



Calhoun: The NPS Institutional Archive
DSpace Repository

Faculty and Researchers

Faculty and Researchers' Publications

2023

Railways, Development, and Literacy in India

Chaudhary, Latika; Fenske, James

University of Warwick

Chaudhary, Latika, and James Fenske. "Railways, development, and literacy in India." *Journal of Economic History* (2023).

<https://hdl.handle.net/10945/72084>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>

TITLE PAGE

Railways, Development, and Literacy in India

Latika Chaudhary
Associate Professor
Naval Postgraduate School
1 University Circle
Monterey, CA 93943
lhartman@nps.edu

James Fenske
Professor
Department of Economics
University of Warwick
Gibbet Hill Road
Coventry, CV4 7EN
United Kingdom
j.fenske@warwick.ac.uk

ACKNOWLEDGEMENTS

This paper was previously circulated as “Did Railways Affect Literacy? Evidence from India.” We are grateful to the Editor and referees for their valuable feedback. We also thank 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 TorVergata, 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.

Railways, Development, and Literacy in India

ABSTRACT

We study the effect of railroads, the single largest public investment in colonial India, on human capital. Using district-level data on literacy and two different identification strategies, we find railroads had positive effects on literacy, in particular on male and English literacy. 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.

INTRODUCTION

By 1900, the rail network in colonial India was the fourth largest in the world, covering almost 25,000 miles (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). And, 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, income, and other labor market opportunities, railways may have increased demand for schooling, even in the absence of supply-side, government interventions. Our paper asks 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 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 with 52% of British Indian districts connected to a railroad by 1881. This increased to 87% in 1901 and then 96% in 1921. Since literacy in the early censuses (1881 to 1901) cannot be compared to each other, or to later censuses, due to changes in enumeration, we use two 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.

Our first approach estimates the differential effect of exposure to railroads across cohorts within districts using the 1911 and 1921 censuses, years with comparable literacy data. The cohorts in our data are age bins for which the census reports literacy – 0 to 10, 10 to 15, 15 to 20,

and 20 and above. Since 94% of districts are connected to the railroad by 1911, we construct railroad exposure as the cumulative number of years a railroad was present in a district before the youngest member of the cohort of interest reached age 6, the start of primary school. Using this measure in a panel framework, we include district fixed effects, cohort \times province and census year \times province fixed effects. Such rich fixed effects control for time-invariant district characteristics, and provincial and national factors that affect cohort literacy flexibly over time.

Our second cross-sectional approach uses 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 an 1852 plan that predates railways construction and favors low-cost routes over gentle terrain compared to direct routes (Davidson, 1868). Our second instrument exploits military reasons for building railroads by measuring the distance between a district and a tree connecting 67 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. Our exclusion restriction assumes distance to military cantonments and the lines in the 1852 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.

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. In

our cross-sectional regressions, we find positive and significant effects of railroad exposure. Standardised coefficients suggest effect sizes ranging from 0.07 to 0.48 standard deviations depending on the measure of literacy, the census year used, and the specific statistical model. Are these effects large? Our effect sizes 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.

Why did railroads increase education? The proximate mechanism is more school enrolment. Using data on primary and secondary school enrollment, we find a one standard deviation increase in railroad exposure increases secondary enrolment by 0.2 to 0.55 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 data limitation. Many “secondary” schools were in fact elite schools with attached primary classes, and so our results may be driven by more enrolment in elite primary classes.

What deeper mechanisms link railroads with greater secondary schooling and literacy? We offer tentative answers using OLS mediation techniques (Imai et al., 2010, 2011). Past work has shown that railroads increased agricultural income (Donaldson, 2018), which in theory can increase literacy if schooling is a normal good. Yet, we find agricultural income is not a significant mediator. Rather, income taxes, urbanisation, and service sector employment are key mediators by which railroads increase literacy and enrolment. Because we cannot disaggregate the 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.

Contribution. Our paper contributes to three literatures. First, we contribute 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 (for example, 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).¹ Indian railroads have also featured in debates on colonisation. Critics argue that the financing of Indian railroads 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, railroads 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

¹ See Andrabi and Kuehlwein (2010) and McAlpin (1974) for other work on prices and income.

income volatility and mitigated the effects of famine (Burgess and Donaldson, 2017). Much of this work focuses on prices, trade, income and market integration.² 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. 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 the National Rural Employment Guarantee Act in India (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 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.

² In the case of US railroads, scholars have looked at a wider range of outcomes such as urbanisation, banking, and schooling, among others (Atack et al., 2010, 2011, 2014).

³ Aggarwal (2018) and Adukia et al. (2020) use contemporary road and highway construction projects in India to examine their effects on schooling.

HISTORICAL BACKGROUND

Literacy in British India. As British rule spread in India, former indigenous schools were either replaced or incorporated into the new colonial education system. 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 Crown took control from the East India Company, increased from 1.5% of the colonial budget in 1882 to 5.2% in 1921, a sum that was still below 1% of national income (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. We know indigenous village schools were common in the early nineteenth 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 however from these estimates because we do not know if Adam's districts were positively or negatively selected compared to the Indian average.

Beginning with the 1881 census, we know that male literacy increased from 6% in 1881 to 12% in 1921, while female literacy increased from under 1% to 2% (Census of India, 1931). Literacy increased because more children went to school. Indeed, enrolment 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.

The increase in enrolment mirrors the increase in schools. 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 publicly funded schools almost doubled from 26 to 41 per 100,000 people (Chaudhary, 2010a). Public funding was used to increase the number of schools and reduce fees. Although public primary education was not free, fees were not a significant barrier for skilled workers. For example, annual primary school fees in 1900 represented less than 0.5% of the annual wages of a skilled labourer (Chaudhary, 2016).

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 big differences in public spending across provinces, for example between the coastal provinces of Bombay and Bengal. 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 (Chaudhary, 2010a).⁴ 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 outcomes, 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%

⁴ 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 drivers of these patterns (Chaudhary, 2010a).

compared to 11% in the interior for Central Provinces and United Provinces. Apart from regional differences, caste ranked higher in the Hindu caste hierarchy were more literate and better educated. 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%.⁵

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.

Railroads in colonial India. Unlike schooling, the British were early promoters of railroads in India, building an extensive rail network.⁶ 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 railroads would lower the costs of exporting raw cotton from India and of importing British manufactured goods to new Indian markets (Thorner, 1951). 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 railroads were built by British firms with British financing, albeit subsidised by a

⁵ This discussion draws on 1931 literacy for individuals aged 10 and over from Chaudhary (2016).

⁶ There is a large literature on Indian railroads. Edited volumes by Kerr (2001, 2007) offer an excellent introduction to the main issues, while Sanyal (1930) offers a detailed history of railway development.

guaranteed dividend backed by the Government of India (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. Such 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 in 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 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 railroads, 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 cognizant 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 railroads 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. Finally, a few small lines were built connecting plains to the hill stations. While not random, the railroad network across districts was not uniformly indicative of positive or negative selection in affecting the subsequent increase in literacy. 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.

Conceptual Framework. To motivate the empirical exercise, we describe a simple framework linking railroad exposure to schooling in this sub-section. In colonial India, parents usually made the decision to send their children to school and to keep them in school. What affected this decision? Among rural households, cultivators and tradesmen sent their boys to the village primary school as did some labourers. For cultivating families, the opportunity cost of child labor was the biggest cost because children helped their families especially in the sowing and harvest seasons (Hartog (1929); Sharp (1914, 1919)). This was less of a concern for tradesmen.

There was little legal compulsion to send your child to school. Weak compulsory schooling laws were introduced in a few towns in the late 1910s, but with rare enforcement (Nurullah and Naik, 1951). Fees were common in primary schools, but as noted earlier, they were low. Scholarships were also available to defray the costs though landless families were probably less aware of such opportunities compared to tradesmen and landed cultivators (Sharp, 1914). If rural families wanted to send their children for more schooling beyond the local or

neighbouring village school, they would send them to a secondary school in a nearby town.

Some urban schools had hostels, but it was more common for children attending an urban school to stay with extended family. Fewer than 10 percent of children in urban schools stayed in a hostel.⁷

Learning to read, write and count effectively was an important skill for families engaged in trade and commerce. With basic literacy, rural men could better perform their occupational tasks. They could also work as teachers and postmen in their villages, earning wages well above those of skilled labor in most provinces (Chaudhary, 2016). Attending a secondary school led to further opportunities in the colonial bureaucracy, law, and professional employment. This was a draw for landed *zamindar* families looking to transition from rural to urban living.

Among urban households, richer families would send their children to the primary stages of middle or English high schools. Fees were higher at these schools, which nonetheless did not deter the growing demand for English education among rich and middle-class Indian families of this period (Basu, 1974).⁸ Learning English was a necessary skill to secure well-paying government, and other service sector jobs. Some girls would also attend urban schools, albeit with fewer job prospects. Teaching was a common occupation for literate women of poor means, most of whom would have been educated in a town. Indeed, girls from rural families were less likely to attend primary school, or move to the nearby town for more schooling.

How did railroads affect this household decision-making? By reducing price dispersion

⁷ See Sharp (1914) and Richey (1923).

⁸ Average annual fees at a secondary English school for boys were Rs. 12.7 in 1906 (Sharp, 1919), which accounted for around 13% of the annual wages of a skilled labourer in UP compared to 5% in Bombay using wages from Chaudhary (2016).

and increasing trade, railroads increased agricultural incomes in colonial India (Donaldson, 2018). This in turn would have generated income and substitution effects. Increasing agricultural incomes would lead families to consume more schooling if schooling is a normal good (i.e., income effect). For rural families, higher incomes would also relax credit constraints on sending children, mostly boys, to higher quality secondary schools.

Yet, rising agricultural incomes would increase the opportunity cost of child labor leading some families, in particular cultivators and landless labor, to send fewer children to school. More agricultural income could also lead to an expansion of rural primary schools if land taxes increased. But, there is no evidence of such a supply side response because land taxes did not increase in sync with agricultural incomes (Kumar, 1983). Apart from agriculture, railroads also had small and positive effects on city growth (Fenske et al., 2021). This would increase the returns to education due to agglomeration, and more service and professional employment. While agricultural and non-agricultural channels likely affected male literacy, non-agricultural channels may have been more important for female literacy.

Unlike transportation costs that fall immediately with the opening of a railway, many of these conceptual links from railroads to literacy take time to develop. Indeed, spillovers from railroads are likely to grow as exposure to railroads increase. For income effects to operate in the agricultural sector, channels such as price discovery (Aker, 2010), the formation of new links between buyers and sellers (Jensen and Miller, 2018), learning about alternative crops (Munshi, 2004), and learning about the return to education in agriculture (Foster and Rosenzweig, 1995), all involve frictions that prevent them from being immediate. They require the development of new networks and relationships. Credit constraints, similarly, would only ease gradually, because

rural credit markets characterised by high interest rates and scarcity (Nath, 2022) may be slow to transform.

Similarly, opportunities in the non-agricultural sector would increase gradually with a railway connection. Such growth requires firms to establish, relocate and grow, all of which takes time. The growth of the bureaucracy in any city connected to the railway would not be immediate. Urbanisation also would not increase the number of workers and consumers immediately. Furthermore, if railroads contribute to agglomeration effects, they would also affect city growth rates and hence the growth of urban returns to education. Such conceptual links suggest that the duration of railroad exposure is perhaps better suited to capture the effect of railroads on schooling because it allows for differential effects between a place connected to a railway for one year compared to another that is connected for ten years. A simple indicator for the presence of railroads may not capture these links.

Against this background, more exposure to colonial railroads could lead to more literacy and schooling if (1) the income effect of rising agricultural incomes exceeded the substitution effect of increasing the opportunity cost of child labor, (2) if rising incomes relaxed household income constraints of sending children to schools, (3) if railroads increased opportunities in the non-agriculture sector such as in industry and services, and (4) if rising urbanisation increased the returns to education.

DATA

We construct a new district-level dataset for British India from 1881 to 1921 to test the

relationship between railroads and education.⁹ 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.

Measuring Literacy. The colonial census reports literacy by gender and age bins. 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). Age bins 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.” Yet again, enumerators were not given official guidance on measuring literacy. Some provinces used a rigorous standard while others enumerated individuals as literate if they could sign their name (Gait, 1913). It was only in 1911 that a uniform standard, the ability to read and write a short

⁹ Colonial India encompassed British India with territories that were under direct British rule and Princely States that were governed by Indian rulers. We study only British India because the provision of education varied between British and Princely India, and among the many Princely States. In addition, the data coverage of the Princely States is incomplete and inconsistent up to 1911 (Census of India, 1901; 1911). Our analysis focuses on British Indian provinces with largely intact borders between 1881 and 1921 namely Ajmer-Merwara, Assam, Bengal, Bihar and Orissa, Bombay, Central Provinces, Coorg, Madras, North West Frontier Province, Punjab, and United Provinces.

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. To get around these issues, our panel analysis uses cohort literacy in the 1911 and 1921 censuses, when literacy was uniformly measured. Our cross-sectional analysis uses total literacy in each census year.

Appendix Table A1 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. Men were more literate than women, though this gender gap narrowed over time. English literacy was low in absolute terms, but sizeable as a share of total 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. Appendix Table A2 shows total, male, and female literacy for each cross-section, while Appendix Figure 1 shows the distribution of total, male and female literacy. While the distribution of literacy was highly skewed in 1881, it became more dispersed by 1921.

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. Rather, they are reported in many district gazetteers, which are less uniform. Nonetheless, we construct a series on primary

and secondary enrolment between 1894 and 1911, years with the most uniform data.¹⁰ This is an unbalanced panel, as most provinces report enrolment for a subset of years.

Primary school enrolment is recorded as the number of children enrolled in primary schools divided by the cohort under age 10. It averages 4% in 1901 and 1911 compared to 3% for secondary enrollment, which is children in schools other than primary schools, divided by the cohort aged 10 to 15.¹¹ A shortcoming of these data 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 increasing to more than 60% in Assam and Eastern Bengal (Sharp, 1914). 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 some measurement error in enrolment, we are unaware of district-level enrollment data for all primary school children, regardless of school type.

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

¹⁰ For 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.

¹¹ 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.

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 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.

We use the date of opening to measure the duration of railroad exposure in a district, which as we noted earlier better captures the conceptual links from railroads to schooling. We refine this idea further in our panel analysis and assume the duration of railroad years affects literacy only up to the beginning of elementary school. 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 would have finished primary school. So, the cohort 20 and above would have no exposure to railroads (coded as 0). In contrast, the arrival of railroads in 1893 would affect cohorts under age 10 in 1901 because many of them would not have presumably started elementary school as railroads arrived.

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. Our 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 at age 6.¹²

Denote this number of years since schooling began as $y(c)$. For cohorts aged 20 and above, $y(c)$

¹² We use age 6 as the beginning of elementary school because the Indian compulsory school schemes in the 1910s used age 6 as their entry point. Sharp (1914) notes primary school could begin as early as age 5.

is 14. For cohorts 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:

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

It may well be that railroads continue to affect schooling up to when a child ends elementary school. We therefore construct a second measure, which 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.¹³ 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 and assume that railroads affect literacy only up to the beginning or end of elementary school. 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.¹⁴

¹³ 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 primary school.

¹⁴ 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, for example 2 years old, may be incorrect. Using two measures of railroad 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 Appendix Table A2, 50% of districts were connected to a railroad by 1881, increasing to 96% by 1921 with much of the increase happening before 1901. The railroad years measure thus better illustrates the variation across districts. For example, the number of railroad years averaged 7.6 years across districts in 1881, increasing to 22 years in 1901, and 40 years by 1921.¹⁵

Geographic and Socioeconomic Controls. We construct a rich set of geographic variables to control for the geographical selection into railroad exposure. 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. 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

¹⁵ Apart from duration, two common methods of measuring railroad access are (1) a simple indicator 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). We discuss robustness checks using these two measures.

railroads using the population of cities enumerated in Chandler and Fox (1974) circa 1850. These cities range in population from 11,000 to 580,000 across 52 districts. This effectively controls for more urban districts that were likely to be connected with railroads before less urban districts. We also control for the religious and caste composition of a district including the share of Brahmans, traditional Hindu elites, the shares of Muslims, Christians, and tribal groups. Such shares are intended to capture historical differences in education among groups that may be correlated with railroad access. These data are taken from the colonial censuses.

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.

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

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

In this model, $LiteracyRate_{cdt}$ is literacy for cohort c in district d and census year t . We use log literacy because it is a highly skewed variable, as shown in Appendix Figure 1. 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. Such flexible controls address most measurement

concerns related to literacy and account flexibly for omitted variables at the province and cohort level that may change over time. We cluster standard errors by district to account for serial correlation over time.

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 \times cohort fixed effects. 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.

Cross-Section. We complement the panel methods with two cross-sectional models as follows.

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 in $t \in \{1881, 1891, 1901, 1911, 1921\}$:

$$\ln(\text{LiteracyRate}_{dt}) = \beta \text{RailroadYears}_{dt} + \gamma' x_{dt} + \delta_p + \epsilon_{dt} \quad (3)$$

In this equation, $\ln(\text{LiteracyRate}_{dt})$ 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. $\text{RailroadYears}_{dt}$ is the number of years district d in year t has had a railroad. This is 0 if the district is unconnected in t . The vector x_{dt} includes the GIS controls, pre-rail urbanisation, and social controls described in the previous section. We also include province fixed effects captured by δ_p . Finally, ϵ_{dt} is the error term.

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 biased upwards 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, then our estimates would likely have a negative bias. To address such endogeneity concerns, we employ instrumental variables.

Instrumental Variables: We construct two instruments for $RailroadYears_{dt}$ that exploit different sources of variation. First, we build a tree spanning the 67 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 67 military cantonments. Figure 2 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. In addition, the northern Indian plains were more vulnerable to attack as compared to hill areas, which in turn motivated the establishment of cantonments (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.

Major Kennedy was the Consulting Engineer for the GOI and pushed for building low-cost railroads that, in his view, would confer innumerable benefits “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 3 to the actual network in Figure 1. In other cases, however, more expensive routes were selected. 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}*. We report results using the two instruments in the same regression and results of each instrument used individually.

RESULTS

Synthetic Panel. Table 1 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.¹⁶

As seen in Table 1, 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 1 by the standard deviation of cohort railroad years, 17.6 years, and dividing by the relevant standard deviation of log literacy) range from $\frac{0.0202 \times 17.63}{1.250} = 0.29$ standard deviations for total to $\frac{0.0224 \times 17.63}{1.267} = 0.31$ standard deviations for male literacy in the top panel. An alternative approach to magnitudes is to consider a counterfactual in which a railroad is connected ten years earlier. This would predict literacy rates, on average, to increase by 0.20 percentage points in the 0-10 age bin, by 1.38 percentage points in the 10-15 age bin, by 1.91 percentage points in the 15-20 age bin, and by 1.76 percentage points in the 20+ age bin.

We find smaller effects on male English compared to male non-English literacy with standardised coefficients at 0.25 for English and 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.¹⁷

This exercise includes the cohort aged 20 and above. Individuals aged 20 in this cohort

¹⁶ Appendix Table A3 shows regressions using spatially adjusted Conley (1999) standard errors.

¹⁷ Our results are robust to dropping the four cities of Bombay, Calcutta, Delhi and Madras as seen in Appendix Table A4. Appendix Table A5 shows they are robust to dropping one province at a time.

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 Appendix Table A6, the results are similar, albeit with stronger results for non-English literacy compared to English literacy.

As outlined earlier, this measure of railroad exposure uses an index age in a cohort based on the beginning of elementary school at age 6. As a robustness check, a second method of constructing exposure uses an index age for a cohort based on the completion of elementary school at age 12. Appendix Table A7 shows these results. We find similar results with positive and significant effects only for male literacy. In terms of magnitude, they are marginally smaller at 0.26 standard deviations for total literacy and 0.29 standard deviations for male literacy.

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 1, we exploit across cohort within district variation using the 1901 census in Appendix Table A8. 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 0.095 standard deviations and male English literacy by 0.24 standard deviations. We again find small and insignificant effects on female literacy.

Cross Section. We turn next to cross-sectional results. Table 2 reports OLS estimates for each

census year. While we report robust standard errors in this table, we show in Appendix Table A9 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.

In specifications (3) to (6) that include the controls and province fixed effects, standardised β coefficients range from 0.08 to 0.22 standard deviations, with those for English and female literacy being on the higher end of the range. For example, the 1921 standardised coefficient for female literacy at 0.15 is larger than for male literacy at 0.069. The effect sizes for English literacy are also larger than for male at 0.16 standard deviations of English literacy.¹⁸ Finally, the consistent estimates between the five cross-sectional years from 1881 to 1921 are reassuring in that they suggest that one time mortality shocks alone such as the 1917 influenza epidemic are not driving the results.¹⁹

Table 3 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 Appendix Table A12. Columns (1)-(6) correspond to the same outcomes and controls as in Table

¹⁸ We computed the p-values of these coefficient comparisons, and they are significantly different.

¹⁹ We find positive effects of railroads in a nearest neighbor matching model for the 1881 and 1891 cross-sections (Appendix Table 10). We also calculate a district-level market access variable following Donaldson and Hornbeck (2016) and use that in place of railroad years (see Appendix Table A11). Market access is statistically uncorrelated with literacy when we include all the controls.

2. 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.²⁰

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 0.48 standard deviations, on female literacy are 0.43 standard deviations, and on male literacy are 0.41 standard deviations. We find similar patterns on effect sizes in the other years. These estimates are larger than the OLS estimates reported in Table 2. These IV estimates are local average treatment effects (LATE), namely the effect of increasing railroad years for those districts that gained access 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.

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

²⁰ Appendix Table A11 shows the first stage results corresponding to Table 3, while Appendix Tables A12 and A13 show the second stage results of using the single instrument.

between 1881 and 1891 predicts 16% of the actual increase in literacy.²¹ It may well be infrastructure expansions have larger spillovers in more developed countries.

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. A 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 401,245 additional literates. Converting this into per capita terms suggests a cost of around 2,100 rupees to make one additional person literate. This is a simple, 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

²¹ 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 2, 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 3 is larger than the OLS estimate, and increases the share of the actual literacy increase explained by railroad exposure to 23%.

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 railroad exposure. 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.

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. 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.

Third, statistically, fixed effects approaches like ours can exacerbate attenuation bias due to measurement error. One reason why literacy in general, and female literacy in particular, may be mis-measured is what Dyson (1989, p. 165) identifies as “female age shifting into the reproductive span” in the colonial censuses. Women’s ages were sometimes misstated towards their main reproductive years. Since the cross-sectional regressions study total male and female literacy, they circumvent measurement error in age enumeration. Further, the standard deviation of 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.

MECHANISMS

In this section, we document the proximate mechanism through which railroads increased literacy – greater school enrolment. We then provide suggestive evidence on the deeper mechanisms linking railroads to schooling.

Enrolment. Table 4 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 railroads 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 0.55 standard deviations in specification (6) of the panel analysis and by 0.42 standard deviations of secondary enrolment for the 1911 IV specification. 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 high drop out rates with many children leaving primary school before completing 3 to 4 years of schooling, which educationists in this period argued was necessary to become literate (Parulelar, 1939).

Agricultural Income and Land Taxes. Railroads had a large effect on price convergence, trade, and agricultural income in India (Donaldson, 2018). Are rising agricultural incomes then a mediator from railroads to higher literacy? In Tables 5 and 6, we conduct a mediation analysis suitable for an OLS framework (Imai et al., 2010, 2011). Similar to enrolment and literacy, the

mediators are logged in these regressions. Table 5 shows the mediation results for total literacy and Table 6 shows these results for secondary enrolment.²² 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 out a link from agricultural income to education via public funding. Surcharges on existing land taxes were a key funding source for district boards that managed rural primary education. While there could be a 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 because land taxes were revised infrequently in most parts of the country in the late nineteenth and early twentieth centuries (Kumar, 1983). While railroad years are correlated with agricultural income, they are uncorrelated with land taxes.²³ There is no evidence that rising agricultural incomes mediate our railroad results.

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. 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. 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

²² We focus on secondary enrolment because railroads did not affect primary enrolment as seen in Table 4.

²³ We show the direct correlation between railroads and potential mediators in Appendix Table A15.

returns to education.

Using data on income taxes from the District Gazetteers for 1901 and 1911, specifications (5) and (6) in the top panel of Tables 5 and 6 show that income taxes have a positive and significant coefficient for both literacy and secondary enrolment. They 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, we look at urbanisation, and the share of workers in industry and services. To measure urbanisation, we compute the share of the population in a district living in towns with at least 5,000 persons using census data on city populations from Fenske et al. (2021). We use labor force data from Fenske et al. (2022) that construct these measures using the decennial census. 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.

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.

While a large literature has estimated the contribution of railroads to economic growth, most studies using social savings or other methodologies ignore spillovers on other parts of the economy. To accurately assess the costs and benefits of railroads that for many countries were their single most expensive public investment of the nineteenth century, we need to account for these spillovers, positive and negative. Our findings suggest the social savings estimates of Indian railways would indeed be higher if we account for their positive spillovers on human capital.

Our findings also speak to current policy debates that pit demand against supply policies to improve schooling in developing countries. Our back of the envelope exercise illustrates that even a large and positive demand shock such as railways had a much smaller impact on Indian illiteracy, relative to its cost, than increasing public education investments.

REFERENCES

Adam, William and James Long (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*. Calcutta: Home Secretariat Press.

- Adukia, Anjali, Sam Asher, and Paul Novosad (2020). “Educational investment responses to economic opportunity: Evidence from Indian road construction.” *American Economic Journal: Applied Economics* 12(1), 348–76.
- Aggarwal, Shilpa (2018). “Do rural roads create pathways out of poverty? Evidence from India.” *Journal of Development Economics* 133, 375–395.
- Aker, Jenny C (2010). “Information from markets near and far: Mobile phones and agricultural markets in Niger.” *American Economic Journal: Applied Economics* 2(3), 46–59.
- Alvaredo, Facundo, Augustin Bergeron, and Guilhem Cassan (2017). “Income concentration in British India, 1885–1946.” *Journal of Development Economics* 127, 459 – 469.
- Andrabi, Tahir and Michael Kuehlwein (2010). “Railways and price convergence in British India.” *The Journal of Economic History* 70(02), 351–377.
- Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A Margo (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, Jeremy, Michael Haines, and Robert A. Margo (2011). “Railroads and the Rise of the Factory: Evidence for the United States, 1850-1870.” In Rosenbloom Joshua Rhode, Paul and David Weiman (Eds.), *Economic Evolution and Revolutions in Historical Time*, pp. 162–179. Palo Alto, CA: Stanford University Press.
- Atack, Jeremy, Matthew Jaremski, and Peter L Rousseau (2014). “American Banking and the Transportation Revolution before the Civil War.” *The Journal of Economic History* 74(4), 943–986.
- Atack, Jeremy, Robert Margo, and Elisabeth Perlman (2012). “The impact of railroads on school enrollment in nineteenth century America.” *Unpublished Working Paper, Boston University*.
- Banerjee, Abhijit and Esther Duflo (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. New York: Public Affairs.
- Basu, Aparna (1974). *The growth of education and political development in India, 1898-1920*. Oxford University Press.
- Bogart, Dan and Latika Chaudhary (2016). “Railways in Colonial India: An Economic Achievement?” In Latika Chaudhary, Bishnupriya Gupta, Tirthankar Roy, and Anand Swamy (Eds.), *A New Economic History of Colonial India*, Chapter 9, pp. 140–160. London: Routledge.
- Bogart, Dan and Latika Chaudhary (2019). “Extractive institutions? Investor returns to Indian railway companies in the age of high imperialism.” *Journal of Institutional Economics* 15(5), 751–774.
- Burgess, Robin and Dave Donaldson (2017). “Railroads and the Demise of Famine in Colonial India.” *Working Paper*.
- Chandler, Tertius and Gerald Fox (1974). *Three Thousand Years of Urban Growth*. New York: Academic Press.
- Chaudhary, Latika (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, Latika (2010b). “Taxation and educational development: Evidence from British India.” *Explorations in Economic History* 47(3), 279–293.

- Chaudhary, Latika (2016). "Caste, Colonialism and Schooling: Education in British India." In Latika Chaudhary, Bishnupriya Gupta, Tirthankar Roy, and Anand Swamy (Eds.), *A New Economic History of Colonial India*, Chapter 10, pp. 161–178. London: Routledge.
- Collins, William J. (1999). "Labor Mobility, Market Integration, and Wage Convergence in Late 19th Century India." *Explorations in Economic History* 36(3), 246–277.
- Conley, Timothy G (1999). "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92(1), 1–45.
- Davidson, Edward (1868). *The Railways of India: with an account of their rise, progress, and construction*. E. & FN Spon.
- Dharampala (1983). *The beautiful tree: Indigenous Indian education in the eighteenth century*. Biblia Impex.
- Dippel, Christian, Andreas Ferrara, and Stephan Heblich (2020). "Causal mediation analysis in instrumental-variables regressions." *The Stata Journal* 20(3), 613–626.
- Donaldson, Dave (2018). "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review* 108(4-5), 899–934.
- Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics* 131(2), 799–858.
- Dubey, Vinod (1965). "Railways." In V. B. Singh (Ed.), *An Economic History of India, 1857-1956*, pp. 327–347. Allied Publishers, Bombay.
- Dyson, Tim (1989). "The historical demography of Berar, 1881-1980." In Tim Dyson (Ed.), *India's Historical Demography*, Chapter 6, pp. 150–196. London: Curzon Press Ltd.
- East India (1887). (Statistical Abstracts) *Statistical abstract relating to British India from 1876/7 to 1885/6*. London: Her Majesty's Stationary Office.
- Ellis, Catriona (2020). *History of Colonial Education: Key Reflections*, pp. 1–27. Singapore: Springer Singapore.
- Feng, Xue, Amilcare Porporato, and Ignacio Rodriguez-Iturbe (2013). "Changes in rainfall seasonality in the tropics." *Nature Climate Change* 3(9), 811–815.
- Fenske, James, Bishnupriya Gupta, and Song Yuan (2022). "Demographic shocks and women's labor market participation: evidence from the 1918 influenza pandemic in India." *The Journal of Economic History* 82(3), 875–912.
- Fenske, James, Namrata Kala, and Jinlin Wei (2021). "Railways and Cities in India." *Centre for Competitive Advantage in the Global Economy Working Paper 559/2021*.
- Fetzer, Thiemo (2020). "Can workfare programs moderate conflict? Evidence from India." *Journal of the European Economic Association* 18(6), 3337–3375.
- Fogel, Robert William (1964). *Railroads and American economic growth*. Johns Hopkins Press Baltimore.
- Foster, Andrew D and Mark R Rosenzweig (1995). "Learning by doing and learning from others: Human capital and technical change in agriculture." *Journal of Political Economy* 103(6), 1176–1209.
- Gait, Edward A (1913). *Census of India, 1911*. Vol. I: India. Part I: Report. Superintendent Government Printing, Calcutta.
- Glewwe, Paul and Karthik Muralidharan (2016). "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications." In *Handbook of the Economics of Education*, Volume 5, pp. 653–743. Elsevier.

- Hartog, Philip (1929). *Interim Report of the Indian Statutory Commission: Review of Growth of Education in British India by the Auxiliary Committee Appointed by the Commission*. Calcutta, Government of India: Central Publication Branch.
- Hicks, Raymond and Dustin Tingley (2011). “Causal mediation analysis.” *The Stata Journal* 11(4), 605–619.
- Hsiang, Solomon 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, John M. (1983). “Railways,” Volume 2 of *The Cambridge Economic History of India*, pp. 737–761. Cambridge University Press.
- Hurd II, John (1975). “Railways and the expansion of markets in India, 1861–1921.” *Explorations in Economic History* 12(3), 263–288.
- Imai, Kosuke, Luke Keele, and Dustin Tingley (2010). “A general approach to causal mediation analysis.” *Psychological methods* 15(4), 309–334.
- Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto (2011). “Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies.” *American Political Science Review* 105(4), 765–789.
- Jensen, Robert (2012, 03). “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.
- Jensen, Robert and Nolan H Miller (2018). “Market integration, demand, and the growth of firms: Evidence from a natural experiment in India.” *American Economic Review* 108(12), 3583–3625.
- Jha, Saumitra (2013). “Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia.” *American Political Science Review* 107(4), 806–832.
- Kerr, Ian J (2001). *Railways in modern India*. New Delhi: Oxford University Press.
- Kerr, Ian J (2007). *27 Down: New Departures in Indian Railway Studies*. Orient Longman.
- Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs (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, Dharma (1983). “The Fiscal System.” In *The Cambridge Economic History of India, Volume 2: c. 1757-c. 1970*, pp. 905–944. Cambridge: Cambridge University Press.
- Li, Tianshu and Sheetal Sekhri (2020). “The Spillovers of Employment Guarantee Programs on Child Labor and Education.” *The World Bank Economic Review* 34(1), 164–178.
- Matsuura, Kenji and Cort J Willmott (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, Michelle Burge (1974). “Railroads, Prices, and Peasant Rationality: India 1860–1900.” *The Journal of Economic History* 34(03), 662–684.
- Mukherjee, Mukul (1980). “Railways and Their Impact on Bengal’s Economy, 1870- 1920.” *The Indian Economic & Social History Review* 17(2), 191–209.
- Munshi, Kaivan (2004). “Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution.” *Journal of Development Economics* 73(1), 185–213.

- Nath, Maanik (2022). "Credit risk in colonial India." *The Economic History Review* 75(2), 396–420.
- Nunn, Nathan and Diego Puga (2012). "Ruggedness: The blessing of bad geography in Africa." *Review of Economics and Statistics* 94(1), 20–36.
- Nurullah, Syed and Pangal Jayendra Naik (1951). *History of Education in India during the British period* (2 ed.). Bombay: Macmillan & Co. Ltd.
- Papers, Parliamentary (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.
- Parulekar, R.V. (1939). *Literacy in India*. London: MacMillan & Co. Ltd.
- Rao, Parimala V. (2020). *Beyond Macaulay: Education in India, 1780-1860*. Routledge.
- Redding, Stephen J. and Matthew A. Turner (2015). "Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity." In Gilles Duranton, J. Vernon Henderson, and William C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of Handbook of Regional and Urban Economics, pp. 1339 – 1398. Elsevier.
- Richey, J.A. (1923). *Progress of Education in India 1917-1922*. Vol. I: Eighth quinquennial review. Superintendent Government Printing, Calcutta.
- Sanyal, Nalinaksha (1930). *Development of Indian Railways*. University of Calcutta, Calcutta.
- Satya, Laxman D (2008, 22). "British Imperial Railways in Nineteenth Century South Asia." *Economic and Political Weekly* 43(47), 69–77.
- Sharp, Henry (1914). *Progress of Education in India 1907-1912*. Vol. I. Superintendent Government Printing, Calcutta.
- Sharp, Henry (1919). *Progress of Education in India 1912-1917*. Vol. I and II. Superintendent Government Printing, Calcutta.
- Sharp, Henry (1920). *Selections from Educational Records*. Calcutta: Superintendent Government Printing, India.
- Studer, Roman (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.
- Tang, John P (2017). "The engine and the reaper: Industrialization and mortality in late nineteenth century Japan." *Journal of Health Economics* 56, 145–162.
- Thorner, Daniel (1951). "Great Britain and the development of India's railways." *The Journal of Economic History* 11(4), 389–402.
- Zimran, Ariell (2020). "Transportation and health in the antebellum United States, 1820–1847." *The Journal of Economic History* 80(3), 670–709.

FIGURES

FIGURE 1. Rail Network 1881-1921



1881



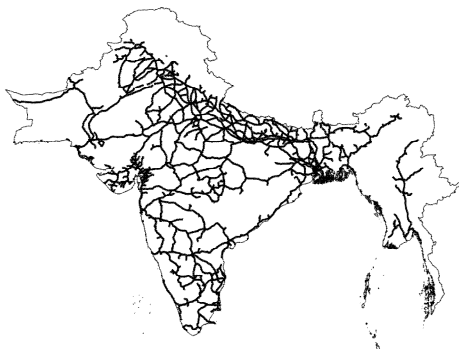
1891



1901

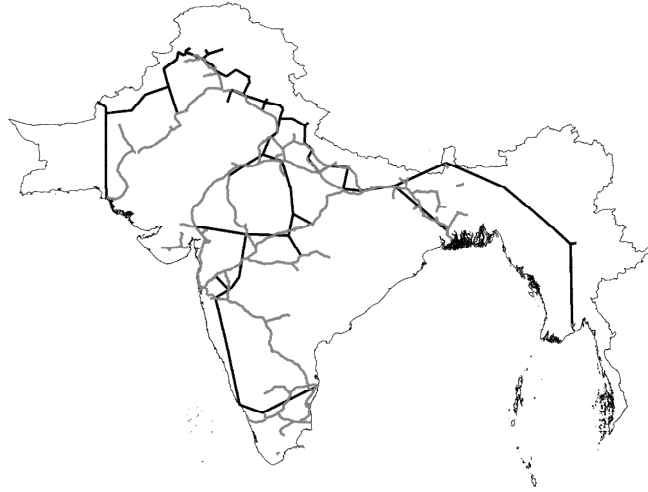


1911



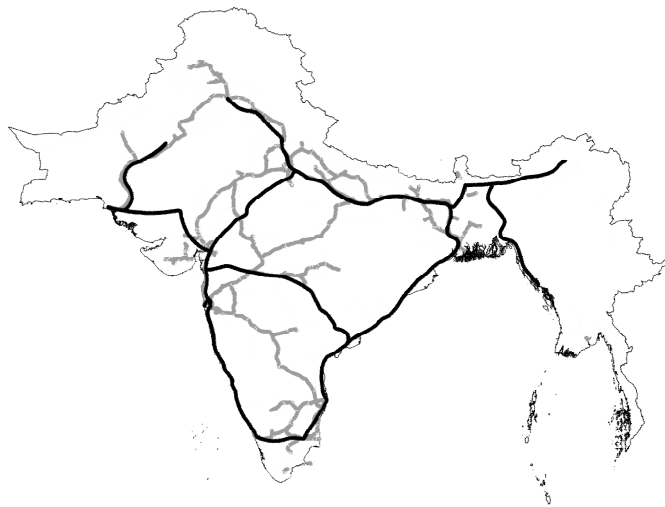
1921

FIGURE 2. Map of Military Cantonment Spanning Tree



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

FIGURE 3. Map of 1852 Kennedy Plan



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