Taxation and Educational Development: Evidence from British India

Latika Chaudhary*
Email: *latikac1@stanford.edu*

October 2008

Abstract

This paper uses a new dataset to study the effects of public educational expenditures on literacy rates across districts of British India in the early 20th century. Using an instrumental variables strategy, I find that 1911 colonial investments on primary education had positive and statistically significant effects on 1921 literacy rates in the population aged 15 to 20 controlling for province fixed effects and other observable differences across districts. A 10 percent increase in per-capita spending or 44 additional primary schools in 1911 would have translated into a 2.6 percentage point increase in the 1921 literacy rate among the population aged 15 to 20. However, there are substantial differences by gender with males being more responsive to public investments as compared to females. India's historical experience thus suggests that building more schools would not have solved the problem of female illiteracy that continues to persist even today.

*Economics Fellow, Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305. Ran Abramitzky, Stan Engerman, Christina Gathmann, Claudia Goldin, Tim Guinnane, Steve Haber, Lakshmi Iyer, Anjini Kochar, Aprajit Mahajan, Doug McKee, Bob Margo, Steve Nafziger, Jean-Laurent Rosenthal and several seminar participants provided helpful comments. I thank Bryan Yoo for research assistance and Dave Donaldson for graciously sharing his data on Indian Railways.

1. Introduction

A vast literature in economics and education has studied the effects of public educational investments on outcomes such as test scores, enrollment rates, literacy and wages.[1] This literature includes studies of US school finance reforms (Hoxby 2001),[2] government mid-day meal programs in India (Misra and Behera 2003) and school construction projects in Indonesia (Duflo 2001) with the papers exhibiting remarkable heterogeneity in terms of geographic coverage and findings that range from negative to large positive effects of school investments.[3] A key reason for the range of estimates in the literature is due to differences in the treatment of demand side conditions that may influence both outcomes and the degree of public investment.

To evaluate the causal effect of public investments, we need to understand the context through which expenditures interact with other factors that may impact schooling. Investment in education is an individual calculation based on costs and benefits, and family background characteristics such as parental income and education play a key role in the decision-making process. Given the potentially high social returns to education, states can also play a vital role in providing education especially when the private market under-supplies schooling either due to low private returns or lack of accurate information on returns.[4] Moreover, public schools can have a large impact in areas where families are credit constrained or where coordination issues between parents and children perhaps due to inter-generational commitment problems lead to sub-optimal choices. However, disentangling public expenditures from other endogenous variables remains the key empirical challenge because higher spending may be a partial response to greater private demand.

[1] See Hanushek (1986) and Glewwe (2002) along with the references therein for an overview of the key issues surrounding public expenditures in the US and developing countries respectively.
[2] Downes (1992), Murray, Evans and Schwab (1998) and Downes, Dye and McGuire (1998) are among several other studies that have evaluated US school finance equalizations.
[3] Glewwe (2002) notes that estimates from studies in developing countries range from negative to large positive effects on the order of 2 standard deviation increases in test scores due to the introduction of blackboards and repairs to school roofs in Ghana. In India, the impact of better teacher quality has been positive but small (Kingdom 1996).
[4] Schultz (1962, 1983) has written extensively about the importance of education as human capital and the numerous benefits a more literate society can confer on economic growth and development even for rural agricultural economies. Becker (1964), Mincer (1974) and Drèze and Sen (1999) among others have also written on this topic.

In this paper, I explore the effects of colonial public investments on literacy rates in British India in the early 20th century.[5] This is a particularly important and interesting setting for three reasons. First, although British India was one of the largest colonies of the British Empire, we know very little about the impact of public expenditures other than the fact that education spending was very low relative to world standards and other countries that perhaps faced similar returns.[6] It is thus unclear whether low public spending especially on primary education was a response to a perceived inefficacy of expenditures, budget constraints or conflicting priorities. Given the strong historical roots of India's current service-led economic growth (Broadberry and Gupta 2008), a better understanding of the impact of spending on outcomes has important implications for India's growth path in the colonial and post-independence period.

Second, the contemporary variation in literacy rates across different states of India bears remarkable similarities to the early 20th century. For example, in 1991 the western states of Gujarat and Maharashtra had over 60 percent literacy as compared to 38 percent in the eastern state of Bihar. Although literacy was clearly lower in the early 1900's, the regional patterns are almost the same with literacy rates in Bombay Presidency, which roughly represents contemporary Gujarat and Maharashtra, being almost twice as high as in Bihar and Orissa (11 percent versus 5 percent in 1931). How can we account for this initial variation in outcomes during the colonial period? Did public investments contribute to the differentials? Only by answering these types of questions can we begin to understand the persistence of the regional patterns into the 1990's.[7]

Third and finally, the Indian institutional setting offers a unique opportunity to disentangle

[5] British India refers to approximately two-thirds of the Indian sub-continent that was under direct colonial control. The remaining one-third of the territories (Princely States) was under the rule of native kings who deferred to the British with regard to defense and foreign policy, but managed their own local affairs. See figure 1 - Map of India.
[6] See Davis and Huttenback (1986). Whitehead (2005) emphasizes the limited quantitative literature on colonial Indian education.
[7] Development economists have extensively discussed the contemporary regional variation in outcomes within India. See Roy and Datta (1993), Drèze and Sen (1998), Probe Team (1999), Drèze and Kingdon (2001), ASER (2005), Kingdon (2007), Pal and Ghosh (2007) among others that have commented on the differences. However the historical roots of the differences are less well known. Ramachandran (1997) is one of the few studies to emphasize the history factors namely the positive role of the state in contributing to Kerala's early success in education.

the role of demand side factors from public spending. Colonial educational expenditures varied dramatically both across and within provinces of British India. Due to historical circumstances, certain provinces such as Bombay had among the highest public investments per-capita whereas other provinces such as Bengal, Bihar and Orissa received relatively low public revenues and consequently spent less on education. Within provinces, differences in land revenues were primarily responsible for differences in public spending, especially on primary education because additional surcharges on land revenues were the main source of income for rural district boards that were responsible for the provision of primary schooling. Subjective forces unrelated to education played an important role in the computation of land revenues and lead to quasi-random variation across districts controlling for objective differences in soil, rainfall and economic conditions. Land revenues thus offer an instrument for public spending because they influenced expenditures on primary education but were unlikely to be systematically correlated with unobservables driving literacy after controlling for observable differences across districts.

I use a new historical dataset for the empirical exercise, which links district-level information on public expenditures from the Indian district gazetteers to estimates of literacy from the colonial censuses. Controlling for province fixed effects and other observable differences across districts, the OLS and IV estimates find a positive and statistically significant effect of 1911 district board expenditures on 1921 literacy rates in the population aged 10 to 20 although the OLS estimates are biased upwards. While district board expenditures do not capture total public investments on education, they were an important category of spending specifically directed toward primary schooling and other public spending did not offset the variation in these expenditures.[8] The findings suggest that colonial public investments on education did in fact translate into better outcomes

[8] Provincial governments and urban municipal boards were also responsible for public provision of education. Provincial governments largely focused on secondary education and transferred grants to district boards to aid them with the provision of rural primary education, and urban municipal boards managed public education in urban towns and cities that incorporated 8 to 10 percent of the total population over this period.

namely higher literacy but only for males. After addressing potential endogeneity problems, the IV estimates of public spending are statistically correlated with male literacy and are uncorrelated with female literacy.

To address concerns regarding the validity of the instrument, I perform a series of robustness checks. First, I use different cohorts in 1921 that were unexposed to public spending in 1911 either because they were too old or too young as control outcomes. Second, I use English literacy as another control outcome because public spending on primary education should not positively influence English literacy. District board expenditures were allocated to vernacular primary education, while English medium instruction was more common in secondary schools. The IV results on expenditures are statistically uncorrelated with both the younger and older cohort-specific literacy rates as well as English literacy. Third, I subject the IV estimates to additional tests by adding controls for different land tenure systems and public expenditures on non-educational services that could perhaps be correlated with land revenues and other unobservables affecting literacy. Finally, I explore the effects of the change in spending between 1901 and 1911 on the production of literates namely the change in literacy rates between 1921 and 1911. This is akin to a quasi-first difference estimation and also finds strong and positive effects of public spending on male literacy.

Although the empirical exercise emphasizes the positive effects of increased public spending for male literacy in British India, more money alone would not have solved the problem of female illiteracy that continues to persist even today. Qualitative evidence suggests that areas with higher public spending had a larger number of better quality public primary schools that charged lower fees and effectively offered a larger public subsidy to the rural population with boys being the dominant beneficiaries. Low colonial spending however cannot account for the low levels of female literacy and India's historical experience suggests that alleviating contemporary female illiteracy may require a more nuanced policy approach that accounts for the strong social and cultural barriers to female education in India.

The remainder of the paper is organized as follows: the next section provides a brief overview of educational expenditures in the colonial period; section 3 lays out the empirical strategy; section 4 describes the district-level data; section 5 discusses the results and section 6 concludes.

2. District board expenditures and land revenues in India

This section discusses colonial public investments on primary education and argues that taxes on land (or land revenues) provide a plausibly exogenous source of variation to identify the effects of public expenditures on educational outcomes. Although public spending may partially respond to demand side factors such as returns to schooling, land revenues in this context offer an instrument for public expenditures because they were unlikely to be systematically related to the demand for education conditional on observable difference across Indian districts in the early 20th century.

During the colonial period, the East India Company followed by the British Crown introduced a new state system of education in the 19th century, which largely replaced the former indigenous system of schooling.[9] From 1858 to 1919, the Crown through the Government of India directly controlled education policy although the public financing and management of schools was decentralized to provincial governments beginning in the 1870's. The provision of primary education among other local services such as infrastructure was further decentralized to rural district and urban municipal boards in the 1880's.[10] In principle, one would like to evaluate the effects of total public educational investments but the historical sources do not always report district-level expenditures by provincial governments and urban boards and consequently, the analysis relies exclusively on spending by rural district councils. This is not a significant constraint on the empirical exercise because district boards capture a large proportion of public expenditures on rural primary education in early 20th century India.

[9] Under the former indigenous system there were elite religious schools serving a small number of students interested in a lifetime of higher learning and local elementary schools to meet the more mundane needs of the village community. See Nurullah and Naik (1951) and Basu (1982) for details.

[10] Rural district boards were constituted by Lord Ripon in 1882 and were primarily responsible for the management of primary education, local infrastructure and medical services in rural areas. See Chand (1947) and Tinker (1968) for details.

Table 1 provides a snapshot view of district board educational expenditures and their chief income sources in 1912-13 for British India as a whole and some of the individual provinces. As seen in the top panel, public educational expenditures by district councils represented 15 percent of total educational spending, 49 percent of total spending on primary education and 65 percent of public spending on primary education.[11] District boards were thus responsible for majority of the public expenditures on primary education in rural India.[12] In terms of levels, district boards spent 3.5 pounds per 1000 of the population on education, which is less than 0.5 percent of per-capita GDP in 1912-13.[13] Although expenditures increased over time, they always remained under 1 percent of per-capita GDP until Indian independence in 1947. Spending was thus very low in the colonial period.[14]

Besides the low level, there was significant heterogeneity in public spending within India with Bombay spending almost ten times as much as Bihar and Orissa.[15] Even within provinces there were important differences in spending for example with spending per-capita ranging from rupees 0.015 to 0.049 in Bihar and Orissa. As seen in the bottom panel of table 1, the spending differences were largely related to differences in revenues across districts and provinces. Additional surcharges on land revenues known as cesses represented 60 percent of revenues of district boards in the 19th century, but transfers from provincial governments to district boards in the form of grants especially targeted toward primary schooling increased in the early 20th century and by 1929-30 they

[11] The estimate of district board spending as a proportion of total public spending on primary education is biased downwards because of differences in reporting years between the historical sources with public primary education expenditures reported for 1916-17 (Progress of Education in India, 1923), while the data on district board spending is from 1912-13 (Statistical Abstracts of India). The Statistical Abstracts also severely undercount the extent of district board spending in the southern province of Madras since they appear to have excluded spending by sub-district boards ('taluk boards') that were in fact a part of district boards. For the empirical analysis, I use district level data obtained from the Indian District Gazetteers that report total spending by district councils including the lower taluk boards.

[12] In the small number of districts without boards, the provincial government managed the provision of public primary education and in less than 10 percent of urban India, urban municipal boards managed primary schooling.

[13] This holds for the range of GDP estimates reported in Kumar, ed. (1982).

[14] In fact, public investments on human capital in British India were among the lowest in the world (Davis and Huttenback 1986) as compared to comparable areas such as the Indian Princely States that spent twice as much on education and other foreign underdeveloped countries that spent five times as much.

[15] The differential patterns in spending are not peculiar to 1912-13 and hold for most of the colonial period. See Kumar, ed. (1982) and Statistical Abstracts of India.

contributed 43 percent to income as compared to 36 percent from the cesses.[16] Cesses and grants jointly accounted for 75 percent of the income of district boards, while the rest came from tolls, school fees and other miscellaneous sources.

Differences in land revenues contributed to significant inter-provincial differences in grants that are included under 'other sources' in the bottom panel of table 1. On account of the Permanent Settlement of 1793 in Bengal and Bihar, land revenue was fixed in cash for perpetuity and so these areas were unable to recover higher revenues as agricultural productivity and prices increased over the 19th and 20th century. Consequently, when land revenues were decentralized to the provinces in the late 19th century, they had lower revenues and were limited in the amount of money they could transfer to district boards via grants.[17] In comparison, land revenues were revised every thirty years in Temporary Settlement areas such as Bombay and Madras to account for changing agricultural conditions and they received higher public revenues.[18] These provinces thus had more public money for grants spending approximately 10 pounds per 1000 of the population.

Within provinces, grants were generally allocated to districts often based on population and sometimes with a stated preference for favoring poorer ('needy') districts or those with larger minority populations.[19] Although land revenues did not significantly influence the distribution of grants within provinces, they contributed to differences in cesses both across and within provinces

[16] Calculations based on Chand (1947).
[17] Areas under the Permanent Settlement were also correlated with the *zamindari* system of land tenure whereby landlords were made responsible for revenue payments as compared to Temporary Settlement districts where individual cultivators or village bodies were responsible for revenue payments. See Baden-Powell (1907, 1972) and Banerjee and Iyer (2005) for details on different land tenure arrangements.
[18] Land revenues in Bombay and Madras were assessed at higher rates as compared to the other Temporary Settlement provinces and hence they had among the highest spending during the colonial period. See Chand (1930, 1931), Misra (1960) and Kumar, ed. (1982) for details on the colonial fiscal system.
[19] See Quinquennial Review of Education (1902-07). Chaudhary (2008) finds a positive correlation between the number of primary schools directly managed by local boards and the population share of minorities such as lower castes and tribes over the 1901 and 1911 decade. According to Chand (1947) some of the earlier grants were distributed based on how much money a district raised (i.e. cess) but over time the rhetoric favored giving more money to 'needy' districts. Some provinces did not have an official policy on how grants were distributed to districts but the empirical analysis in Chaudhary (2008) supports the idea that minority districts did receive more money on average to set up schools.

because cesses were just additional taxes on land revenues. Cesses were usually levied at the rate of 6.25 percent on land revenues. The rate was set at the province level and hence districts within the same province faced the same rate.[20]

Since land revenues had been fixed in Permanent Settlement districts in 1793, the cess assessment in these areas was based on the annual value defined as the "rent paid by the tenant to the landlord" (Chand 1947, page 118). Surveys were conducted to measure annual values, but they were often outdated and inaccurate, and in practice it appears that cesses in Permanent Settlement areas were just based on the notoriously inaccurate land revenue assessments of 1793—the correlation between 1911 per-capita cesses and land revenues is 0.9 in Permanent Settlement districts.[21] Since cesses in Permanent Settlement districts were based on idiosyncratic land assessments conducted in 1793, they should be unrelated to factors that may influence the demand for education almost 150 years later in 1911 or 1921.

In Temporary Settlement districts, the cesses were directly assessed on land revenues that were calculated as a proportion (roughly 50 percent) of monetized net agricultural produce in Madras, a function of general conditions adjusted for different soil types in Bombay and roughly 50 percent of rental assets in the United Provinces and Punjab.[22] Land revenue rates were high in the early to mid-19th century but declined over the late 19th and 20th century and represented less than 5 percent of net agricultural produce by the 1930's (Roy 2000 and Kumar, ed. 1982).[23] The

[20] For example if a district owed 30,000 rupees in land revenues, the cess amount would equal 1875 rupees (0.0625*30,000).

[21] See Chand (1947) for details on cesses in Bengal. In particular, land revenue assessments in Permanent Settlement Bengal were not based on "any area survey, any consideration, that is, of the number, various fertility, or productive power, of the acres held in each case, or of the influence of proximity to markets and facility of communication, on the value of produce" (Baden-Powell 1972, vol.1, page 287). They were largely a function of historical revenues and negotiations between Indian landlords and the East India Company. See Baden-Powell (1972) and Kumar, ed. (1982) for details.

[22] The cess was assessed on double the land revenue in the United Provinces and Punjab. See Baden-Powell (1972) and Chand (1931) for details on the assessment of land revenues in each province. Kumar (1982) highlights the random heterogeneity of the initial 19th century revenue assessment in Madras.

[23] Tax rates were generally the same within provinces and so differences in land assessments were the main reason for differences in land revenues across districts with the same province.

assessments were revised every thirty years making them relatively insensitive to changes in economic and agricultural conditions as well as inelastic to increases in population. Detailed cadastral surveys were conducted at the reassessments, but subjective forces strongly influenced the calculations with the system "lacking in precision and objectivity."[24] Soil type, rainfall and the general level of development were some of the objective measures that influenced the calculations.

While the literature on Indian economic history has discussed potential links between the revenue assessment rates and Indian agricultural distress, the relative inaccuracy of the system has received little attention although both contemporaries of the period and subsequent writers were cognizant of the problems.[25] Settlement officers set the revenue demand at each reassessment and the "idiosyncrasies of the Settlement Officers" lead to substantial quasi-random variation in land revenues both across and within provinces (Chand 1931, 1947). For example, "according to Mr. Mackee, the Settlement Commissioner and Director of Land Records, the pitch of assessment varies in the most unreasonable fashion even from group to group in the same taluka and from district to district, so that the assessment may be half the rent in one place and only one-fifth in another" (Chand 1931, page 61). Baden-Powell's detailed description of the land settlement process also acknowledges the role of "intuitive calculation" and states "with all these different methods, it is apt to be supported that, after all, Settlement is very much a matter of individual taste and opinion" (Baden-Powell 1972, vol. I, page 338) while defending the assessment of land revenues.

While Chand (1931) highlights the problems of assessments based on general considerations such as in Bombay, even in provinces such as United Provinces where rental values formed the basis of assessment, the computation was often random because rents were not readily observed and "there

[24] Chand (1947), page 119. The surveys involved demarcation of boundaries, a record of land holdings and rights, and an assessment of the soil conditions. See Baden-Powell (1907, 1972) and Chand (1931, 1947).

[25] The discussion in Chand (1931) suggests that the marked diversity and inaccuracy of the system was realized only after official evidence was collected by the Taxation Enquiry Committee of 1924-25, which highlighted that practices on the ground were very subjective relative to theory. Chand (1931) reviews the revenue assessment basis in each province and emphasizes the general inaccuracy and randomness of the system. Baden-Powell (1972) reviews the revenue assessment process in each province and although he defends the system his discussion also suggests that Settlement officers along with objective factors such as soil, etc. determined the final assessment.

must necessarily be a point where estimation–guess-work if the term is preferred—comes in (Baden-Powell 1972, vol. I, page 339).[26] Even when rents were directly observed, they only captured the value at the time of the settlement with subsequent rents becoming "largely matters of agreement between landlord and tenants" (Stokes in Kumar, ed. 1982).

This is not to suggest that objective factors did not influence land revenue assessments. Soil type, rainfall and the degree of development were common variables that appeared in revenue assessment discussions.[27] However, the key point is that the subjectiveness of the assessments does not appear to be correlated with education after controlling for differences in objective factors such as soil, rainfall and other socio-economic differences across districts. Anecdotal evidence in Baden-Powell (1972) and Chand (1931) suggests that the subject of schools and education never factored into assessment calculations. Land revenues thus offer an instrument for district board educational expenditures because they are correlated with board revenues (i.e. cesses) but are unlikely to be correlated with other unobservables that may affect educational outcomes.

This overview suggests that colonial policy created marked differences in public investments on primary education both across Indian provinces and districts. While differences in Land Settlements between provinces such as Bombay (i.e. Temporary Settlement) versus Bengal (i.e. Permanent Settlement) lead to extreme disparity in inter-provincial patterns of public spending (i.e. grants), the idiosyncratic computation of land revenues within provinces created quasi-random variation in educational expenditures across districts in the same province. I use this quasi-random variation in the empirical analysis to identify the effects of colonial public investments on male and female cohort specific literacy rates.

[26] Chand (1931) also notes that "much depends upon the personal judgment of the officers in charge of the work of assessment or reassessment" (page 60).
[27] See Baden-Powell (1972) for details. He highlights that soil classification was a very important criterion for assessment in all provinces other than United Provinces and Punjab where rents formed the basis of assessment.

3. Empirical Framework

Although public educational expenditures accounted for 50 percent of total educational spending on primary education in British India, no study has systematically evaluated whether higher spending in the colonial period lead to better educational outcomes namely higher enrollment or greater literacy. In fact, British administrators took a dim view of the potential efficacy of public investments and official rhetoric often emphasized demand-side issues as the primary driver of educational outcomes (Nurullah and Naik, 1951).[28] However, this position may partially reflect official strategy to absolve the colonial Government from any blame for the relatively low levels of public investments. On the other extreme Congress leaders and Indian nationalists bemoaned the low level of public spending and strongly advocated higher expenditures as the key to better outcomes.

This disagreement on the precise role of educational expenditures in British India mirrors the lack of consensus on the causal effect of expenditures both within the US literature and among studies that focus on developing countries. A key obstacle to identifying the effects of spending is related to confounding factors such as income and development that may be positively correlated with both higher public investments and better educational outcomes. Endogeneity concerns are particularly relevant for an outcome such as literacy that is more likely to be correlated with demand (i.e. returns to education) and supply (availability of schools and teacher quality).

At an individual level, the decision to invest in schooling or literacy involves a simple cost and benefit calculation. Both individual characteristics namely ability, family background, parental education, social and religious affiliation along with community or district characteristics such as economic conditions and public educational investments factor into the decision-making process. Outcomes are thus an aggregate measure of the underlying individual decisions assuming every individual who invests in education or literacy becomes literate. In the absence of historical

[28] See Progress of Education in India, Quinquennial Reviews (1886-1937) that highlight official views on educational expenditures with statements such as "There is little to be said for opening schools to which parents will not send their children" (Eighth Quinquennial Review, page 123) without compulsory education or greater demand.

information on test scores or enrollment rates, I focus on literacy as the outcome variable because it offers a reliable measure of the efficacy of public expenditures and historical data is readily available on total and cohort-specific literacy.

British India offers a unique institutional context to evaluate the effects of public investments on outcomes. As discussed in section 2, district board expenditures were largely a function of cesses and grants with a small contribution from school fees that were collected in public schools operated by the district boards. The across-district differences in expenditures were due to both inter- and intra-provincial factors. Although a portion of inter-provincial differences in expenditures was due to historical circumstances i.e. the type of Land Settlement, Indian provinces differed along several other dimensions and exploiting differences across provinces may be especially vulnerable to criticisms of omitted variables. The analysis therefore focuses on variation across districts within the same province by using province dummies. But, a simple OLS estimation of expenditures and outcomes within provinces is still subject to endogeneity problems.

First, if literacy affects public expenditures, then we cannot infer whether higher expenditures lead to higher literacy or higher literacy leads to higher expenditures, a problem that is commonly referred to as reverse causality. We can however address this issue by using lagged values of expenditures. If we assume that a child generally attends primary school between the ages of 5 and 10 where he acquires basic literacy skills, then a primary school pupil in 1911 would on average fall in the cohort aged 15 to 20 in 1921. Consequently, we can examine the effect of 1911 expenditures on 1921 cohort-specific literacy rates in the population aged 15 to 20.

Second, it is important to control for variables that affect both the level of public spending and literacy rates. For example, districts with higher population density i.e. more urbanized may have a higher private demand for education and more public resources to supply schools. A failure to control for population density would then generate an upward bias on the OLS coefficient. To address these concerns, I include three sets of variables to capture differences in the costs and

benefits of literacy across districts. Since parental education and family background are critical inputs into the schooling calculation, I control for the share of the population supported by professionals such as doctors, lawyers, etc. This is perhaps a lower bound on the effect of family education since it limits the educated parental population to the set of professionals but it is the best available measure given data constraints. I also introduce variables to capture the share of the population supported by commerce and industry because economic structure can also affect schooling since the opportunity cost of a rural child's time who worked in the field to assist his agricultural parents was perhaps higher than for other children.[29]

Because of the high degree of social heterogeneity in India and the potential variation in demand for schooling across different groups, the analysis controls for the population share of Brahmans, the traditional educated caste of Hindus, Muslims, Christians, Buddhists, and minorities such as lower castes and tribes. I also include a measure of caste and religious fragmentation, which has a strong negative effect on the supply of private primary schools in the colonial period (Chaudhary 2008) and could potentially affect the demand for education. In addition, I control for population density, access to railways and income to capture differences in economic conditions.

Despite the controls and province fixed effects, the OLS coefficient could still be reflective of some dimension of unobservable heterogeneity across districts, which affects both literacy and public spending. The Indian context, however, offers an instrumental variable in the form of land revenues that were strongly correlated with educational expenditures across districts but were unlikely to be systematically related to the error term. As discussed in section 2, the computation of land revenues was plagued with idiosyncrasies, which created quasi-random variation in both land revenues and cesses—surcharges on land revenues that were the most important source of income for local district councils. Given land revenues determined cesses, we expect them to be strongly correlated with board spending.

[29] The population share supported by agriculture is the omitted category in all the regressions.

Table 2 examines the determinants of 1911 per-capita district board educational expenditures. Land revenues are a key determinant of public spending even after including province dummies that capture unobservable differences across provinces. Although the coefficient on land revenues is higher in specifications without province fixed effects, almost two-thirds of the effect of land revenues on expenditures is driven by within province differences. Controlling for province fixed effects and observable differences in geography and socio-economic factors across districts, a 10 percent increase in per-capita land revenues leads to a 3 percent increase in average educational expenditures (specification 7) with an F-statistic of 12 on land revenues. Although a dummy for Temporary Settlement districts is significant in the across province comparisons (specifications 3 and 4), neither this variable nor the type of land tenure system (specifically, the proportion of non-landlord districts) from Banerjee and Iyer (2005) has any statistically significant effect on expenditures after controlling for province fixed effects (specification 5). This is largely due to the fact that there was limited variation in Land Settlements or land tenure systems within provinces.

While land revenues are strongly correlated with expenditures, a valid instrument also has to be uncorrelated with unobservables that may affect literacy. Land revenues were generally supposed to capture a percentage of the rental assets or agricultural surplus of a district. In principle, soil quality and rainfall were two objective factors, which strongly influenced the assessment calculations. For example, districts with alluvial soil that enjoyed a larger agricultural surplus were assessed higher taxes. Discussions on assessment procedures in Baden-Powell (1972) indicate that settlement officers may have also considered the development and social structure of the district. So after controlling for differences in geography, development, occupational and social structure across districts, land revenues should be uncorrelated with the underlying demand for literacy because revenues are then largely a function of idiosyncratic factors unrelated to education. Given the method of assessing revenues varied across provinces and there may be some concern that unobservable and perhaps endogenous provincial characteristics influenced the assessment method, I focus on within

province variation. Land revenues thus offer a credible instrument and the identifying assumption is that they are uncorrelated with unobservables driving literacy controlling for the above mentioned differences across districts.

The Indian context also offers an alternative strategy to address the endogeneity problem of public spending. Since land revenues were revised infrequently, temporal variation in district board educational expenditures was largely due to increases in provincial grants to district boards. We can thus exploit the temporal variation in expenditures to identify the effects of expenditures on outcomes. The historical sources do not report board expenditures for all the districts in 1921, but, I can examine the effect of changes in expenditures between the 1901 and 1911 cross-sections for which data is available on changes in cohort literacy between 1911 and 1921 controlling for changes in the other time varying control variables over this decade. This is an alternative to the traditional first difference estimation and effectively controls for time invariant district characteristics. Although grants were not randomly assigned across districts, official documents indicate a preference for rewarding more needy (i.e. poorer) districts and those with larger minority populations. This would generate a downward bias and so the estimates from the first difference specifications should perhaps be viewed as a lower bound on the causal effect of expenditures on outcomes.

4. Data

For the empirical analysis, I constructed a new historical dataset that links district-level data from the colonial censuses to information reported in the Indian district gazetteers for 1911 and 1921. The dataset includes districts in all the major provinces of British India namely Assam, Bengal, Bihar and Orissa, Bombay, Central Provinces, Madras, Punjab and United Provinces.[30] Since comparable data on the urban centers of Bombay, Calcutta and Madras are unavailable in the gazetteers, these cities are excluded from the analysis. Indian districts were fairly large administrative units with an

[30] District boards were not constituted in some districts of Assam, Bengal, Bihar and Orissa. Consequently, these districts are excluded from the analysis.

average population of over one million in 1911. Given the district size and potential heterogeneity within districts, the district level aggregation will lead to some efficiency loss in the analysis relative to an analysis at a lower level of aggregation (Brown and Guinnane 2007).[31] Nonetheless, there does appear to be significant variation across Indian districts to identify the effects of spending on literacy.[32]

I restrict the analysis to the 1911 and 1921 cross-sections primarily due to data constraints. Literacy was inconsistently enumerated in the pre-1911 censuses and a uniform definition of literacy across all provinces was adopted in the 1911 census—an individual was recorded as literate if he or she could read and write a short letter to a friend. Census enumerators were required to test for literacy and officials believed the enumeration was relatively accurate in subsequent censuses relative to the pre-1911 censuses when provinces tested for literacy in different and often inconsistent ways.[33] Other than the short reading and writing test, no schooling was necessary in order to be recorded as literate. In principle a child could become literate at home but this was perhaps true for only a very small share of the Indian population because literacy was very low in this period and a vast majority of literates appear to have attended some school.[34] The regressions thus focus on the following outcomes in 1911 and 1921 namely total literacy rates, cohort-specific literacy rates for the population aged 10 to 20 and 15 to 20 and literacy rates disaggregated by gender.[35]

Although the literacy variable became relatively accurate beginning in 1911, the same cannot be said for the data on age enumeration. Official discussions indicate that the age enumeration was

[31] In general the aggregation should make it more difficult to find statistically significant results. And, it would also lead to higher R-squared's relative to an individual or village level analysis.
[32] District board education expenditures in 1911 range from rupees 0.015 to 0.34 per capita across districts, while cohort literacy rates among ages 10 to 20 range from 1.9 percent to 23 percent in 1911.
[33] In the pre-1911 censuses no specific guidelines were given to enumerators to test for literacy, which lead to substantial variation in the methods adopted across provinces. Although officials point to certain problems with the post-1911 enumeration such as enumerators on occasion adopting school standards, they do indicate that "the simple criterion laid down was easily understood and sensibly interpreted" (Census of India 1921, Volume I – Report, Chapter VIII).
[34] See Census of India (1911 and 1921).
[35] The 10 to 20 and 15 to 20 cohorts are up to but not including ages 10 and 20.

plagued with inaccuracies because individuals often did not know their age.[36] This is perhaps to be expected for a country like India in the early 20th century because numeracy and literacy generally go together. This inaccuracy however introduces measurement error in cohort-specific literacy rates. Fortunately, the census discussions highlight that uncertainty was the primary culprit for incorrect enumerations and so the measurement error arising from people being uncertain about their age is likely to be random or classical in nature. Classical measurement error in the dependent variable yields consistent estimates but may lead to larger standard errors (Angrist and Krueger, 1999).

In addition to literacy, I constructed measures of development such as population density, and occupational structure from the 1911 and 1921 censuses. Given the accuracy concerns of the smaller occupational categories, I focused on broad occupational types—the share of the population supported by agriculture, commerce, industry and professionals—to minimize measurement error. To control for differences in social structure across districts that may impact both literacy and the level of public spending, I also extracted information on important castes and religions namely upper caste Brahmans that traditionally worked as priests and teachers, Muslims, Christians, Buddhists, Animistic tribes and lower castes who were also referred to as the untouchables. To identify lower castes, I used the province-specific social precedence tables reported in the census of 1901, which list the specific lower castes and used these lists to calculate the number of lower castes in 1911 and 1921.[37] The castes enumerated as lower castes are the same as the castes enumerated as Scheduled

[36] Although ignorance was the most importance reason for inaccuracies in enumeration, official discussions also allude to parents being superstitious about revealing the true age of an infant, young married couples overstating their age and older men (especially widowers and bachelors) understating their age. See Census of India 1931, Volume I, Part I – Report, Chapter IV for specific details.

[37] The 1901 census compiled social precedence tables on caste rankings in consultation with locals well versed with the caste hierarchy. Due to a variety of reasons outlined in Dirk (2001), subsequent censuses discontinued this practice partially because several castes, especially those in the middle, that believed they were higher in the hierarchy than their stated position petitioned the government to change their ranking. In general, the colonial caste censuses have generated substantial critiques of British interpretations of caste and the subsequent impact of the censuses on the Hindu caste system. See Cohn (1990), Dirks (2001), and Srinivas (1996) for details. The caste and religious data are self-reported measures, which could introduce measurement error if individuals in non-upper castes changed their name to enumerate themselves as upper castes. This measurement error would attenuate the estimates on the caste variables toward zero.

Castes in post-independence India and for the province of Punjab I used the Scheduled Caste lists from 1950 to indentify lower castes since the colonial censuses did not provide this information.[38] A separate data appendix is available upon request that lists the specific castes enumerated as lower caste in each province.

Using the information on caste and religion, I constructed a measure of caste and religious fragmentation (CRFI), similar to the ethnic-fragmentation index used in the literature. CRFI is a Herfindahl-based index equal to $1 - \sum s_i^2$, where s_i is the population share of each caste or religious group in 1911 or 1921 respectively. As per Banerjee and Somanathan (2007), I restricted the data to Hindu castes larger than 1 percent of the province population along with Muslims, Christians, Buddhists, Sikhs, Jains, Animistic Tribes and Others that include castes that did not constitute 1 percent of the province population and small religious groups such as the Jews and Parsies.

Given the importance of geography for the calculation of land revenues, I used the World Bank Agricultural and Climate dataset to construct dummy variables for black, red and alluvial soil.[39] Moreover, I extracted district-level data on normal rainfall from the 1911 census and created a dummy for coastal districts based on a visual inspection of historical maps from this period. Since railways may have also factored into revenue calculations, I constructed a variable that captures the number of years between 1911 and the first year a railway line was constructed through a district.[40]

I used the Indian district gazetteers of this period to obtain information on expenditures and revenues. The district gazetteers are a unique source of data with two parts, A and B, for each district series: part A describes the history, geography, economy, culture and administration of the district, while part B contains statistical tables that complement the discussion in part A. Although the statistical tables do not consistently report total public expenditures for all districts, the expenditure

[38] To ensure that the Scheduled Caste lists were in fact picking up lower castes in Punjab, I double checked these castes against their occupations and most of them were associated with menial occupations common to lower castes.
[39] Data on soil type is from the post-independence period.
[40] Dave Donaldson provided the necessary data on Indian railways to construct this variable.

statements of rural district boards report educational expenditures. Given rural boards managed the provision of primary schooling, district board expenditures were an important category of public spending as discussed in section 2. While this data are consistently available at the district level for the 1911 cross-section, several districts stop reporting this information after 1916 and therefore the regressions only focus on the 1911 and 1921 cross-sections.[41]

I also extracted information on land tax and income tax revenues collected in each district. For districts in the United Provinces, the gazetteers do not report the exact land revenues collected but rather the revenue demand or what the districts were expected to collect in specific years between 1905 and 1911 with the reported year varying by district. Since the revenue demand was fixed for 20 to 30 years, the lack of consistency in the reporting year does not pose any significant problems. In principle, there could be differences between the revenue demand and revenue collected but the analysis controls for both geographic and economic factors in addition to province fixed effects, which should address any problems that may arise from this reporting difference.

In the absence of information on median income, I use per-capita income tax revenues as a proxy for income. The taxes were levied on a small share of the population such as government employees, and members of the formal non-agricultural sector with substantially high incomes.[42] Income taxes are the only information on income reported in the gazetteers and should be interpreted as more representative of the higher tail of the income distribution. This variable is consistently available for the 1911 cross-section but is unavailable for the districts of Assam, Bengal, Bihar and Orissa in 1921. Consequently, the regressions on the 1921 cross-section use the 1911 per-capita figures for these districts.[43]

Table 3 presents annual summary statistics of the variables for 1911 and 1921. Literacy rates were fairly stagnant between 1911 and 1921 increasing from just over 5 percent to 6 percent. Male

[41] For Madras districts, data on expenditures is reported for 1902-03 and 1912-13.
[42] See Kumar, ed. (1982) for details on income taxes. Income tax rates were very low in the colonial period.
[43] The analysis is robust if we use the 1911 per-capita income taxes for all districts in the 1921 cross-section.

literacy rates were higher and averaged 10 percent as compared to female literacy that averaged only 1 percent across Indian districts. High rates of dropout, wastage and relapses into illiteracy were frequently blamed for the lack of progress in the early 20th century. Many of the occupational and development controls such as population density and commercial population also exhibit no remarkable changes between 1911 and 1921 with urbanization rates barely increasing from 9.6 percent to 10.7 percent over the decade. In terms of social structure, upper caste Brahmans represented 5 percent of the population on average, lower castes comprised almost 16 percent of the population and there was a high degree of fragmentation across Indian districts with a 75 percent probability that two random individuals drawn from a district would belong to different castes or religions.

5. Results

In this section, I first present the OLS results, followed by IV estimates of expenditures on literacy, robustness checks and finally the quasi-first difference results.

5.1 Ordinary Least Squares Results

Table 4 presents OLS estimates of district board expenditures on overall literacy and cohort-specific literacy rates among the population aged 10 to 20 and 15 to 20. The regressions include province fixed effects as well as controls for soil, rainfall, coastal districts, income, social structure, population density and railway development. Panel A focuses on the 1911 cross-section and reports the effects of 1901 expenditures on 1911 outcomes, while panel B focuses on the 1921 cross-section and uses 1911 expenditures as the key measure of public investments. The control variables in each cross-section are for the same year as the outcome variables.

Although grants and surcharges on land revenues accounted for the bulk of district board income, the boards also received school fees charged in public schools as additional income. Moreover, districts with higher private returns to literacy may have lobbied the boards to increase public spending on education. A failure to adequately control for differences in the returns to literacy

can thus generate an upward bias on the OLS coefficient because districts with higher private demand would have more children attending public schools, paying fees and exhibiting higher literacy rates. The OLS regressions therefore include the extensive set of controls precisely to mitigate these concerns of omitted variables.

The results are similar across the 1911 and 1921 cross-section with a 10 percent increase in per-capita expenditures in 1901 and 1911 translating into literacy gains of almost 4 percentage points for the cohort-specific literacy rate among the 10 to 20 age group. These effects of public spending, almost on the order of one standard deviation of literacy rates, are large and suggest that endogeneity problems may be generating an upward bias on the OLS estimates. To test for endogeneity directly, table 5 explores the effects of 1911 expenditures on cohorts in both 1911 and 1921 that in principle should be unaffected by public spending on primary education in 1911 because they were not exposed to any primary school in 1911. According to official sources, a child became proficient in reading and writing (i.e. literate) after completing anywhere from two to four years of primary schooling. If we assume that a child begins primary school at age 5, in principle 1911 educational expenditures should be uncorrelated with 1911 cohort literacy in the population aged 10 to 20 and especially those over 20 unless expenditures are endogenous to literacy. Moreover, 1911 expenditures should be unrelated with 1921 cohort specific literacy rates for the population under age 10 and over age 20 that were too young and too old respectively to be in primary school in 1911.

Table 5 reports strong correlations between spending in 1911 and literacy rates in 1911 (panels A and B), which are troubling because they suggest that the OLS estimates are indeed endogenous to demand side factors. In principle, 1911 expenditures could be picking up the effect of spending in previous years since expenditures did not change dramatically from year to year, but this cannot account for the large and statistically significant effects of 1911 expenditures on cohort-specific literacy rates for the population aged 20 and over (panel B) that attended primary school at

least ten years earlier in 1901 when expenditures were quite different.[44] Moreover, 1911 expenditures are also correlated with 1921 literacy in cohorts that in principle were 'not treated' to public spending on primary education in 1911 (panel C). In general, the endogeneity problems appear more severe for male literacy than female literacy, which is perhaps unsurprising in the historical Indian context where women were less likely to participate in the labor market due to social and cultural barriers thereby making them probably less responsive to private returns to education and literacy.

5.2 Instrumental Variables Results

To address the endogeneity problems of the OLS estimates, the next set of results focus on instrumental variables to identify the causal effect of expenditures on literacy. A valid instrument has to be correlated with educational expenditures but uncorrelated with the unobservables (i.e. error term) affecting literacy. In the Indian context, land revenues per-capita were an important determinant of educational expenditures both across and within provinces as seen in table 2. Moreover, after controlling for differences in agricultural (i.e. soil, rainfall, coastal districts) and economic conditions, the variation in land revenues was largely due to idiosyncratic differences in computation that should be unrelated to unobservable factors that may influence literacy (i.e. the error term). Land revenues thus offer an instrument for district board spending.

Table 6 presents the IV results of 1911 expenditures for both the 1911 and 1921 cross-sections controlling for province fixed effects and other observable differences in geography, income, development and social structure across districts.[45] In these specifications, 1911 expenditures per-capita are instrumented using 1911 per-capita land revenues. As seen in panel A, the IV estimates of 1911 per-capita spending are no longer statistically correlated with any measure of 1911 literacy. The lack of findings on 1911 outcomes are a bit reassuring because we would

[44] Beginning in 1905, provincial governments allocated larger grants to district boards to improve the quality and quantity of primary education.

[45] The IV estimates are similar although a bit larger in magnitude in specifications without province fixed effects.

expect 1911 expenditures to be unrelated to 1911 cohort-specific literacy rates if the instrument is successful in addressing the endogeneity problems of the OLS estimates seen in tables 4 and 5.

The findings on 1921 literacy are reported in panel B and indicate that 1911 expenditures lead to higher literacy among the 10 to 20 and 15 to 20 aged cohorts that were in primary school 10 years earlier and would presumably have been the most affected 'treatment group' in 1911. The findings on total literacy rates in 1921 thus appear to be largely driven by the positive effects of spending on these 'treated' cohorts. In almost all the specifications, the IV estimates in table 6 are smaller in magnitude as compared to the OLS estimates reported in table 4 and confirm that the OLS coefficients are indeed biased upwards. Nonetheless, the magnitude of the IV estimates is important—a 10 percent increase in 1911 per-capita district board spending, which roughly translates into 44 additional primary schools using average 1911 district population (1,112,131) or an 8 percent increase in school enrollment, leads to a 2.6 percentage point increase in 1921 cohort-specific literacy rates for the population aged 15 to 20 with substantially larger effects for male literacy although the latter results are marginally significant at the 12 percent level.[46] Although expenditures can contribute to higher literacy by increasing access to schools, they can also improve outcomes by moving existing pupils to better quality schools or keeping them in school longer. With this view a 10 percent increase in 1911 per-capita spending would allow 4 percent of pupils to move from a low quality private school with an untrained teacher to a higher quality public school with a trained teacher.[47]

While the OLS results found a statistically significant effect of public spending on 1921 female literacy, the IV results indicate that higher spending did not lead to better outcomes for women. Thus, there was some truth to official discussions on the importance of changing social

[46] Data on average cost of primary schools and enrollment rates used in the calculations are for 1911-12 obtained from the Progress of Education, Eighth Quinquennial Review (1923).
[47] Average annual expenditures for the two school types (public board and private unaided) are for 1911-12 and were obtained from the Progress of Education, Eighth Quinquennial Review (1923).

attitudes toward the education of women in the absence of which higher spending was unlikely to increase female literacy.[48] In addition to public investments, tables 4 and 6 also report the coefficient on the share of the population supported by professionals such as doctors, lawyers, teachers, etc. This is a key demand side factor, which is highly correlated with both male and female literacy as well as cohort-specific literacy. Other variables that also positively affect literacy include income tax revenues, being a coastal district and a lower degree of social fragmentation. Differences in returns to literacy as captured by these variables were thus also critical to outcomes in the colonial period.

5.3 Robustness Checks

Colonial public investments on primary education lead to substantial gains in literacy based on the IV estimates reported in table 6. Although land revenues strongly influenced public spending on education, there may still be some concern that these revenues are related in a systematic manner to the unobservable factors affecting literacy. After controlling for objective differences in geography, income and development across districts, the underlying assumption of the IV strategy is that the variation in revenues was largely a function of historical idiosyncrasies that are uncorrelated with the error term. Since this strategy does not identify the precise nature of the underlying variation in land revenues, this section subjects the IV results to a variety of robustness checks to alleviate concerns regarding the plausibility of land revenues as an instrument.

Similar to table 5 that tested for endogeneity in the OLS estimates, the first strategy uses certain cohort-specific literacy rates in 1921 as a control group because they were either too old or too young to be in primary school ten years earlier in 1911. Consequently, the IV coefficients on 1911 public spending should not affect outcomes for these cohorts because they were not exposed to 1911 expenditures. The top panel of table 7 (panel A) presents the IV coefficients on 1911 public spending against 1921 cohort-specific literacy rates for the cohort under age 10 and cohort over age 20 as the dependent variable. Individuals in the 1921 cohort under age 10 were not born in 1911 and

[48] Progress of Education, Quinquennial Reviews (1886-1937).

so could not have attended any primary school, while cohorts aged 20 and over were perhaps a bit too old for primary school in 1911. As seen in panel A, the estimates on district board expenditures are statistically insignificant for these control groups as we would expect.

The second strategy exploits the same idea and uses English literacy as another control outcome because public primary school expenditures should not increase literacy in the English language unless the instrument is not valid. Primary schools in this period only offered instruction in the vernacular medium as compared to secondary schools where English medium instruction was more common.[49] If the unobservable factors affecting the returns to English language literacy are similar to those for literacy in any other language, then we can use English literacy as a control outcome. The IV estimates on board expenditures should be uncorrelated with English literacy unless per-capita land revenues are related to some unobservable factors that lead to an increase in general demand for education. The first three specifications in panel B of table 7 report the findings on total English language literacy and disaggregated for the population aged 10 to 20 and 15 to 20 respectively. As is evident, the IV estimates on 1911 expenditures are in fact negatively correlated with English language literacy. Moreover, we can reasonably infer that the demand for English and non-English literacy was affected by similar forces because the share of professionals are positively related to both overall and English literacy. While this finding is reassuring, the test is not conclusive because the unobservables driving English literacy may be different from those for overall literacy.

Specifications 4, 5 and 6 in panel B of table 7 test for additional endogeneity problems with the IV strategy. Although public spending on education leads to better outcomes, there may be some concern about how this spending relates to expenditures on other public goods that may also positively influence literacy. Rural district boards were primarily responsible for the provision of education, medical services and local infrastructure. Districts with higher land revenues presumably would have more money to spend on these other local services or alternately districts could substitute

[49] See Wood's Education Despatch of 1854 and Progress of Education in India, Quinquennial Reviews (1886-1937).

between the different services. If areas with high public educational expenditures also spend more money on other services due to higher land revenues, then this could generate an upward bias on the IV estimates because higher quality medical services may have an independent effect on literacy.[50] However, the IV estimates on educational spending are robust to the inclusion of 1911 medical and infrastructure expenditures as additional control variables. I only report the results on cohort-specific literacy for ages 15 to 20 but the findings are unchanged for other measures of literacy as well.

Specification 6 in panel B tests for another variable that may be correlated with land revenues, expenditures and literacy namely the land tenure system. In Bengal, Bihar and Orissa, the British made landlords responsible for revenue payments (*zamindari*), whereas in other parts of the country they dealt directly with individual cultivators (*raiyatwari*) or with village bodies (*mahalwari*). Since non-landlord areas also enjoyed higher land revenues, it could be the case that these areas generated both higher revenues and a higher demand for basic literacy because cultivators in these districts negotiated directly with British offices. Banerjee and Iyer (2005) in an influential paper find significant differences in agricultural productivity and public investments between landlord and non-landlord districts in post-independence India. Given there is limited variation in land tenures within provinces, this should not raise any problems but nonetheless specification 6 includes the Banerjee and Iyer (2005) variable for non-landlord districts as an additional control. The results on expenditures are essentially unchanged although they are a bit larger in magnitude.

While the robustness tests thus far have focused on the IV estimates, the Indian institutional context offers another estimation strategy. Although the historical sources do not always report expenditures in 1921, I was able to obtain 1901 district board educational expenditures for all the districts used in the analysis. Since there is temporal variation in expenditures between 1901 and 1911 due to increases in grants from provincial governments to district boards, I can exploit this

[50] In theory, we could expect districts to allocate differential spending across local services i.e. high for education and low for medical services depending on the preferences of the district. However, in the Indian context, it appears that districts with high public spending on education spent more on all other public services on average.

increase in spending to identify its effects on changes in literacy for the relevant cohorts. Panel C in table 7 focuses on the production of literates namely changes in the overall and cohort-specific literacy rate for the population aged 15 to 20 between 1921 and 1911. Specifications 1 to 3 report the findings on changes in overall literacy and the coefficients indicate that changes in expenditures between 1901 and 1911 are positively correlated with changes in literacy between 1921 and 1911 controlling for changes in all the other control variables other than time invariant variables such as soil type that are not included in these regressions.

The results on the production of literates in the cohort aged 15 to 20 are reported in specifications 4 to 6, which highlight that the findings on overall literacy are driven by increases in the cohort-specific literacy rates. If I had access to consistently enumerated literacy data for 1901 or perhaps expenditure data for 1921, I could run a first difference regression. In the absence of that data, these regressions are a second best option and in a sense are quasi-first difference estimations that are essentially controlling for time invariant district characteristics. The main idea is that changes in spending between 1901 and 1911 should affect outcomes for the 15 to 20 aged cohort between 1911 and 1921 that was in primary school between 1901 and 1911 and thus was most exposed to public spending on primary education. The magnitude of the coefficient suggests that a 10 percent increase in spending between 1901 and 1911 would lead to a 1.5 percentage point increase in the cohort-specific literacy rate for the population aged 15 to 20. The magnitude of this effect is smaller than the IV estimates but we may want to view these first difference estimates as a lower bound on the effect of public spending. Although public grants were often disbursed on a per-capita basis to districts within provinces, there was some discussion of giving more money to poorer or more backward districts that would generate a downward bias on the quasi-first difference estimates. Nonetheless, these results along with the IV estimates broadly indicate that public spending lead to economically significant increases in male literacy but had no appreciable impact on female literacy.

5.4 Discussion

Now that the empirical analysis has established that higher public spending in British India would have translated into better outcomes at least for men, the next logical question is how we should interpret the magnitude of these effects. Relative to the contemporary US literature on school finance reforms and test scores where higher expenditures have either insignificant or relatively small positive effects on test scores, the estimates on colonial public expenditures may seem some what large.[51] But, the key difference is that our outcome of interest is basic literacy that is probably more responsive to higher public spending as opposed to math test scores in the United States.

An alternate way to interpret the magnitude of the results is to use the estimates on public spending to arrive at a rough calculation of the costs associated with increasing mass literacy in British India. Even though higher public expenditures on primary education increased literacy, the colonial government may have been reluctant to increase aggregate public spending perhaps because it was not cost-effective. A 10 percent increase in 1911 per-capita district board expenditures translates into an increase of Rs. 7107 using the average 1911 district population size of 1,112,131 from the data. Since a 10 percent increase in expenditures leads to a 2.6 percentage point increase in 1921 literacy among the population aged 15 to 20, this increase in spending is equivalent to making 2399 additional people, or rather males, literate . This back of the envelope calculation suggests that it would have cost the colonial government roughly 3 rupees to make an additional person literate.[52]

Unfortunately, it is difficult to assess returns to basic literacy in the colonial period because the historical sources do not report information on wages disaggregated by education or literacy. One way to assess the cost of 3 rupees is to consider whether a reallocation of the entire education budget toward primary education would have generated substantial increases in literacy. During the colonial period more than 50 percent of total expenditures were directed to secondary schools and

[51] Hanushek's (1986, 1996) evaluation of the US education literature suggests that there is a wide range of estimates but in general even on the high end of positive effects they are of small magnitude.
[52] 3 rupees corresponds to just over 3 years of annual primary school fees at an average primary school in 1911-12.

colleges and a redirection of these expenditures to primary education would have increased overall literacy by 2.4 percentage points in 1921 from say 7.5 percent to 9.9 percent.[53] Although this does reflect an increase in literacy, it also emphasizes that a simple reallocation of funds within the education budget was not the answer to achieving mass literacy. The colonial government would have had to substantially increase public spending to achieve significant gains in male literacy.

Before concluding, I want to suggest a potential mechanism linking public spending to literacy based largely on qualitative evidence. Chaudhary (2008) finds that districts with higher land revenues per-capita, a proxy for higher public spending on primary education, had more publicly funded and managed primary board schools. Discussions in the Quinquennial Reviews of Education indicate that these schools were often of higher quality with a larger proportion of trained teachers and charged lower fees as compared to privately managed schools. Due to more trained teachers and better facilities these schools had higher operating expenditures, but they did not pass on these higher expenses to the students. Higher expenditures thus translated into better outcomes for males via a larger number of public schools with better teachers that offered a larger public subsidy to the rural population in the form of lower fees.

6. Conclusion

In this paper, I explore the effects of colonial public investments on literacy rates across districts of British India in the early 20th century. Using a novel dataset on expenditures and outcomes, I find large and positive effects of spending on cohort-specific literacy rates using an instrumental variables strategy. A 10 percent increase in spending or 44 additional primary schools lead to a 2.6 percentage point increase in literacy rates for the population aged 15 to 20. The findings are robust to a variety of falsification tests and an alternate estimation strategy akin to a first difference estimation that exploits variation in spending within districts. Qualitative evidence

[53] Data on 1922 expenditures on secondary schools is from the Progress of Education in India, Eighth Quinquennial Review (1917-22).

suggests that greater reliance on higher quality public schools with lower fees is perhaps a key mechanism from higher spending to literacy.

These findings have three important implications for the development of education in British India. First, more public money would have lead to better outcomes for male literacy. Second, the colonial educational budget in the early 20th century was insufficient to tackle the problem of mass illiteracy even if the entire budget was targeted to primary education. To increase literacy the British Government would have had to either raise taxes or reallocate revenues from other sources toward education spending. Third, higher public spending would not have had any appreciable effect on female literacy rates that were more responsive to demand side variables. Social barriers against the education of women were perhaps so strong in British India that building more schools was not the solution to the problem of female illiteracy. Finally, the results from this paper also have broader implications beyond Indian economic history.

A cursory examination of regional literacy patterns in contemporary India suggests that they are very similar to those observed in the colonial period. Cochin, which corresponds to the contemporary state of Kerala, enjoyed relatively high levels of literacy as early as 1911 (15 percent as compared to the national average of 5.6 percent) and is close to achieving universal literacy today. Western and southern India generally had above average literacy rates historically as compared to Northern and Eastern India and this phenomenon has continued into the post-independence period. Indian districts appear to have inherited a certain educational endowment at Independence and differences in colonial investments were partially responsible for these differences in endowments along with key demand side factors such as the share of professionals. Understanding the historical circumstances and endowments is thus a necessary pre-condition for current development policies that hope to remedy the poor educational performance in places such as Bihar and Orissa that unfortunately have strong and persistent historical roots.

References

Angrist, Joshua D., and Alan B. Krueger. "Empirical Strategies in Labor Economics", Chapter 23 in *Handbook of Labor Economics*. Orley Ashenfelter and David E. Card. New York: Elsevier, 1999.

Baden-Powell, Baden H. *A Short Account of the Land Revenue and its Administration in British India; with a Sketch of the Land Tenures*. Oxford: Clarendon Press, 1907.

Baden-Powell, Baden H. *The Land-Systems of British India; Being a Manual of the Land-Tenures and of the Systems of Land-Revenue Administration Prevalent in the Several Provinces.* New York: Johnson Reprint Corp., 1972.

Banerjee, Abhijit and Lakshmi Iyer. "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India", *American Economic Review* 95, no. 4 (2005), 1190-1213.

Banerjee, Abhijit, and Rohini Somanathan. "The Political Economy of Public Goods: Some Evidence from India", *Journal of Development Economics* 82, no. 2 (2007):287-314.

Basu, Aparna. *The Growth of Education and Political Development in India 1898-1920*. New Delhi: Oxford University Press, 1974.

_____. *Essays in the History of Indian Education*. New Delhi: Concept Publishing, 1982.

Becker, Gary. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: National Bureau of Economic Research, Second Edition, 1964.

Blyn, George. *Agricultural Trends in India, 1891-1947: Output, Availability and Productivity*. Philadelphia: University of Pennsylvania Press, 1966

Broadberry, Stephen and Bishnupriya Gupta. "The Historical Roots of India's Service-led Development: A Sectoral Analysis of Anglo-Indian Productivity Differences, 1870-2000", working paper, June 2008.

Brown, John C. and Timothy W. Guinnane. "Regions and time in the European Fertility Transition: Problems in the Princeton Project's Statistical Methodology", *Economic History Review* 60, no. 3(2007):574-595.

Chand, Gyan. *Financial System of India*. London: Kegan Paul, Trench, Trubner & Co. Ltd., 1926.

_____. *The Essentials of Federal Finance.* London: Oxford University Press, 1930.

_____. *Some Aspects of Fiscal Reconstruction in India.* London: Oxford University Press, 1931.

_____. *Local Finance in India.* Allahabad: Kitabistan, 1947.

Chaudhary, Latika. "Determinants of Primary Schooling in British India", *Journal of Economic History*, forthcoming.

Davis, Lance E. and Robert A. Huttenback. *Mammon and the Pursuit of Empire: The Political Economy of British Imperialism, 1860-1912*. New York: Cambridge University Press, 1986.

Dirks, Nicholas. *Castes of Mind: Colonialism and the Making of Modern India*. Princeton University Press, 2001.

Downes, Thomas A. "Evaluating the impact of school finance reform on the provision of public education: The California case", *National Tax Journal* 45, no. 4 (1992), 405-419.

Downes, Thomas A., Richard Dye, Therese McGuire. "Do limits matter? Evidence on the effects of tax limitations on student performance", *Journal of Urban Economics* 43, no. 3 (1998), 401-417.

Drèze, Jean and Amartya Sen. *India: Economic Development and Social Opportunity.* Clarendon, Oxford, 1998.

Drèze, Jean and Geeta Kingdon. "School Participation in Rural India", *Review of Development Economics* 5, no. 1 (2001), 1-33.

Duflo, Esther. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review* 91, no. 4 (2001), 795-813.

Ghosh, Suresh. C. *The History of Education in Modern India 1757-1998*. New Delhi: Orient Longman, 2000.

Glewwe, Paul. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes", *Journal of Economic Literature* 40, no. 2 (2002), 436-82.

Government of India. *Indian Education Commission, Report and Appendices.* Calcutta, 1883-84.

_____. *Progress of Education in India: Quinquennial Review.* Volumes 1886-1937.

_____. *Census of India, 1911.* Calcutta: Office of the Superintendent of Government Printing, India, 1913.

_____. *Census of India, 1921.* Calcutta: Office of the Superintendent of Government Printing, India, 1923.

_____. *India, district and provincial gazetteers*. Zug, Switzerland: Inter Documentation, 1979.

Guha, Sumit, ed. *Growth, Stagnation or Decline? Agricultural Productivity in British India.* New Delhi: Oxford University Press, 1992.

Hanushek, Eric. "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature* 24, no. *3* (1986), 1141-77.

Hoxby, Caroline. "All school finance equalizations are not created equal", *Quarterly Journal of Economics* 116, no. 4 (2001), 1189-1231.

Kingdon, Geeta. "Student Achievement and Teacher Pay: A Case Study of India", STICERD working paper 74, London School Econ. Polit. Science, 1996.

Kingdon, Geeta G. "The Progress of School Education in India", *Oxford Review of Economic Policy* 23, no. 2 (2007), 168-195.

Kumar, Dharma, ed. *The Cambridge Economic History of India vol. 2*. Cambridge: Cambridge University Press, 1982.

Mincer, Jacob. *Schooling, Experience and Earnings*. New York: Columbia University Press for National Bureau of Economic Research, 1974.

Misra, Babu Ram. *Indian Federal Finance.* Calcutta: Orient Longmans, 1960.

Misra, Surendra N. and Manaranjan, Behera. "Child Nutrition and Primary Education: A Comparative Study of Mid-day Meal Programme in Orissa and Tamil Nadu", *Indian Journal of Social Development* 3, no. 2 (2003), 267-99.

Mukhopadhyay, Gopal. *Mass Education in Bengal (1882-1914)*. Calcutta: National Publishers, 1984.

Murray, Sheila E., William Evans and Robert Schwab. "Education-Finance Reform and the Distribution of Education Resources", *American Economic Review* 88, no. 4 (1998), 789-812.

Nurullah, Syed and J.P. Naik. *A History of Education in India (During the British Period)*. London: Macmillan, 1951.

Pal, Sarmistha and Sugata Ghosh. "Elite Dominance and Under-Investment in Mass Education: Disparity in the Social Development of the Indian States, 1960-92", IZA Discussion Paper Series No. 2852 (2007).

Pratham. "ASER 2005—Annual Status of Education Report", New Delhi, Pratham, 2006.

Probe Team. *Public Report on Basic Education in India*, New Delhi, Oxford University Press, 1999.

Ramachandran, V.K., "On Kerala's Development Achievements", in Drèze and Sen (edited) *Indian Development: Selected Regional Perspectives.* Oxford: Clarendon Press, 1997.

Roy, Choudhury and Uma Datta, "Inter-state and Intra-state Variations in Economic Development and Standards of Living", *Journal of the Indian School of Political Economy* 5, no. 1 (1993).

Roy, Tirthankar. *The Economic History of India 1857-1947*. New Delhi: Oxford University Press, 2000.

Schultz, Theodore W. "Reflections on Investment in Man", *Journal of Political Economy*, 70, no. 5 (1962), S1- S8.

_____. *Transforming Traditional Agriculture*. Chicago: University of Chicago Press, 1983.

Srinivas, M.N. *Village, Caste, Gender and Method: Essays in Indian Social Anthropology*. Oxford University Press, 1996.

Tinker, Hugh. *The Foundations of Local Self-Government in India, Pakistan and Burma*. New York: Frederick A. Praeger, Publishers, 1968.

Whitehead, Clive. "The Historiography of British Imperial Education Policy, Part I: India", *History of Education* 34, no. 3 (2005), 315-29.

Figure 1 - Map of British India and Princely States

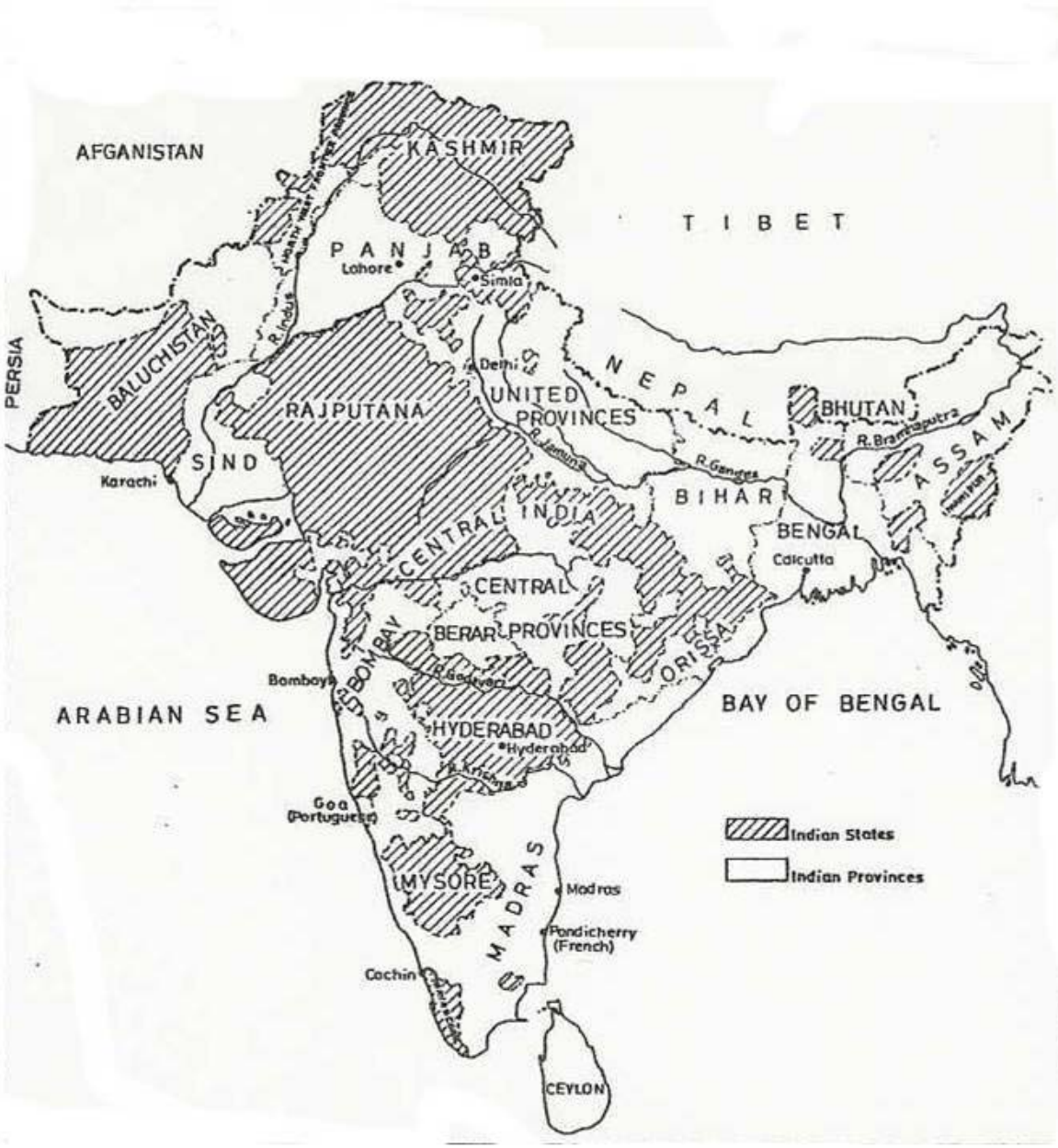

TABLE 1: DISTRICT BOARD EDUCATIONAL EXPENDITURES AND INCOME

	DB Educational Expenditures in 1912-13			
	Per 1000 of the population (£)	As % of Total Educ. Exp.	As % of Total Primary Educ. Exp	As % of Total Public Primary Educ. Exp.
British India	**3.59**	**15%**	**49%**	**65%**
Assam	5.60	30%	84%	87%
Bengal	2.43	8%	40%	80%
Bihar and Orissa	1.65	13%	34%	61%
Bombay	10.27	20%	42%	59%
Central Provinces and Berar.	4.58	25%	61%	66%
Madras	3.03	12%	34%	38%
Punjab	5.12	18%	85%	93%
United Provinces	3.55	21%	93%	100%

	DB Revenues in 1912-13 Per 1000 of the population(£)			
	Cesses (Additional taxes on land revenues)	Other Sources (for e.g. Provincial grants)	Total Income	% of Cesses and Other Sources to Total Income
British India	**6.53**	**5.96**	**16.65**	**75%**
Assam	7.40	7.02	23.11	62%
Bengal	4.28	4.35	10.51	82%
Bihar and Orissa	4.56	3.08	9.66	79%
Bombay	10.84	10.23	25.78	82%
Central Provinces and Berar	4.99	9.34	16.07	89%
Madras	10.80	10.13	31.85	66%
Punjab	10.05	6.75	22.92	73%
United Provinces	5.20	4.26	11.53	82%

Source: Statistical Abstracts of India and Progress of Education in India, Eighth Quinquennial Review (1917-1922). The statistical abstracts separate educational expenditures into direct expenditures on different levels of education and indirect expenditures on buildings, etc. I constructed total primary education expenditure as the sum of direct expenditures on primary schools and a proportion of indirect expenditures on buildings with the proportion equal to the direct primary school expenditures relative to total direct expenditures on education. Information on public primary education expenditures is from 1916-17 and was obtained from the Quinquennial Review, Volume II, Supplemental Tables, page 125. In the bottom panel, provincial contributions or grants to district boards are included under "other sources" along with school fees and miscellaneous income sources. The omitted category of income is tolls.

TABLE 2: DETERMINANTS OF 1911 DISTRICT BOARD EDUCATIONAL EXPENDITURES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Land Tax Revenues per-capita	0.0289***	0.0161***	0.0267***	0.0287***	0.0214***	0.0173***	0.0178***
	[0.0046]	[0.0055]	[0.0050]	[0.0074]	[0.0071]	[0.0055]	[0.0051]
Temporary Settlement			0.0169***	0.0138***	-0.0034		
			[0.0043]	[0.0042]	[0.0071]		
Proportion of Non-Landlord (Banerjee and Iyer 2005)				0.0146*	0.002		
				[0.0076]	[0.0087]		
Province Fixed Effects	No	Yes	No	No	Yes	Yes	Yes
Geographic Controls	No	No	No	No	No	Yes	Yes
Social and Economic Controls	No	No	No	No	No	No	Yes
Constant	0.0218***	0.0519***	0.0114***	0.003	0.0455***	0.0294	0.0667*
	[0.0060]	[0.0119]	[0.0043]	[0.0069]	[0.0159]	[0.0199]	[0.0392]
Observations	197	197	197	153	153	197	197
Adjusted R-squared	0.51	0.72	0.53	0.58	0.79	0.72	0.80
F-test: Land Tax Revenues = 0	38.97	8.66	27.92	15.14	9.13	9.82	12.33
Prob > F	0.0000	0.0037	0.0000	0.0002	0.003	0.0020	0.0006

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Geographic controls include dummies for coastal districts, black, red and alluvial soil, and normal rainfall. Social and economic controls include population share supported by industry, population share supported by commerce, populations share supported by professionals, population density, number of years since first railway line, income tax revenues per-capita, share of Brahmans, Christians, Buddhists, Muslims, Tribes, low castes and a measure of caste and religious fragmentation.

TABLE 3: SUMMARY STATISTICS

Variable	1911 Mean	1911 Std. Dev	1921 Mean	1921 Std. Dev
Census Variables				
Overall literacy rate	5.1%	2.8%	6.0%	3.2%
Overall male literacy rate	9.3%	4.8%	10.4%	5.2%
Overall female literacy rate	0.8%	1.2%	1.2%	1.5%
Literacy rate aged 10-20	6.3%	3.6%	7.4%	4.0%
Male literacy rate aged 10-20	10.5%	5.7%	11.9%	6.0%
Female literacy rate aged 10-20	1.4%	1.9%	2.2%	2.5%
Literacy rate aged 15-20	7.4%	3.7%	9.0%	4.4%
Male literacy rate aged 15-20	12.8%	6.2%	15.0%	7.0%
English literacy	0.5%	0.5%	0.7%	0.6%
English literacy aged 10-20	0.7%	0.7%	1.0%	0.9%
Fraction Urban	9.6%	8.4%	10.7%	9.8%
Population Density	376	250	380	258
Fraction Commerce	7.0%	3.2%	6.7%	3.1%
Fraction Industry	12.3%	6.2%	11.6%	6.1%
Fraction Professionals	1.6%	0.8%	1.6%	0.9%
Fraction Brahman	5.0%	4.4%	5.1%	4.3%
Fraction Muslim	23.9%	26.3%	24.1%	26.2%
Fraction Christian	0.9%	2.0%	1.1%	2.2%
Fraction Tribes	3.1%	8.2%	2.8%	7.5%
Fraction Low Castes	16.0%	8.1%	14.8%	8.2%
Caste & Religious Fragmentation Index	0.75	0.19	0.74	0.19
District Gazetteer Variables (in Rupees)				
Income Tax Revenues per-capita	0.06	0.10	0.20	0.65
1911 District Board Educ. Exp per-capita	0.06	0.04	.	
1911 Land Tax Revenues per-capita	1.46	1.03	.	
No. of Observations	197	197	197	197

Sources: Census of India (1911 and 1921) and Imperial District Gazetteer Series (1914).

The dataset includes districts in the British Indian provinces of Assam, Bengal, Bombay, Central Provinces and Berar, Madras, Punjab and United Provinces with functioning district boards. Local boards were not constituted in a few districts of Bengal, Bihar and Orissa, and the hill districts of Assam. See text for more details.

Income tax revenues are missing for Garhwal district in United Provinces in 1921.

TABLE 4: OLS ESTIMATES OF PUBLIC EXPENDITURES ON LITERACY RATES

	All ages			Cohort 10-20			Cohort 15-20		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Panel A: 1911 Cross Section									
District Board Educ Exp per-capita in 1901	0.2356**	0.4094**	0.0311	0.3990***	0.6364***	0.0828*	0.3709**	0.5943**	0.0684
	[0.0954]	[0.1801]	[0.0224]	[0.1453]	[0.2410]	[0.0443]	[0.1502]	[0.2564]	[0.0465]
Fraction Professionals	0.9908***	1.6753***	0.2901***	1.1022***	1.6951***	0.4453***	1.2865***	2.0573***	0.5365***
	[0.2625]	[0.4617]	[0.0925]	[0.3696]	[0.5886]	[0.1637]	[0.3653]	[0.6234]	[0.1703]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	197	197	197	197	197	197	197	197	197
Adjusted R-squared	0.80	0.77	0.87	0.75	0.73	0.82	0.75	0.74	0.82
Panel B: 1921 Cross Section									
District Board Educ Exp per-capita in 1911	0.2034***	0.3216***	0.0630**	0.3723***	0.5498***	0.1515***	0.3993***	0.6052***	0.1398***
	[0.0566]	[0.1031]	[0.0280]	[0.0855]	[0.1298]	[0.0522]	[0.0976]	[0.1521]	[0.0522]
Fraction Professionals	1.1692***	1.7879***	0.4848***	1.3191***	1.7954***	0.7086***	1.5775***	2.2691***	0.7320***
	[0.2905]	[0.5127]	[0.1304]	[0.3620]	[0.5915]	[0.1903]	[0.3963]	[0.6887]	[0.2044]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
Adjusted R-squared	0.78	0.75	0.84	0.76	0.73	0.83	0.75	0.73	0.81

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All specifications in panels A and B control for province fixed effects and include geographic controls namely normal rainfall, dummies for black , red and alluvial soil, economic and development controls namely population share supported by industry and commerce, population density, number of years since first railway line, income tax revenues per-capita and social controls namely share of Brahmans, Muslims, Christians, Buddhists, Tribes, low castes and a measure of caste and religious fragmentation.

TABLE 5: ARE OLS ESTIMATES ENDOGENOUS?

Panel A: 1911 Literacy Rate

	All ages			Cohort 10-20		
	Total	Male	Female	Total	Male	Female
District Board Educ Exp per-capita in 1911	0.1852*** [0.0628]	0.3115** [0.1230]	0.0329* [0.0186]	0.3363*** [0.0935]	0.5201*** [0.1586]	0.0843** [0.0366]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	197	197	197	197	197	197
Adjusted R-squared	0.80	0.77	0.87	0.76	0.74	0.82

Panel B: 1911 Literacy Rate

	Cohort under 10			Cohort over 20		
	Total	Male	Female	Total	Male	Female
District Board Educ Exp per-capita in 1911	0.0321** [0.0125]	0.0567*** [0.0194]	0.0059 [0.0104]	0.2092** [0.0814]	0.3497** [0.1629]	0.0307 [0.0197]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	197	197	197	197	197	197
Adjusted R-squared	0.81	0.74	0.89	0.79	0.77	0.85

Panel C: 1921 Literacy Rate

	Cohort under 10			Cohort over 20		
	Total	Male	Female	Total	Male	Female
District Board Educ Exp per-capita in 1911	0.0421** [0.0170]	0.0539** [0.0263]	0.0299*** [0.0111]	0.2308*** [0.0731]	0.3769*** [0.1406]	0.0469 [0.0293]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	196	196	196	196	196	196
Adjusted R-squared	0.74	0.67	0.86	0.77	0.75	0.83

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All specifications in panels A, B and C control for province fixed effects and the same set of controls as in Table 4.

TABLE 6: IV ESTIMATES OF PUBLIC EXPENDITURES ON LITERACY RATES

	All ages			Cohort 10-20			Cohort 15-20		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Panel A: 1911 Cross Section									
District Board Educ Exp per-capita in 1911	0.086	0.157	-0.008	0.179	0.286	0.017	0.164	0.261	0.014
	[0.085]	[0.157]	[0.026]	[0.127]	[0.209]	[0.052]	[0.126]	[0.214]	[0.056]
Fraction Professionals	1.075***	1.811***	0.329***	1.198***	1.845***	0.494***	1.377***	2.207***	0.577***
	[0.248]	[0.438]	[0.088]	[0.350]	[0.559]	[0.160]	[0.347]	[0.587]	[0.170]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	197	197	197	197	197	197	197	197	197
Adjusted R-squared	0.80	0.77	0.86	0.75	0.73	0.82	0.75	0.73	0.82
Panel B: 1921 Cross Section									
District Board Educ Exp per-capita in 1911	0.127*	0.207	0.024	0.256**	0.375**	0.091	0.261**	0.339	0.106
	[0.076]	[0.134]	[0.034]	[0.103]	[0.163]	[0.063]	[0.130]	[0.218]	[0.071]
Fraction Professionals	1.262***	1.926***	0.532***	1.460***	2.007***	0.782***	1.745***	2.591***	0.772***
	[0.269]	[0.474]	[0.122]	[0.342]	[0.557]	[0.178]	[0.374]	[0.655]	[0.194]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
Adjusted R-squared	0.78	0.75	0.84	0.76	0.72	0.82	0.75	0.72	0.81

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions in panels A and B control for province fixed effects, and include the same set of controls as in table 4 that are for the same year as the outcomes. District board educational expenditures per-capita in 1911 are instrumented using 1911 per-capita land revenues. Specification 7 in table 2 is the first stage IV regression.

TABLE 7: ROBUSTNESS CHECKS

PANEL A: 1921 Literacy Rate

	Cohort under 10			Cohort over 20		
	Total	Male	Female	Total	Male	Female
District Board Educ Exp per-capita in 1911	0.0283	0.0445	0.0107	0.1426	0.2468	0.0038
	[0.0212]	[0.0320]	[0.0138]	[0.0972]	[0.1761]	[0.0363]
Fraction Professionals	0.2240***	0.2787**	0.1689***	1.6449***	2.5738***	0.6275***
	[0.0774]	[0.1182]	[0.0470]	[0.3513]	[0.6454]	[0.1500]
Controls	Yes	Yes	Yes	Yes	Yes	Yes

PANEL B: 1921 Literacy Rate

	English Language Literacy Rate			Literacy Rate		
	Total	10-20	15-20		Cohort 15-20	
District Board Educ Exp per-capita in 1911	-0.0655**	-0.0667*	-0.0937*	0.2419*	0.2516*	0.4977***
	[0.0326]	[0.0379]	[0.0526]	[0.1336]	[0.1378]	[0.1067]
Fraction Professionals	0.4224***	0.5485***	0.7030***	1.5561***	1.7216***	1.2710***
	[0.1367]	[0.1401]	[0.1559]	[0.4429]	[0.4004]	[0.4048]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
1911 Medical Exp.				Yes		
1911 Local Infrastructure Exp.					Yes	
Proportion Non-Landlord (Banerjee and Iyer 2005)						Yes

PANEL C: Δ Literacy Rate (1921-1911)

	Total	Male	Female	15-20	Male 15-20	Female 15-20
Δ Educational Exp (1911-1901)	0.0649**	0.1159**	0.0142	0.1551**	0.2626***	0.0535
	[0.0325]	[0.0475]	[0.0255]	[0.0664]	[0.0954]	[0.0567]
Δ Fraction Professionals	0.1633	0.241	0.0857*	0.4235**	0.7877**	0.1033
	[0.0998]	[0.1672]	[0.0503]	[0.1918]	[0.3418]	[0.1090]
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All specifications include province fixed effects. Panels A and B include the same set of controls as in tables 4 and 5, and 1911 district board expenditures are instrumented using 1911 land revenues. In panel C the control variables are the same as in tables 4 and 5 except that they are expressed as first differences between 1921 and 1911. See text for more details.