxes.** Before railroads, transportation in India was of poor quality, expensive, and unreliable (Hurd, 1983). Railroads had a large effect on price convergence, trade, and agricultural income. According to Donaldson (2018), about half the increase in agricultural income due to the railroad came from falling trade costs. Are rising agricultural incomes a mediator from railroads to higher literacy?

We study agricultural income as a possible mediator using the series on rural income from Donaldson (2018).[35] We begin with our instrumental variables framework following Dippel et al. (2020), who offer a formal mediation analysis nested in instrumental variables. This approach identifies a total effect from a treatment (in our case railroads) to an outcome (here, education) and then decomposes it into an indirect effect via the mediating factor (agricultural income in our case) and a direct effect from treatment to the outcome not via the mediator. According to Dippel et al. (2020), this procedure requires strong instruments in two first stage regressions to identify all three effects. The standard first stage regression from instrument to treatment generates a treatment first stage F-statistic, while another first stage regression from instrument and treatment to mediator generates another mediator first stage F-statistic.

Our treatment F-statistics are greater than 10, but we have very low mediator first stage F-statistics. That is, the instrument and treatment together are not strong predictors of the mediator. Although we present results from this exercise in Table A13, we are cautious in drawing strong conclusions because of our weak mediator first stage. The coefficients on total effects for enrolment and literacy are similar to our main IV results shown in Table 5.

In Tables 7 and 8, we conduct an alternative mediation analysis suitable for an OLS framework (Imai et al., 2010, 2011).[36] As is standard in these analyses, this method relies on the sequential ignorability assumptions that railroad years are quasi-randomly assigned, conditional on the geographic and social controls, and that the mediator is ignorable conditional on railroad years and controls. Similar to enrolment and literacy, the mediators are logged in these regressions. Table 7 shows the mediation results for total literacy and Table 8 shows these results for secondary enrolment.[37] As seen in specifications (1) and (2) in the top panel, the coefficient on income is small, negative and insignificant.

In specifications (3) and (4), we also rule rule out a link from agricultural income to education via public funding. Surcharges on existing land taxes were a key funding source for rural district boards that managed rural primary education. While there could be a

[35]Our results are the same if we use nominal income per area, real income, or real income per area.

[36]In particular, we use the implementation developed by Hicks and Tingley (2011).

[37]We focus on secondary enrolment because railroads did not effect primary enrolment as seen in Table 6.

positive link in theory from railroads to agricultural income to land taxes, we find land taxes per capita are not a significant mediator for literacy or secondary enrolment. These results are unsurprising. Land taxes were fixed in nominal terms in eastern India in 1793 (Kumar, 1983). In these areas, land taxes were thus unconnected to late 19^{th} and early 20^{th} agricultural incomes. Even in other parts of the country, land taxes were revised infrequently, around every 30 years (Kumar, 1983).[38] There is no evidence that rising agricultural incomes mediate our railroad results.

We also collected data on unskilled wages to assess their relationship with railroads. Across fifty districts with reported data on agricultural wages, we find no significant correlation between the low-skilled agricultural wage and railroads in 1881 or 1891. If railroads did not increase the low-skilled wage, it is likely that substitution effects driven by rising incomes are not a significant channel from railroads to education.[39]

Finally, we consider whether our results are mechanically driven by a larger supply of railway schools in districts with early exposure to railroads. Railway companies established separate schools for children of their European and Indian employees. The East Indian Railway Company was among the first to set up schools in places where there was sufficient demand among its employees.[40] Such discussions were just beginning in 1881, so they cannot account for the 1881 results between railroads and literacy. By 1911 there were 200 such schools in British India enrolling 7,500 children of which 42% were European.[41] To put in perspective, they account for 0.16% of total schools and 0.15% of total enrolment in 1911. They are too small to affect our results. We also directly test whether the presence of Europeans is mediating our results in the next section.

6.3. **Non-Agricultural Income, Urbanisation and Returns to Education.** Apart from agricultural income, railroads may have increased non-agricultural income and urbanisation, which in turn would have increased the returns to education thus linking railroads to education. Unfortunately, we are unaware of any data sources reporting wages by location (rural/urban) and level of education or literacy. We indirectly test whether increasing returns to skill, urbanisation and rising non-agricultural incomes play a mediating role by looking at income tax revenues, urbanisation share and the share of workers in industry and services.

Income taxes were assessed on non-agricultural income using a schedule that varied by income source. Salaries and pensions for example came under one schedule, while income

[38]We show the direct correlation between railroads and potential mediators in Table A17. While railroad years are correlated with agricultural income, they are uncorrelated with land taxes.
[39]Results available upon request.
[40]Typically, parents were charged fees with some allowances for low income employees.
[41]Data from *Administration Report on the Railways in India for the Calendar Year 1911* (Government of India, 1911).

from trade, commerce and professional employment came under another schedule (Alvaredo et al., 2017). Since income from agriculture was not taxed, this measure captures income from industrial and professional employment. The share of non-agricultural workers and income taxes are both proxies, then, for returns to education. Both measures conflate the supply and demand for educated labor. However, we believe they are decent proxies for returns. Of the two, income taxes are the better proxy, because the demand for educated labor was inelastic in colonial India. Increases in labor demand would thus lead to larger changes in wages rather than in the number of workers.

We use census data to construct measures of urbanisation in 1901 and 1911. We compute the share of the population in a district living in towns with at least 5,000 persons. To do this, we use data on city populations from Fenske et al. (2021), who have shown the railways had modest positive effects on the growth of cities. And we use labor force data from Fenske et al. (2020) that construct these measures using the decennial census. Income taxes on the other hand were reported in the district gazetteers for certain years with 1901 and 1911 being the most common reporting years. Specifications (5) and (6) in the top panel of Tables 7 and 8 show that income taxes have a positive and significant coefficient for both literacy and secondary enrolment. As seen in the top panel, income taxes mediate 30% to 46% of the effects of railroads on literacy, and 25% to 43% of the effects on secondary enrolment. Rising non-agricultural incomes may have led to income effects and eased liquidity constraints, leading more families to "buy" schooling for their children.

In the bottom panel of Tables 7 and 8, we look at the urbanisation share, and the share of workers in industry and services. Similar to income taxes, urbanisation mediates between 38% and 48% of the effects on total literacy, and a smaller share of secondary enrolment at 9% to 16%. Service sector employment also appears to partially mediate the results, but less so than urbanisation and income taxes. It mediates anywhere from 6% to 16% of the effect of railroads on literacy and secondary enrolment. Lawyers and public administrators among other professionals were part of the service sector. Such workers were more educated than the rest of the population and were paid higher wages than other skilled occupations. These measures do, however, conflate income effects with rising returns to education. We have no way of disentangling these channels and interpret these results as evidence of their joint importance. In Tables A14, A15 and A16 we report the mediation results for male, female, and English literacy. Urbanisation mediates a larger share of the railroad effect for male literacy compared to female literacy. That said, these results are remarkably similar to those on total literacy and secondary enrolment.

Tables A18 and A19 summarise results for other mediators we considered. The top panel shows the correlation between mediators and railroad years, while the bottom panel shows the OLS mediation analysis. Railroads carried both goods and people. So in theory they could have increased migration, which in turn may have increased literacy. However, we find no correlation between railroads and the share of migrants, or any evidence that migration was mediating the effect from railroads to literacy.[42] Indian railroads were built by the British GOI, so it may well be that districts with a larger share of Europeans had more exposure to railroads, which in turn could account for the positive effects on literacy. Yet again, we find no significant relationship between railroads and the share of Europeans, or on the mediating role of Europeans. Christians set up missions all over India and they may have chosen to settle in districts with easy access to railroads. Indeed, we find that railroad years are positively correlated with an indicator for whether a district had a Protestant mission as of 1911. But, these missions did not play a significant mediating role between railroads and literacy or secondary enrolment. In specification (4), we check whether railroad workers are driving the service sector result. Here we subtract railroad workers from the service sector. We find similar results for non-railroad service sector workers as total service sector workers. Similarly, it is unlikely that military personnel explains our result; recruitment of soldiers on the scale of the First World War (as in Vanden Eynde (2016)) will only affect our data from 1921.

7. Conclusion

We study the effects of railroads on Indian literacy and enrolment using district-level data from 1881 to 1921. We find positive and significant effects of railroads on male and English literacy. Our results are robust in both panel models where we exploit variation in railroad exposure across cohorts within districts and in cross-sectional models where we control for the endogeneity of railroad exposure using instrumental variables. Railroads lead to greater literacy via higher secondary enrolment. We find no evidence that agriculture is an important mediator. Rather, non-agricultural income, urbanisation and service sector employment are key mediators of the link between railroads and higher schooling. Railroads generated positive spillovers on education, but their effects were concentrated and not broadly shared.

References

Adam, W. and Long, J. (1868). *Adam's reports on vernacular education in Bengal and Behar, submitted to Government in 1835, 1836 and 1838. With a brief view of its past and present condition*. Home Secretariat Press, Calcutta.

Adukia, A., Asher, S., and Novosad, P. (2020). Educational investment responses to economic opportunity: Evidence from Indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–76.

[42]Migrants are defined as people that are not born in their district of census enumeration.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, 133:375–395.

Alvaredo, F., Bergeron, A., and Cassan, G. (2017). Income concentration in British India, 1885–1946. *Journal of Development Economics*, 127:459 – 469.

Andrabi, T. and Kuehlwein, M. (2010). Railways and price convergence in British India. *The Journal of Economic History*, 70(02):351–377.

Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-1860. *Social Science History*, 34(2):171–197.

Atack, J., Haines, M., and Margo, R. A. (2011). Railroads and the Rise of the Factory: Evidence for the United States, 1850-1870. In Rhode, Paul, R.-J. and Weiman, D., editors, *Economic Evolution andRevolutions in Historical Time*, pages 162–179. Stanford University Press, Palo Alto, CA.

Atack, J., Jaremski, M., and Rousseau, P. L. (2014). American Banking and the Transportation Revolution before the Civil War. *The Journal of Economic History*, 74(4):943–986.

Atack, J., Margo, R., and Perlman, E. (2012). The impact of railroads on school enrollment in nineteenth century America. *Unpublished Working Paper, Boston University*.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in Mexico. *American Economic Review*, 106(8):2046–85.

Balagopalan, S. (2014). Sedimenting Labour through Schooling: Colonial State, Native Elite and Working Children in Early Twentieth-Century India. In *Inhabiting 'Childhood'*, pages 59–86. Springer.

Banerjee, A. and Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs, New York.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, page 102442.

Basu, A. (1974). *The growth of education and political development in India, 1898-1920*. Oxford University Press.

Björkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics*, 105:237–253.

Bogart, D. (2009). Turnpike trusts and property income: New evidence on the effects of transport improvements and legislation in eighteenth-century England. *The Economic History Review*, 62(1):128–152.

Bogart, D. and Chaudhary, L. (2016). Railways in Colonial India: An Economic Achievement? In Chaudhary, L., Gupta, B., Roy, T., and Swamy, A., editors, *A New Economic History of Colonial India*, chapter 9, pages 140–160. Routledge, London.

Bogart, D. and Chaudhary, L. (2019). Extractive institutions? Investor returns to Indian railway companies in the age of high imperialism. *Journal of Institutional Economics*, 15(5):751–774.

Broadberry, S. and Gupta, B. (2010). The historical roots of India's service-led development: A sectoral analysis of Anglo-Indian productivity differences, 1870–2000. *Explorations in Economic History*, 47(3):264–278.

Burgess, R. and Donaldson, D. (2017). Railroads and the Demise of Famine in Colonial India. *Working Paper*.

Cascio, E. U. and Narayan, A. (2022). Who needs a fracking education? The educational response to low-skill-biased technological change. *ILR Review*, 75(1):56–89.

Chandler, T. and Fox, G. (1974). *Three Thousand Years of Urban Growth*. Academic Press, New York.

Chaudhary, L. (2010a). Land Revenues, Schools and Literacy: A Historical Examination of Public and Private Funding of Education. *Indian Economic and Social History Review*, 47(1):179–203.

Chaudhary, L. (2010b). Taxation and educational development: Evidence from British India. *Explorations in Economic History*, 47(3):279–293.

Chaudhary, L. (2016). Caste, Colonialism and Schooling: Education in British India. In Chaudhary, L., Gupta, B., Roy, T., and Swamy, A., editors, *A New Economic History of Colonial India*, chapter 10, pages 161–178. Routledge, London.

Chaudhary, L. and Swamy, A. V. (2017). Protecting the borrower: An experiment in colonial India. *Explorations in Economic History*, 65:36–54.

Cheesman, D. (1997). *Landlord power and rural indebtedness in colonial Sind, 1865-1901*. Curzon Press, Richmond, U.K.

Cogneau, D. and Jedwab, R. (2012). Commodity price shocks and child outcomes: the 1990 cocoa crisis in Cote d'Ivoire. *Economic Development and Cultural Change*, 60(3):507–534.

Collins, W. J. (1999). Labor Mobility, Market Integration, and Wage Convergence in Late 19th Century India. *Explorations in Economic History*, 36(3):246 – 277.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):87–122.

Das, D. and Pathak, M. (2012). The growing rural-urban disparity in India: Some issues. *International Journal of Advancements in Research & Technology*, 1(5):1–7.

Davidson, E. (1868). *The Railways of India: with an account of their rise, progress, and construction*. E. & FN Spon.

Dimico, A., Isopi, A., and Olsson, O. (2017). Origins of the Sicilian Mafia: The Market for Lemons. *The Journal of Economic History*, 77(4):1083–1115.

Dippel, C., Ferrara, A., and Heblich, S. (2020). Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, 20(3):613–626.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Dubey, V. (1965). Railways. In Singh, V. B., editor, *An Economic History of India, 1857-1956*, pages 327–347. Allied Publishers, Bombay.

Dyson, T. (1989). The historical demography of berar, 1881-1980. In Dyson, T., editor, *India's Historical Demography*, chapter 6, pages 150–196. Curzon Press Ltd., London.

East India (1887). *(Statistical Abstracts) Statistical abstract relating to British India from 1876/7 to 1885/6*. Her Majesty's Stationary Office, London.

Edmonds, E. V. and Pavcnik, N. (2005). Child labor in the global economy. *Journal of Economic Perspectives*, 19(1):199–220.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81(3):1046–1070.

Feng, X., Porporato, A., and Rodriguez-Iturbe, I. (2013). Changes in rainfall seasonality in the tropics. *Nature Climate Change*, 3(9):811–815.

Fenske, J., Gupta, B., and Yuan, S. (2020). Demographic Shocks and Women's Labor Market Participation: Evidence from the 1918 Influenza Pandemic in India. *CEPR Discussion Paper No. DP15077*.

Fenske, J., Kala, N., and Wei, J. (2021). Railways and Cities in India. *Centre for Competitive Advantage in the Global Economy Working Paper 559/2021*.

Fetzer, T. (2020). Can workfare programs moderate conflict? Evidence from India. *Journal of the European Economic Association*, 18(6):3337–3375.

Fogel, R. W. (1964). *Railroads and American economic growth*. Johns Hopkins Press Baltimore.

Gait, E. A. (1913). *Census of India, 1911. Vol. I: India. Part I: Report*. Superintendent Government Printing, Calcutta.

Gibson, J. and Olivia, S. (2010). The effect of infrastructure access and quality on non-farm enterprises in rural Indonesia. *World Development*, 38(5):717–726.

Glewwe, P. and Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education*, volume 5, pages 653–743. Elsevier.

Gollin, D., Kirchberger, M., and Lagakos, D. (2021). Do urban wage premia reflect lower amenities? Evidence from Africa. *Journal of Urban Economics*, 121:103301.

Heath, R. and Mobarak, A. M. (2015). Manufacturing growth and the lives of Bangladeshi women. *Journal of Development Economics*, 115:1–15.

Hicks, R. and Tingley, D. (2011). Causal mediation analysis. *The Stata Journal*, 11(4):605–619.

Hindman, H. D., editor (2009). *The world of child labor: An historical and regional survey*. Routledge.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of sciences*, 107(35):15367–15372.

Hurd, J. M. (1983). *Railways*, volume 2 of *The Cambridge Economic History of India*, pages 737–761. Cambridge University Press.

Hurd II, J. (1975). Railways and the expansion of markets in india, 1861–1921. *Explorations in Economic History*, 12(3):263–288.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309–334.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789.

Jacoby, H. G. (2000). Access to markets and the benefits of rural roads. *The Economic Journal*, 110(465):713–737.

Jaworski, T. and Kitchens, C. T. (2019). National policy for regional development: Historical evidence from Appalachian highways. *Review of Economics and Statistics*, 101(5):777–790.

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *The Quarterly Journal of Economics*, 127(2):753–792.

Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*, 107(4):806–832.

Kaukiainen, Y. (2001). Shrinking the world: Improvements in the speed of information transmission, c. 1820–1870. *European Review of Economic History*, 5(1):1–28.

Kerr, I. J. (2001). *Railways in modern India*. New Delhi: Oxford University Press.

Kerr, I. J. (2007). *27 Down: New Departures in Indian Railway Studies*. Orient Longman.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Kulkarni, K. M. (1979). Cantonment towns of India. *Ekistics*, 46(277):214–220.

Kumar, D. (1983). The Fiscal System. In *The Cambridge Economic History of India, Volume 2: c. 1757-c. 1970*, pages 905–944. Cambridge University Press, Cambridge.

Kung, J. K.-S. and Ma, C. (2014). Autarky and the rise and fall of piracy in Ming China. *The Journal of Economic History*, 74(2):509–534.

Li, T. and Sekhri, S. (2020). The Spillovers of Employment Guarantee Programs on Child Labor and Education. *The World Bank Economic Review*, 34(1):164–178.

Matsuura, K. and Willmott, C. J. (2018). Terrestrial air temperature and precipitation: Monthly and annual time series (1900 - 2017) `http://climate.geog.udel.edu/~climate/html_pages/download.html#T2017`.

McAlpin, M. B. (1974). Railroads, Prices, and Peasant Rationality: India 1860–1900. *The Journal of Economic History*, 34(03):662–684.

Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Mukherjee, M. (1980). Railways and Their Impact on Bengal's Economy, 1870-1920. *The Indian Economic & Social History Review*, 17(2):191–209.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1):20–36.

Nurullah, S. and Naik, P. J. (1951). *History of Education in India during the British period.* Macmillan & Co. Ltd, Bombay, 2 edition.

Oster, E. and Steinberg, B. M. (2013). Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics*, 104:123–135.

Papers, P. (1854). *131. Railways (India).* Copy of memorandum by Major Kennedy, on the question of a general system of railways for India, referred to in the minute by the Governor-General in Council of 20 April 1853, relative to railway undertakings in that country*. Vol.XLVIII, 32 pp.*

Parulelar, R. (1939). *Literacy in India.* MacMillan & Co. Ltd., London.

Pérez, S. (2018). Railroads and the rural to urban transition: Evidence from 19th-century Argentina. *Working Paper, University of California, Davis.*

Rao, P. V. (2020). *Beyond Macaulay: Education in India, 1780-1860.* Routledge.

Redding, S. J. and Turner, M. A. (2015). Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1339 – 1398. Elsevier.

Richey, J. (1923). *Progress of Education in India 1917-1922. Vol. I: Eighth quinquennial review.* Superintendent Government Printing, Calcutta.

Sanyal, N. (1930). *Development of Indian Railways.* University of Calcutta, Calcutta.

Satya, L. D. (2008). British Imperial Railways in Nineteenth Century South Asia. *Economic and Political Weekly*, 43(47):69–77.

Sen, S. et al. (1999). *Women and labour in late colonial India: The Bengal jute industry*, volume 3. Cambridge University Press.

Shah, M. and Steinberg, B. M. (2017). Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. *Journal of Political Economy*, 125(2):527–561.

Shah, M. and Steinberg, B. M. (2021). Workfare and Human Capital Investment Evidence from India. *Journal of Human Resources*, 56(2):380–405.

Sharp, H. (1914). *Progress of Education in India 1907-1912. Vol. I.* Superintendent Government Printing, Calcutta.

Studer, R. (2008). India and the great divergence: Assessing the efficiency of grain markets in eighteenth-and nineteenth-century India. *Journal of Economic History*, 68(02):393–437.

Sztern, S. (2018). Railroads to Democracy: Individualism, Rationalism, and Literacy. The Challenge to Autocracy in Late Imperial Russia. *Paper Presented at the PEDD Conference, Muenster University, 15-17 March 2018*.

Tang, J. P. (2017). The engine and the reaper: industrialization and mortality in late nineteenth century Japan. *Journal of Health Economics*, 56:145–162.

Thorner, D. (1951). Great Britain and the development of India's railways. *The Journal of Economic History*, 11(4):389–402.

Thorner, D. (1955). The pattern of railway development in India. *The Journal of Asian Studies*, 14(2):201–216.

Vanden Eynde, O. (2016). Military service and human capital accumulation: Evidence from colonial Punjab. *Journal of Human Resources*, 51(4):1003–1035.

Walby, K. and Haan, M. (2012). Caste Confusion and Census Enumeration in Colonial India, 1871–1921. *Histoire sociale/Social history*, 45(90):301–318.

Zimran, A. (2020). Transportation and health in the antebellum United States, 1820–1847. *The Journal of Economic History*, 80(3):670–709.

8. Tables

Table 1. Summary Statistics: Cohort

	Under 10	10-15	15-20	20 & Above
	1901			
Total	0.77%	4.76%	6.63%	6.45%
Male	1.30%	7.82%	11.68%	12.04%
Female	0.25%	1.02%	1.21%	0.69%
English	0.08%	0.50%	0.84%	0.63%
English Male	0.11%	0.77%	1.43%	1.07%
English Female	0.05%	0.16%	0.20%	0.13%
Non-English	0.69%	4.25%	5.79%	5.82%
Non-English Male	1.19%	7.05%	10.25%	10.97%
Non-English Female	0.19%	0.85%	1.01%	0.56%
Cohorts Years of Railroad Exposure	21.51	18.13	14.31	10.83
	1911			
Total	0.78%	5.67%	7.63%	7.26%
Male	1.21%	8.96%	13.04%	13.09%
Female	0.33%	1.60%	1.83%	1.07%
English	0.12%	0.70%	1.20%	0.90%
English Male	0.13%	1.00%	1.99%	1.48%
English Female	0.10%	0.30%	0.36%	0.23%
Non-English	0.66%	4.97%	6.43%	6.36%
Non-English Male	1.08%	7.96%	11.04%	11.61%
Non-English Female	0.24%	1.30%	1.47%	0.83%
Cohorts Years of Railroad Exposure	30.59	26.88	22.38	18.13
	1921			
Total	1.02%	6.67%	9.42%	8.41%
Male	1.53%	10.08%	15.38%	14.77%
Female	0.51%	2.43%	2.87%	1.62%
English	0.14%	0.98%	2.07%	1.23%
English Male	0.17%	1.41%	3.25%	2.03%
English Female	0.10%	0.43%	0.80%	0.32%
Non-English	0.89%	5.69%	7.35%	7.18%
Non-English Male	1.36%	8.67%	12.13%	12.74%
Non-English Female	0.41%	2.00%	2.07%	1.31%
Cohorts Years of Railroad Exposure	40.12	36.28	31.53	26.88

Note: Literacy rates are as a percentage of the relevant population. Cohorts Years of Railroad Exposure is the number of years a railroad has operated in a district prior to the age at which the youngest member of the cohort would have regularly started elementary school.

Table 2. Summary Statistics by Cross Section

	Mean	SD	Min	Max	N
Literacy 1881	3.15%	2.43%	0.27%	17.66%	198
Literacy 1891	4.18%	3.55%	0.60%	35.23%	199
Literacy 1901	4.77%	3.33%	0.69%	24.82%	203
Literacy 1911	5.35%	3.95%	0.86%	32.13%	203
Literacy 1921	6.25%	4.62%	1.27%	41.88%	203
Male Literacy 1881	5.83%	3.99%	0.52%	30.52%	198
Male Literacy 1891	7.55%	4.80%	1.10%	35.23%	199
Male Literacy 1901	8.67%	5.26%	1.34%	35.99%	203
Male Literacy 1911	9.42%	5.83%	1.65%	42.13%	203
Male Literacy 1921	10.65%	6.53%	2.29%	50.15%	203
Female Literacy 1881	0.27%	0.72%	0.01%	6.33%	197
Female Literacy 1891	0.40%	0.99%	0.04%	8.73%	198
Female Literacy 1901	0.65%	1.41%	0.02%	11.49%	203
Female Literacy 1911	0.99%	1.97%	0.05%	16.45%	203
Female Literacy 1921	1.50%	2.62%	0.12%	24.30%	203
English Literacy 1901	0.50%	1.11%	0.00%	10.31%	203
English Literacy 1911	0.72%	1.62%	0.01%	14.20%	203
English Literacy 1921	0.99%	1.97%	0.04%	19.15%	203
Railroad Years 1881	7.58	9.18	0	28	198
Railroad Years 1891	13.86	12.62	0	38	199
Railroad Years 1901	21.51	15.18	0	48	203
Railroad Years 1911	30.59	16.65	0	58	203
Railroad Years 1921	40.12	17.62	0	68	203
Railroad Indicator 1881	52.02%	50.09%	0	1	198
Railroad Indicator 1891	73.37%	44.32%	0	1	199
Railroad Indicator 1901	87.19%	33.50%	0	1	203
Railroad Indicator 1911	93.60%	24.54%	0	1	203
Railroad Indicator 1921	96.06%	19.50%	0	1	203

Note: Literacy rates are as a percentage of the relevant population. We do not have English literacy before 1901. Railroad Years is the number of years a railroad has operated in a district.

Table 3. Synthetic Panel: Cohort, District and Year Fixed Effects

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years	0.0202***	0.0224***	0.0079
of Railroad Exposure	(0.0070)	(0.0071)	(0.0078)
Obs.	1,609	1,609	1,608
	English Literacy		
Cohort Years	0.0234***	0.0266***	0.0050
of Railroad Exposure	(0.0078)	(0.0086)	(0.0079)
Obs.	1,598	1,597	1,536
	Non-English Literacy		
Cohort Years	0.0212***	0.0235***	0.0080
of Railroad Exposure	(0.0074)	(0.0075)	(0.0081)
Obs.	1,607	1,607	1,606
Years	1911-1921	1911-1921	1911-1921

Note: Robust standard errors clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort-year level. All the regressions include fixed effects for district, cohort × province and year× province.

Table 4. Cross-Section: Ordinary Least Squares

	(1)	(2)	(3)	(4)	(5)	(6)
		Literacy		Male	Female	English
Year 1881						
Years of Railroad Exposure	0.0269***	0.0191***	0.0073***	0.0073***	0.0210***	
	(0.0047)	(0.0036)	(0.0022)	(0.0021)	(0.0047)	
Obs	198	198	198	198	197	
Year 1891						
Years of Railroad Exposure	0.0140***	0.0119***	0.0053***	0.0049***	0.0140***	
	(0.0034)	(0.0026)	(0.0018)	(0.0017)	(0.0038)	
Obs	199	199	199	199	198	
Year 1901						
Years of Railroad Exposure	0.0105***	0.0105***	0.0047***	0.0043**	0.0128***	0.0144***
	(0.0029)	(0.0028)	(0.0017)	(0.0017)	(0.0034)	(0.0035)
Obs	203	203	203	203	203	203
Year 1911						
Years of Railroad Exposure	0.0094***	0.0101***	0.0037**	0.0034**	0.0087***	0.0135***
	(0.0026)	(0.0023)	(0.0015)	(0.0014)	(0.0026)	(0.0028)
Obs	203	203	203	203	203	203
Year 1921						
Years of Railroad Exposure	0.0073***	0.0090***	0.0027*	0.0021	0.0073***	0.0088***
	(0.0025)	(0.0022)	(0.0015)	(0.0014)	(0.0026)	(0.0026)
Obs	203	203	203	203	203	203
Controls	No	No	Yes	Yes	Yes	Yes
FE	No	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table 5. Cross-Section: 1852 Kennedy Plan and Military Cantonment IVs

	(1)	(2)	(3)	(4)	(5)	(6)
		Literacy		Male	Female	English
			Year 1881			
Years of	0.0350***	0.0235***	0.0107**	0.0110**	0.0308***	
Railroad Exposure	(0.0080)	(0.0054)	(0.0044)	(0.0044)	(0.0103)	
KPF	44.78	34.99	26.56	26.56	26.55	
P-value over-id test	0.0844	0.881	0.194	0.116	0.522	
			Year 1891			
Years of	0.0216***	0.0163***	0.0091**	0.0089**	0.0256***	
Railroad Exposure	(0.0058)	(0.0048)	(0.0037)	(0.0037)	(0.0073)	
KPF	43.22	33.92	23.60	23.60	23.59	
P-value over-id test	0.0278	0.699	0.143	0.0905	0.585	
			Year 1901			
Years of	0.0187***	0.0191***	0.0127***	0.0128***	0.0280***	0.0376***
Railroad Exposure	(0.0048)	(0.0043)	(0.0042)	(0.0043)	(0.0079)	(0.0085)
KPF	36.09	29.40	18.71	18.71	18.71	18.71
P-value over-id test	0.258	0.217	0.228	0.163	0.329	0.004
			Year 1911			
Years of	0.0190***	0.0191***	0.0091***	0.0089**	0.0179***	0.0316***
Railroad Exposure	(0.0047)	(0.0041)	(0.0035)	(0.0035)	(0.0063)	(0.0076)
KPF	33.54	26.74	17.17	17.17	17.17	17.17
P-value over-id test	0.0629	0.289	0.343	0.208	0.636	0.426
			Year 1921			
Years of	0.0172***	0.0186***	0.0120***	0.0109***	0.0233***	0.0272***
Railroad Exposure	(0.0047)	(0.0042)	(0.0038)	(0.0037)	(0.0069)	(0.0071)
KPF	32.31	25.44	17.15	17.15	17.15	17.15
P-value over-id test	0.0409	0.240	0.937	0.698	0.170	0.879
Controls	No	No	Yes	Yes	Yes	Yes
FE	No	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table 6. Enrolment

	(1)	(2)	(3)	(4)	(5)	(6)
	Enrolment	Enrolment	Primary Enrolment	Primary Enrolment	Secondary Enrolment	Secondary Enrolment
Panel: District and Year Fixed Effects						
Years of Railroad Exposure	-0.0042 (0.0148)	0.0020 (0.0107)	-0.0130 (0.0173)	-0.0068 (0.0121)	0.0236* (0.0139)	0.0305*** (0.0116)
Obs	1,051	652	1,051	652	1,051	652
Year	All	1894/1897 1901/1905 /1911	All	1894/1897 1901/1905 /1911	All	1894/1897 1901/1905 /1911
Cross-Section						
1901						
Years of Railroad Exposure	0.0035 (0.0022)	0.0158*** (0.0060)	0.0028 (0.0025)	0.0123* (0.0063)	0.0115*** (0.0031)	0.0266*** (0.0085)
Obs	179	179	179	179	179	179
Model	OLS	IV	OLS	IV	OLS	IV
1911						
Years of Railroad Exposure	0.0031* (0.0017)	0.0039 (0.0046)	0.0014 (0.0018)	-0.0004 (0.0049)	0.0128*** (0.0030)	0.0253*** (0.0088)
	178	178	178	178	178	178
Model	OLS	IV	OLS	IV	OLS	IV

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table 7. Mediators: Total Literacy, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0030*	0.0037***	0.0041**	0.0046***	0.0027*	0.0022
	(0.0016)	(0.0013)	(0.0017)	(0.0014)	(0.0016)	(0.0014)
Ln(Ag Income)	-0.0180	-0.0196				
	(0.0473)	(0.0408)				
Ln(Land Taxes Per-Capita)			0.0134	0.0565		
			(0.0443)	(0.0403)		
Ln(Income Taxes Per-Capita)					0.1228***	0.1507***
					(0.0386)	(0.0300)
% of Total Effect Mediated	-0.059	-0.032	0.016	0.031	0.308	0.459
Obs	163	157	188	188	190	187
	Share Urbanisation		Share Workers in Industry		Share Workers in Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0024	0.0023	0.0032*	0.0038***	0.0027*	0.0035**
	(0.0017)	(0.0015)	(0.0017)	(0.0014)	(0.0016)	(0.0013)
Ln(Share Urbanisation)	0.0643***	0.0580***				
	(0.0112)	(0.0122)				
Ln(Shared Workers, Industry)			-0.0244	0.0790		
			(0.0656)	(0.0578)		
Ln(Shared Workers, Services)					0.2503***	0.2378***
					(0.0751)	(0.0636)
% of Total Effect Mediated	0.477	0.380	0.0103	0.0182	0.159	0.0973
Obs	203	203	187	187	187	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The outcome is log literacy in the respective year. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table 8. Mediators: Secondary Enrolment, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0132***	0.0116***	0.0106***	0.0126***	0.0065**	0.0094***
	(0.0030)	(0.0032)	(0.0032)	(0.0033)	(0.0028)	(0.0031)
Ln(Ag Income)	-0.1062	0.0826				
	(0.0898)	(0.1116)				
Ln(Land Taxes Per-Capita)			0.1494*	0.0710		
			(0.0805)	(0.0968)		
Ln(Income Taxes Per-Capita)					0.4697***	0.2570***
					(0.0659)	(0.0668)
% of Total Effect Mediated	-0.068	0.013	0.072	0.017	0.430	0.250
Obs	151	143	178	177	179	177

			Share Workers in			
	Share Urbanisation		Industry		Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0104***	0.0108***	0.0125***	0.0130***	0.0104***	0.0124***
	(0.0031)	(0.0030)	(0.0032)	(0.0032)	(0.0031)	(0.0031)
Ln(Share Urbanisation)	0.1029***	0.1636***				
	(0.0339)	(0.0352)				
Ln(Shared Workers, Industry)			0.2350*	0.1583		
			(0.1257)	(0.1355)		
Ln(Shared Workers, Services)					0.5710***	0.4118***
					(0.1409)	(0.1541)
% of Total Effect Mediated	0.091	0.156	-0.014	0.019	0.154	0.064
Obs	179	178	174	172	174	172

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The outcome is log secondary enrolment in the respective year. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

9. Figures

Figure 1. Rail Network 1881-1921

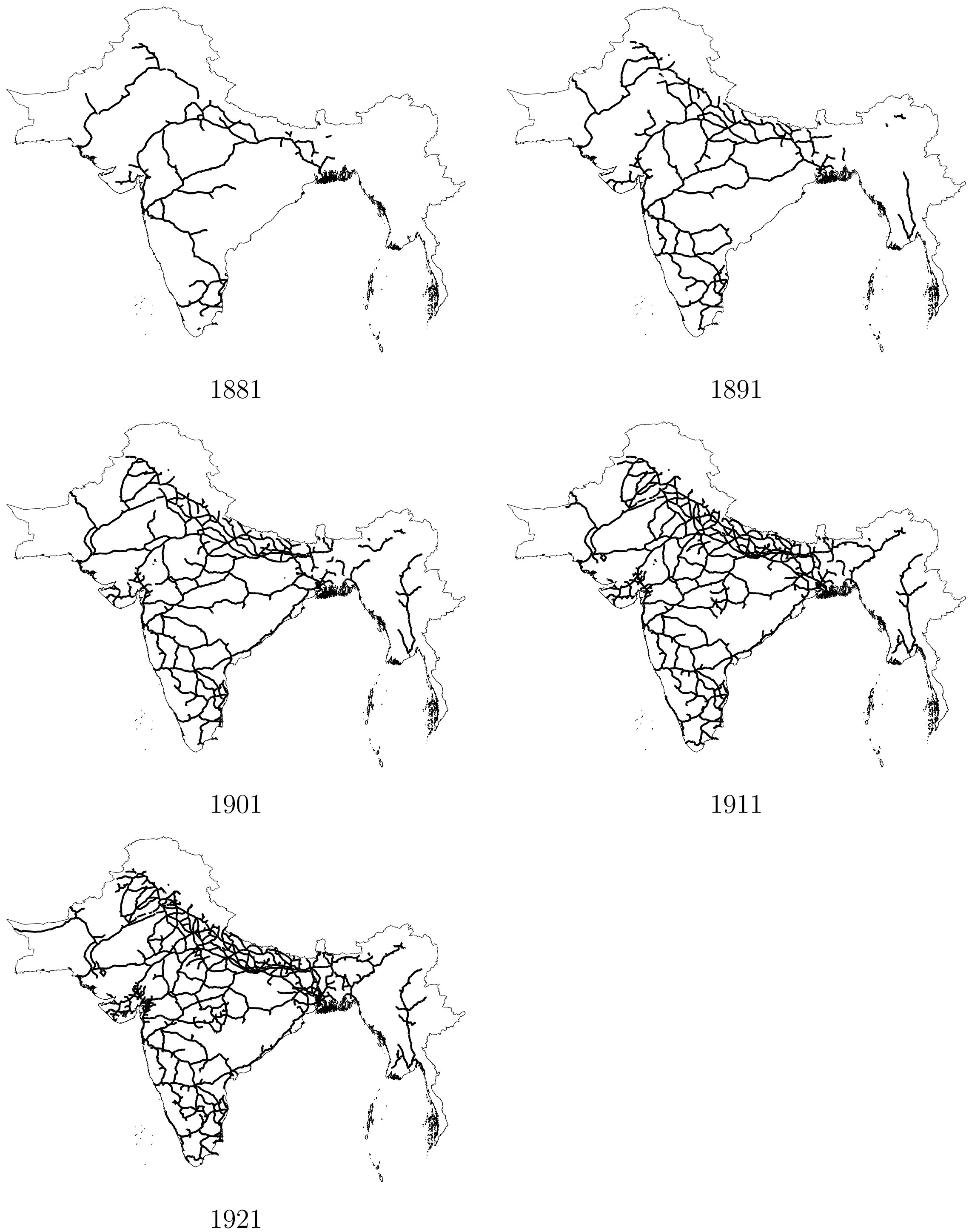
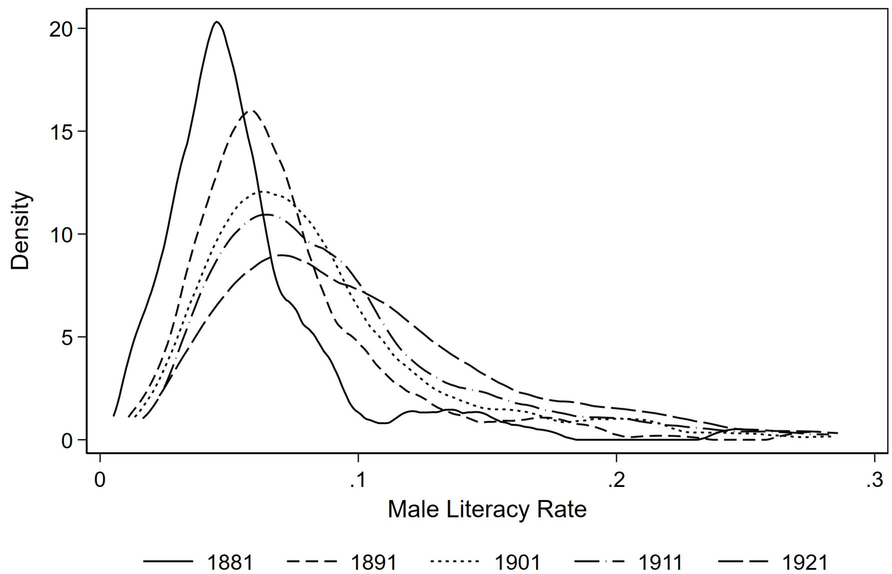


Figure 2. Distribution of Total Literacy

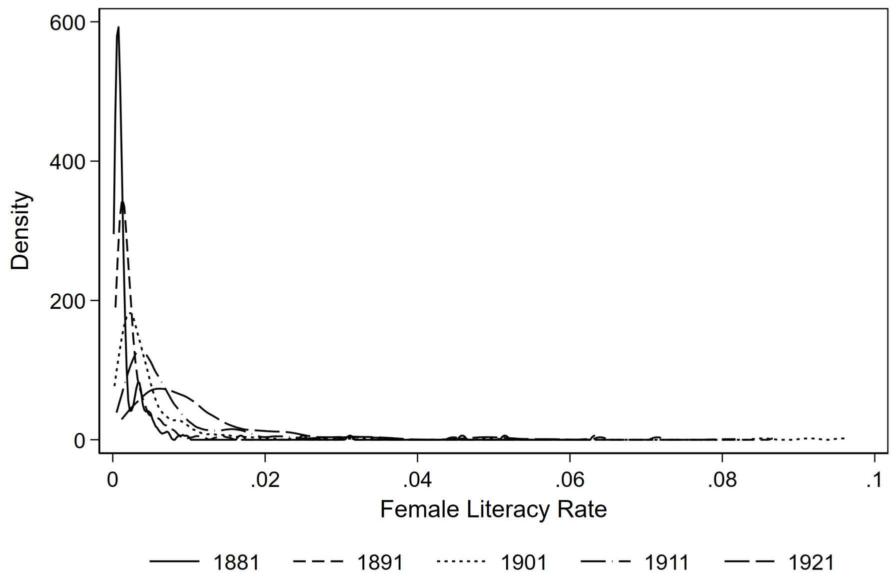
Distribution truncated at 25%. Fewer than 1% of observations are above this cutoff.

FIGURE 3. Distribution of Male Literacy



Distribution truncated at 30%. Fewer than 1% of observations are above this cutoff.

FIGURE 4. Distribution of Female Literacy



Distribution truncated at 10%. Fewer than 1% of observations are above this cutoff.

Figure 5. Map of Total Literacy, 1881-1921, Quintiles

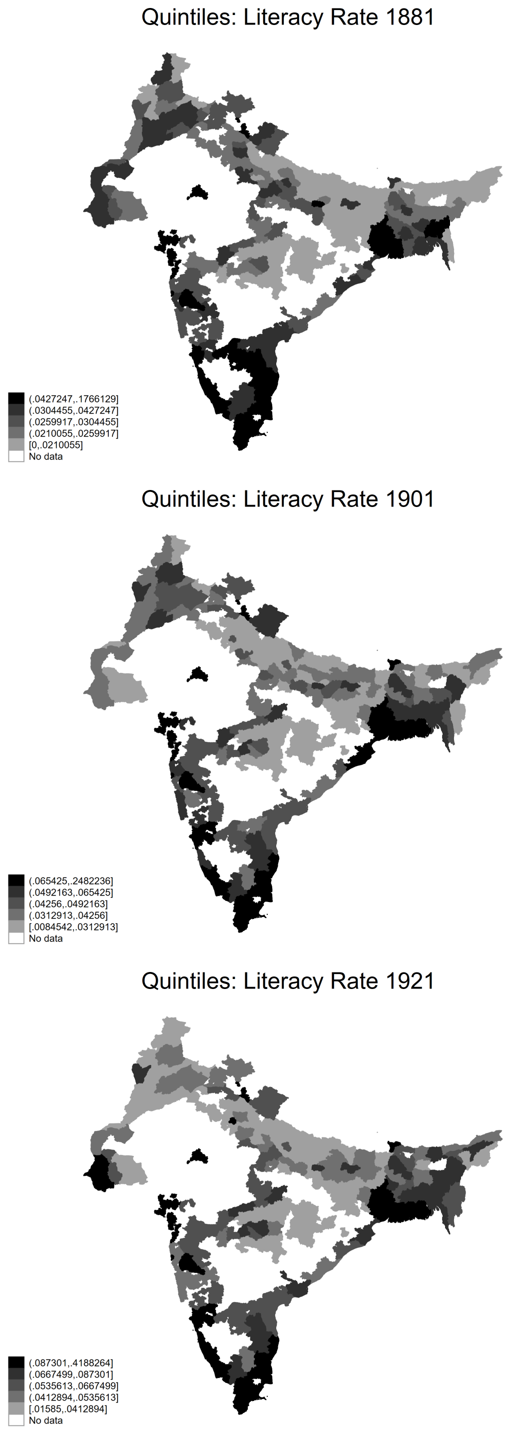


Figure 6. Scatterplot of Railroad Years and Literacy, 1881

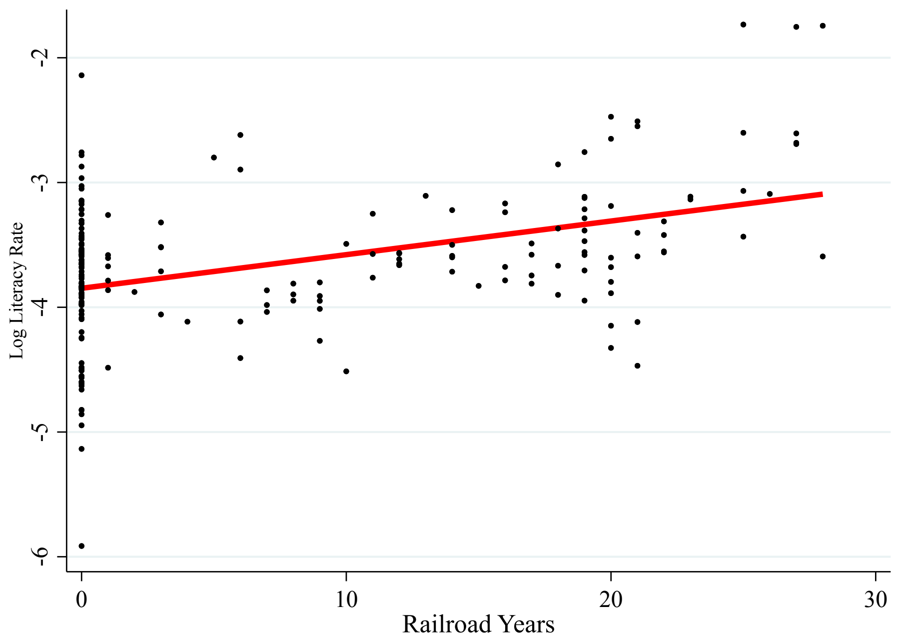


Figure 7. Scatterplot of Railroad Years and Literacy, 1921

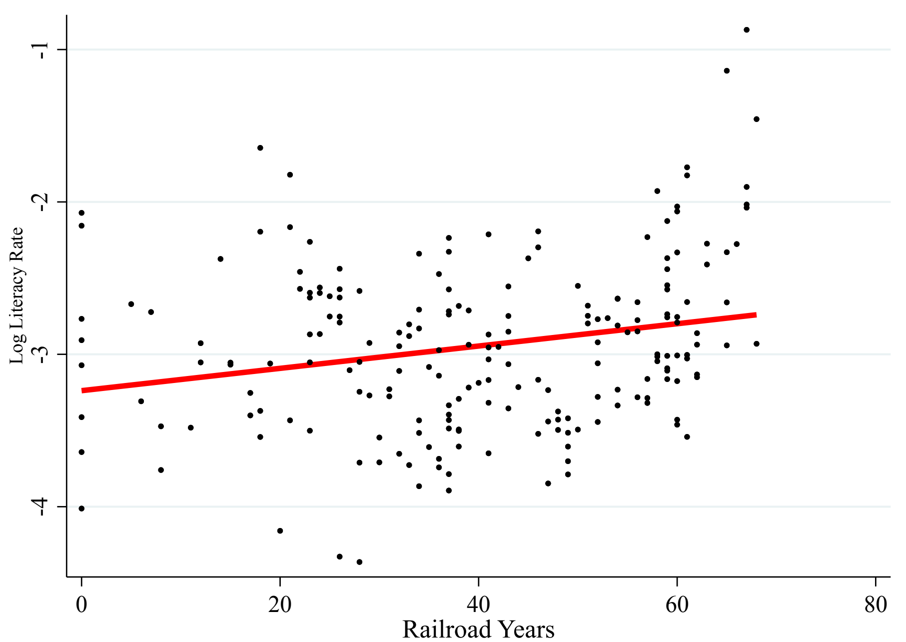
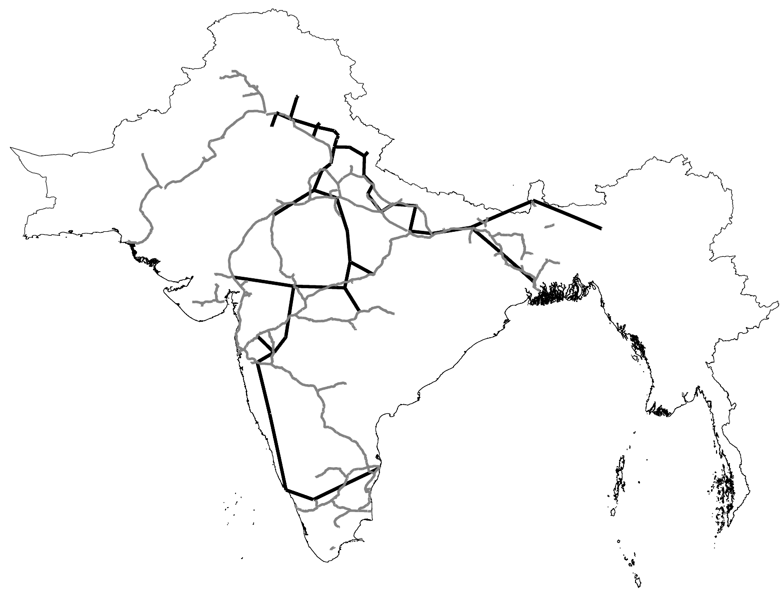
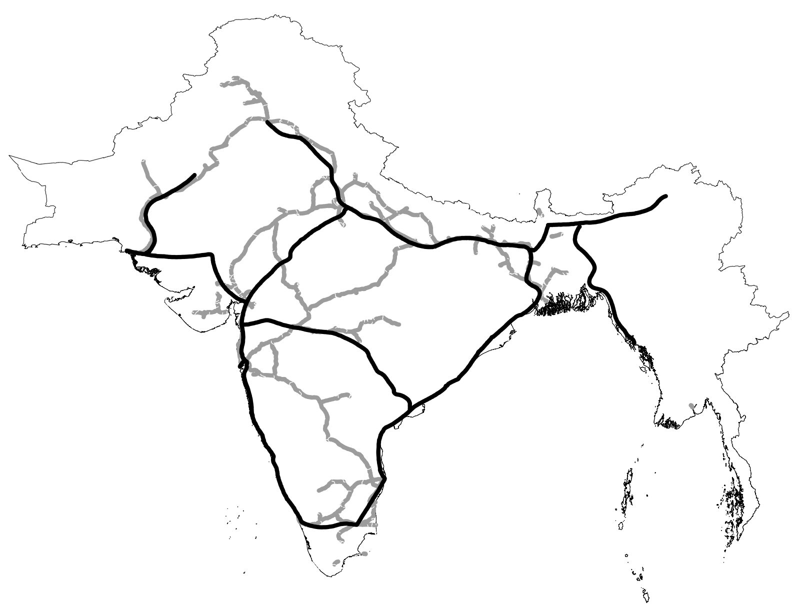


Figure 8. Map of Military Cantonment Spanning Tree



Spanning tree drawn in black. 1881 railway network drawn in grey.

Figure 9. Map of 1852 Kennedy Plan



1852 Kennedy Plan drawn in black. 1881 railway network drawn in grey.

Appendix A. Appendix Tables

Table A1. Synthetic Panel: Conley (1999) Standard Errors, Cutoff 200 km

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years	0.0202***	0.0224***	0.0079
of Railroad Exposure	(0.0058)	(0.0058)	(0.0061)
Obs.	1,609	1,609	1,608
	English Literacy		
Cohort Years	0.0234***	0.0266***	0.0050
of Railroad Exposure	(0.0061)	(0.0064)	(0.0061)
Obs.	1,598	1,597	1,536
	Non-English Literacy		
Cohort Years	0.0212***	0.0235***	0.0080
of Railroad Exposure	(0.0059)	(0.0060)	(0.0063)
Obs.	1,607	1,607	1,606
Years	1911-1921	1911-1921	1911-1921

Note: Conley (1999) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort*year level. All the regressions include fixed effects for district, cohort× province and year× province.

Table A2. Synthetic Panel: Dropping Bombay, Calcutta, Delhi and Madras

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years of Railroad Exposure	0.0191*** (0.0071)	0.0212*** (0.0072)	0.0075 (0.0079)
Obs.	1,577	1,577	1,576
	English Literacy		
Cohort Years of Railroad Exposure	0.0221*** (0.0076)	0.0252*** (0.0085)	0.0044 (0.0078)
Obs.	1,566	1,565	1,504
	Non-English Literacy		
Cohort Years of Railroad Exposure	0.0200*** (0.0075)	0.0222*** (0.0076)	0.0076 (0.0082)
Obs.	1,576	1,576	1,575
Years	1911-1921	1911-1921	1911-1921

Note: Robust standard errors clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort*year level. All the regressions include fixed effects for district, cohort× province and year× province.

Table A3. Synthetic Panel: Dropping Provinces, Total Literacy

	(1)	(2)	(3)	(4)
Cohort Years	0.0202***	0.0147**	0.0176**	0.0196**
of Railroad Exposure	(0.0070)	(0.0071)	(0.0072)	(0.0077)
Drop Province	Ajmer	Assam	Bengal	Bihar & Orissa
Cohort Years	0.0204***	0.0216***	0.0202***	0.0205***
of Railroad Exposure	(0.0078)	(0.0080)	(0.0070)	(0.0072)
Drop Province	Bombay	CP	Coorg	Madras
Cohort Years	0.0207***	0.0260***	0.0204***	
of Railroad Exposure	(0.0075)	(0.0071)	(0.0071)	
Drop Province	NWFP	Punjab	UP	

Note: Robust standard errors clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort*year level. All the regressions include fixed effects for district, cohort× province and year× province.

Table A4. Synthetic Panel: Excluding Cohort 20 and Above

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years of Railroad Exposure	0.0235***	0.0271***	0.0131
	(0.0090)	(0.0093)	(0.0109)
Obs.	1,206	1,206	1,205
	English Literacy		
Cohort Years of Railroad Exposure	0.0262	0.0305*	-0.0027
	(0.0160)	(0.0164)	(0.0131)
Obs.	1,195	1,194	1,133
	Non-English Literacy		
Cohort Years of Railroad Exposure	0.0246**	0.0285***	0.0128
	(0.0096)	(0.0099)	(0.0109)
Obs.	1,204	1,204	1,203
Years	1911-1921	1911-1921	1911-1921
Years	1911-1921	1911-1921	1911-1921

Note: Robust standard errors clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort*year level. All the regressions include fixed effects for district, cohort × province and year× province.

Table A5. Synthetic Panel: Alternative Age Band for Cohort Years

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years	0.0186**	0.0206**	0.0028
of Railroad Exposure	(0.0087)	(0.0085)	(0.0118)
Obs.	1,609	1,609	1,608
	English Literacy		
Cohort Years	0.0211**	0.0231**	0.0054
of Railroad Exposure	(0.0107)	(0.0115)	(0.0162)
Obs.	1,598	1,597	1,536
	Non-English Literacy		
Cohort Years	0.0187**	0.0207**	0.0030
of Railroad Exposure	(0.0088)	(0.0087)	(0.0125)
Obs.	1,607	1,607	1,606
Years	1911-1921	1911-1921	1911-1921

Note: Robust standard errors clustered at district level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1 The unit of analysis is log literacy at the cohort*year level. All the regressions include fixed effects for district, cohort × province and year× province.

Table A6. Synthetic Cohort: 1901 Census Year

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years	0.0066	0.0081	-0.0005
of Railroad Exposure	(0.0053)	(0.0052)	(0.0057)
Obs.	812	812	811
	English Literacy		
Cohort Years	0.0344***	0.0293***	-0.0095
of Railroad Exposure	(0.0103)	(0.0102)	(0.0104)
Obs.	809	805	746
	Non-English Literacy		
Cohort Years	0.0061	0.0070	0.0033
of Railroad Exposure	(0.0053)	(0.0055)	(0.0058)
Obs.	812	812	811
Years	1901	1901	1901

Note: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$ The unit of analysis is log literacy for the 1901 cohort. All the regressions include district and cohort × province fixed effects.

Table A7. Cross-Section: OLS with Conley (1999) SE

	(1)	(2)	(3)	(4)	(5)	(6)
		Literacy		Male	Female	English
Year 1881						
Years of Railroad Exposure	0.0269*** (0.0063)	0.0191*** (0.0034)	0.0073*** (0.0022)	0.0073*** (0.0021)	0.0210*** (0.0033)	
Obs	198	198	198	198	197	
Year 1891						
Years of Railroad Exposure	0.0140*** (0.0047)	0.0119*** (0.0029)	0.0053** (0.0021)	0.0049** (0.0020)	0.0140*** (0.0028)	
Obs	199	199	199	199	198	
Year 1901						
Years of Railroad Exposure	0.0105*** (0.0040)	0.0105*** (0.0033)	0.0047*** (0.0018)	0.0043** (0.0017)	0.0128*** (0.0029)	0.0144*** (0.0032)
Obs	203	203	203	203	203	203
Year 1911						
Years of Railroad Exposure	0.0094*** (0.0035)	0.0101*** (0.0026)	0.0037*** (0.0012)	0.0034*** (0.0012)	0.0087*** (0.0021)	0.0135*** (0.0023)
Obs	203	203	203	203	203	203
Year 1921						
Years of Railroad Exposure	0.0073** (0.0033)	0.0090*** (0.0023)	0.0027** (0.0013)	0.0021* (0.0012)	0.0073*** (0.0024)	0.0088*** (0.0028)
Obs	203	203	203	203	203	203
Controls	No	No	Yes	Yes	Yes	Yes
FE	No	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A8. Cross-Section: Grid Cell Fixed Effects

	(1) Total	(2) Male	(3) Female	(4) English
		Year 1881		
Years of Railroad Exposure	0.0054* (0.0029)	0.0055* (0.0028)	0.0178** (0.0074)	
Obs	198	198	197	
		Year 1891		
Years of Railroad Exposure	0.0044** (0.0021)	0.0038** (0.0019)	0.0174*** (0.0045)	
Obs	199	199	198	
		Year 1901		
Years of Railroad Exposure	0.0024 (0.0020)	0.0019 (0.0019)	0.0120*** (0.0043)	0.0173*** (0.0048)
Obs	203	203	203	203
		Year 1911		
Years of Railroad Exposure	0.0033** (0.0016)	0.0026 (0.0016)	0.0116*** (0.0029)	0.0176*** (0.0036)
Obs	203	203	203	203
		Year 1921		
Years of Railroad Exposure	0.0030* (0.0017)	0.0020 (0.0017)	0.0114*** (0.0030)	0.0144*** (0.0034)
Obs	203	203	203	203
Controls	Yes	Yes	Yes	Yes
FE	Grid Cell	Grid Cell	Grid Cell	Grid Cell

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A9. Matching ATET Estimates

	(1)	(2)	(3)
	Total	Male	Female
	1881		
Railroads Indicator	0.2268	0.2310*	0.4393***
	(0.1411)	(0.1194)	(0.0191)
	1891		
Railroads Indicator	0.2176***	0.2403***	0.0268
	(0.0134)	(0.0343)	(0.4327)

Note: Abadie Imbens standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We match districts using the following controls: the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; the city population recorded in Chandler and Fox (1974) circa 1850.; and province indicators.

Table A10. First Stage: 1852 Kennedy Plan & Military Cantonment Instruments

	(1) 1881	(2) 1891	(3) 1901	(4) 1911	(5) 1921
Ln (Distance from a line in Kennedy Plan)	-0.6011***	-0.6704***	-0.6346**	-0.6238**	-0.6370**
	(0.1312)	(0.1780)	(0.2042)	(0.2294)	(0.2389)
Ln (Distance from Military Cantonment)	-0.4682***	-0.6732****	-0.7695***	-0.8107***	-0.8193***
	(0.1191)	(0.1618)	(0.1835)	(0.1918)	(0.1957)
GIS controls	Yes	Yes	Yes	Yes	Yes
Religious controls	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes
KPF	26.56	23.6	18.71	17.17	17.15
Obs	198	199	203	203	203

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The dependent variable in the first stage is railroad years. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; and GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

TABLE A11. Cross-Section: Instrument Military Cantonment

	(1)	(2)	(3)	(4)	(5)	(6)
		Literacy		Male	Female	English
Year 1881						
Years of	0.0192*	0.0223**	0.0159***	0.0174***	0.0368**	
Railroad Exposure	(0.0114)	(0.0091)	(0.0061)	(0.0062)	(0.0143)	
Obs	198	198	198	198	197	
KPF	25.69	19.40	17.81	17.81	17.64	
Year 1891						
Years of	0.0099	0.0182***	0.0131***	0.0135***	0.0288***	
Railroad Exposure	(0.0074)	(0.0066)	(0.0047)	(0.0048)	(0.0093)	
Obs	199	199	199	199	198	
KPF	31.96	22.84	20.15	20.15	20.12	
Year 1901						
Years of	0.0134**	0.0259***	0.0160***	0.0167***	0.0333***	0.0573***
Railroad Exposure	(0.0064)	(0.0077)	(0.0053)	(0.0055)	(0.0093)	(0.0131)
Obs	203	203	203	203	203	203
KPF	33.70	23.30	20.01	20.01	20.01	20.01
Year 1911						
Years of	0.0099*	0.0235***	0.0118***	0.0118***	0.0171**	0.0280***
Railroad Exposure	(0.0056)	(0.0065)	(0.0044)	(0.0045)	(0.0076)	(0.0085)
Obs	203	203	203	203	203	203
KPF	35.94	24.17	19.95	19.95	19.95	19.95
Year 1921						
Years of	0.0115**	0.0240***	0.0111***	0.0116***	0.0198***	0.0358***
Railroad Exposure	(0.0058)	(0.0068)	(0.0042)	(0.0043)	(0.0070)	(0.0093)
Obs	203	203	203	203	203	203
KPF	34.70	23.34	20.11	20.11	20.11	20.11
Controls	No	No	Yes	Yes	Yes	Yes
FE	No	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A12. Cross-Section: Instrument 1852 Kennedy Plan

	(1)	(2)	(3)	(4)	(5)	(6)
		Literacy		Male	Female	English
			Year 1881			
Years of	0.0441***	0.0242***	0.0067	0.0061	0.0264**	
Railroad Exposure	(0.0103)	(0.0073)	(0.0054)	(0.0054)	(0.0121)	
Obs	198	198	198	198	197	
KPF	41.87	37.76	26.13	26.13	26.51	
			Year 1891			
Years of	0.0324***	0.0146**	0.0044	0.0034	0.0219**	
Railroad Exposure	(0.0085)	(0.0068)	(0.0050)	(0.0049)	(0.0099)	
Obs	199	199	199	199	198	
KPF	30.51	29.53	18.86	18.86	18.83	
			Year 1901			
Years of	0.0259***	0.0117*	0.0073	0.0063	0.0193	0.0046
Railroad Exposure	(0.0086)	(0.0066)	(0.0057)	(0.0058)	(0.0120)	(0.0123)
Obs	203	203	203	203	203	203
KPF	21.51	22.49	13.20	13.20	13.20	13.20
			Year 1911			
Years of	0.0300***	0.0132**	0.0052	0.0037	0.0144	0.0234*
Railroad Exposure	(0.0087)	(0.0061)	(0.0053)	(0.0052)	(0.0102)	(0.0124)
Obs	203	203	203	203	203	203
KPF	19.80	19.91	10.16	10.16	10.16	10.16
			Year 1921			
Years of	0.0293***	0.0122**	0.0123**	0.0091	0.0350***	0.0257**
Railroad Exposure	(0.0089)	(0.0060)	(0.0061)	(0.0057)	(0.0131)	(0.0128)
Obs	203	203	203	203	203	203
KPF	18.06	17.92	9.923	9.923	9.923	9.923
Controls	No	No	Yes	Yes	Yes	Yes
FE	No	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

TABLE A13. Mediator: Agricultural Income, Instrumental Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Literacy		Male Literacy		English Literacy	
	1901	1911	1901	1911	1901	1911
Total Effect	0.0218***	0.0203***	0.0217***	0.0196***	0.0302***	0.0265***
(Railroads)	(0.0069)	(0.0054)	(0.0069)	(0.0053)	(0.0114)	(0.0088)
Direct Effect	0.0351	0.0419	0.0345	0.0407	0.0430	0.0431
(Unmediated)	(0.0465)	(0.1477)	(0.0468)	(0.1462)	(0.0418)	(0.1014)
Indirect Effect	-0.0137	-0.0205	-0.0137	-0.0203	-0.0110	-0.0137
(Mediated)	(0.0348)	(0.1088)	(0.0349)	(0.1076)	(0.0294)	(0.0740)
Obs	163	157	163	157	163	157
Controls	All	All	All	All	All	All
FE	Province	Province	Province	Province	Province	Province
	Enrolment		Primary Enrolment		Secondary Enrolment	
	1901	1911	1901	1911	1901	1911
Total Effect	0.0156*	0.0105	0.0132	0.0091	0.0296***	0.0202**
(Railroads)	(0.0083)	(0.0065)	(0.0087)	(0.0065)	(0.0115)	(0.0097)
Direct Effect	0.0124*	0.0040	0.0103	0.0017	0.0257**	0.0144*
(Unmediated)	(0.0073)	(0.0073)	(0.0072)	(0.0079)	(0.0113)	(0.0084)
Indirect Effect	0.0032	0.0072	0.0027	0.0078	0.0049	0.0078
(Mediated)	(0.0108)	(0.0267)	(0.0091)	(0.0290)	(0.0164)	(0.0296)
Obs	151	143	151	143	151	143
Controls	All	All	All	All	All	All
FE	Province	Province	Province	Province	Province	Province

Note: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$ These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A14. Mediators: Male Literacy, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0022	0.0029**	0.0035**	0.0040***	0.0025	0.0019
	(0.0016)	(0.0013)	(0.0017)	(0.0013)	(0.0017)	(0.0014)
Ln(Ag Income)	-0.0085	-0.0092				
	(0.0466)	(0.0399)				
Ln(Land Taxes Per-Capita)			0.0125	0.0530		
			(0.0440)	(0.0391)		
Ln(Income Taxes Per-Capita)					0.0924**	0.1256***
					(0.0390)	(0.0293)
% of Total Effect Mediated	-0.0359	-0.0206	0.0167	0.0335	0.264	0.448
Obs	163	157	188	188	190	187
			Share Workers in			
	Share Urbanisation		Industry		Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0020	0.0019	0.0027	0.0032**	0.0022	0.0029**
	(0.0017)	(0.0014)	(0.0017)	(0.0013)	(0.0016)	(0.0013)
Ln(Share Urbanisation)	0.0633***	0.0577***				
	(0.0113)	(0.0119)				
Ln(Shared Workers, Industry)			-0.0332	0.0531		
			(0.0654)	(0.0561)		
Ln(Shared Workers, Services)					0.2258***	0.1896***
					(0.0755)	(0.0623)
% of Total Effect Mediated	0.510	0.417	0.0170	0.0140	0.163	0.0909
Obs	203	203	187	187	187	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The outcome is log secondary enrolment in the respective year. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

TABLE A15. Mediators: Female Literacy, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0108*** (0.0032)	0.0096*** (0.0024)	0.0119*** (0.0033)	0.0097*** (0.0023)	0.0090*** (0.0032)	0.0063*** (0.0023)
Ln(Ag Income)	-0.0370 (0.0938)	-0.0245 (0.0730)				
Ln(Land Taxes Per-Capita)			0.0293 (0.0868)	0.0532 (0.0691)		
Ln(Income Taxes Per-Capita)					0.3019*** (0.0752)	0.2543*** (0.0506)
% of Total	-0.034	-0.016	0.012	0.014	0.255	0.337
Effect Mediated	163	157	188	188	190	187
			Share Workers in			
	Share Urbanisation		Industry		Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0087*** (0.0032)	0.0063** (0.0024)	0.0120*** (0.0033)	0.0094*** (0.0024)	0.0106*** (0.0032)	0.0087*** (0.0022)
Ln(Share Urbanisation)	0.1163*** (0.0207)	0.0926*** (0.0201)				
Ln(Shared Workers, Industry)			0.0532 (0.1298)	0.3077*** (0.0988)		
Ln(Shared Workers, Services)					0.5649*** (0.1472)	0.5786*** (0.1065)
% of Total	0.317	0.264	-0.010	0.030	0.105	0.096
Effect Mediated	203	203	187	187	187	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The outcome is log secondary enrolment in the respective year. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A16. Mediators: English Literacy, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0099*** (0.0036)	0.0127*** (0.0027)	0.0134*** (0.0034)	0.0137*** (0.0029)	0.0109*** (0.0034)	0.0084*** (0.0026)
Ln(Ag Income)	0.0601 (0.1050)	0.1720** (0.0815)				
Ln(Land Taxes Per-Capita)			-0.0399 (0.0893)	-0.0088 (0.0843)		
Ln(Income Taxes Per-Capita)					0.2480*** (0.0791)	0.3884*** (0.0570)
% of Total Effect Mediated	0.049	0.071	-0.017	-0.003	0.188	0.367
Obs	163	157	188	188	190	187
			Share Workers in			
	Share Urbanisation		Industry		Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0097*** (0.0036)	0.0105*** (0.0028)	0.0123*** (0.0036)	0.0141*** (0.0028)	0.0110*** (0.0035)	0.0131*** (0.0025)
Ln(Share Urbanisation)	0.1326*** (0.0235)	0.1205*** (0.0231)				
Ln(Shared Workers, Industry)			0.0257 (0.1421)	0.1917 (0.1171)		
Ln(Shared Workers, Services)					0.5585*** (0.1625)	0.7219*** (0.1215)
% of Total Effect Mediated	0.319	0.221	-0.006	0.012	0.100	0.082
Obs	203	203	187	187	187	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The outcome is log secondary enrolment in the respective year. These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A17. Railroads and Mediators

	(1)	(2)	(3)	(4)	(5)	(6)
	Ag Income		Land Taxes Per-Capita		Income Taxes Per-Capita	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0091***	0.0057**	0.0053*	0.0027	0.0102***	0.0125***
	(0.0030)	(0.0027)	(0.0029)	(0.0027)	(0.0033)	(0.0034)
Obs	163	157	188	188	190	187
			Share Workers in			
	Share Urbanisation		Industry		Services	
	1901	1911	1901	1911	1901	1911
Years of Railroad Exposure	0.0353***	0.0252**	-0.0019	0.0009	0.0022	0.0016
	(0.0119)	(0.0106)	(0.0022)	(0.0018)	(0.0017)	(0.0018)
Obs	203	203	187	187	187	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
These cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely area, latitude, longitude, altitude, precipitation, slope, temperature, ruggedness, malaria transmission, seasonality, indicators for coastal districts, rivers, and medieval ports, and suitability for specific crops such as cotton, dryland rice, wetland rice, wheat and tea; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A18. Other Mediators, Total Literacy

	(1)	(2)	(3)	(4)
	Share Migrants	Share Europeans	Indicator Protestant Mission	Share Workers, Non-Rail Service
	1901	1901	1911	1911
Years of Railroad Exposure	-0.0011	-0.0000	0.0041**	0.0018
	(0.0007)	(0.0000)	(0.0017)	(0.0018)
Obs	203	203	203	187
		Total Literacy		
Years of Railroad Exposure	0.0042**	0.0048**	0.0035**	0.0034**
	(0.0018)	(0.0018)	(0.0016)	(0.0013)
Share Migrants	-0.4293***			
	(0.1507)			
Share Europeans		5.6732		
		(8.4639)		
Indicator, Protestant Mission			0.0555	
			(0.0689)	
Ln(Share Workers, Non-Rail Services)				0.2374***
				(0.0631)
% of Total Effect Mediated	0.104	-0.0251	0.0548	0.110
Obs	203	203	203	187

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely latitude, longitude, altitude, ruggedness, precipitation, distance to the coast, distance to a river and suitability for specific crops such as cotton, dryland rice, wetland rice, and wheat; and the city population recorded in Chandler and Fox (1974) circa 1850.

Table A19. Other Mediators, Secondary Enrolment

	(1)	(2)	(3)	(4)
	Share Migrants	Share Europeans	Indicator Protestant Mission	Share Workers, Non-Rail Service
	1901	1901	1911	1911
Years of Railroad Exposure	-0.0011	-0.0000	0.0041**	0.0018
	(0.0007)	(0.0000)	(0.0017)	(0.0018)
Obs	203	203	203	187
		Secondary Enrolment		
Years of Railroad Exposure	0.0108***	0.0121***	0.0129***	0.0125***
	(0.0030)	(0.0030)	(0.0033)	(0.0031)
Share Migrants	-1.3657***			
	(0.3755)			
Share Europeans		60.9846***		
		(13.4640)		
Indicator, Protestant Mission			-0.0113	
			(0.1512)	
Ln(Share Workers, Non-Rail Services)				0.3929**
				(0.1545)
% of Total Effect Mediated	0.0535	-0.0562	-0.00403	0.0536
Obs	179	179	178	172

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The cross-sectional models include province fixed effects; social controls namely the share of Brahmans, Christians, Muslims and Tribes; the GIS controls namely latitude, longitude, altitude, ruggedness, precipitation, distance to the coast, distance to a river and suitability for specific crops such as cotton, dryland rice, wetland rice, and wheat; and the city population recorded in Chandler and Fox (1974) circa 1850.