Taxation and Educational Development:
Evidence from British India

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

Version: June 2008

Abstract

This paper uses a novel district-level dataset to study the effects of public educational expenditures on literacy outcomes across districts of British India in the early 20th century. Using an instrumental variables strategy, I find that 1911 colonial public 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 expenditures translates into a 2.4 percentage point increase in the literacy rate among the population aged 15 to 20 with larger effects for male literacy. The results are robust to a variety of robustness tests and an alternate estimation strategy, which exploits the variation in expenditures within districts.

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 across space and time. Interest in the efficacy of public expenditures on education also extends beyond the academic realm and occupies a central position in policy debates within the US and other developing countries. The central issue surrounding the academic and policy debates is whether or not educational expenditures translate into better outcomes controlling for differences in demand side conditions.

To evaluate the effects 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 information on the returns to education.[3] 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 a key empirical challenge because higher spending may partially be a response to greater private demand for schooling.

---

[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] 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.[4] 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 overall education spending was very low relative to world standards.[5] A better understanding of the impact of public spending on outcomes has strong 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 appears to have important historical roots. For example, in 1991 the western states of Gujarat and Maharashtra had over 60% literacy as compared to 38% in the eastern state of Bihar. Although literacy was much lower in the early 20th century, the regional variation was very similar to that observed in 1991. Literacy rates in Bombay Presidency, which roughly represents contemporary Gujarat and Maharashtra, were almost twice as high as in Bihar. 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.[6] A better insight into the historical record can thus potentially guide current development policy as India strives to achieve universal literacy.

Third and finally, the Indian institutional setting offers a unique opportunity to disentangle demand side factors from public spending on education. Public educational expenditures varied dramatically both across and within provinces of British India. Due to historical circumstances,

---

[4] 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.
[5] See Davis and Huttenback (1986). Whitehead (2005) emphasizes the limited quantitative literature on colonial Indian education.
[6] 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.

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 played an important role in the computation of land revenues and lead to quasi-random variation across districts controlling for objective differences in agricultural and economic conditions. Land revenues thus offer an instrument for public spending because they strongly influenced expenditures on primary education but were unlikely to be systematically related to unobservable demand side factors after controlling for differences in soil, rainfall and development across districts.

I use a new historical dataset for the empirical exercise, which links district-level information on public expenditures from the Imperial district gazetteers to estimates of literacy from the colonial censuses. Controlling for province fixed effects and other observable differences across districts, both 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 with a significant upward bias on the OLS estimates. Although the 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.[7] The findings thus suggest that colonial public investments on education did in fact translate into better outcomes namely higher literacy.

To address concerns regarding the validity of the instrument, I perform a series of robustness

---

[7] 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% of the total population over this period.

checks. First, I use English literacy as a control outcome because in principle it should be uncorrelated with spending and land revenues because 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 English literacy. Second, 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. Third, I drop the districts of Bombay presidency where the computation of land revenues was perhaps less subjective than in other provinces. The IV results are robust to these tests. Finally, I explore the effects of spending on the production of literates namely the change in literacy rates between 1921 and 1911. Both, the level of spending in 1911 and the change in spending between 1901 and 1911 are positively correlated with changes in subsequent literacy.

The empirical results emphasize the positive benefits of increased public spending for literacy in British India. Qualitative evidence suggests that areas with higher district board expenditures had a larger number of better quality public schools that charged lower fees and effectively offered a bigger public subsidy to the population. Differences in colonial expenditures thus lead to marked differences in subsequent literacy across Indian districts with important ramifications for educational development in colonial and post-independence India. Broadly, this paper contributes to both the economics of education and the institutions literature by analyzing the specific historical circumstances that lead to disparate educational endowments across Indian districts during the colonial period, which closely resemble the contemporary variation in educational outcomes.

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.

4

2. Institutional Background

The East India Company followed by the British Crown introduced a new state system of education in British India in the 19th century, which largely replaced the former indigenous system of schooling.[8] From 1858 to 1919, the Crown through the Government of India directly controlled education policy although the financing and management of schools was decentralized to provincial governments beginning in the 1870's. Over this period, the provision of education was managed by British administrators such as provincial governors subject to financial constraints from the central government. However, after 1919 education was transferred to the control of elected Indian ministers in provincial legislatures (Dyarchy). The Indian ministers worked with somewhat constrained financial resources in the 1920's but with the introduction of Provincial Autonomy in 1935 education became a provincial subject and the legislatures operated with greater fiscal independence.[9]

Given the low level of mass education and high costs of developing a broad education system, the colonial government relied on both public and private revenues in the late 19th century to increase schooling in British India.[10] Table 1A shows nominal educational expenditures per-capita from 1881 to 1931 for British India as a whole and separately for some of the larger provinces.[11] There was a significant increase in total per-capita expenditures over this period with spending in 1931 being almost ten times higher than in 1881. Since prices increased anywhere from 22% to 59% between 1881 and 1931, spending thus increased in real terms as well.[12] Although this increase was significant, it should be interpreted with caution because British India had the lowest expenditures on

---

[8] 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.

[9] See Chand (1930), Misra (1960), and Kumar, ed. (1982) for details on the colonial fiscal system.

[10] See Nurullah and Naik (1951), Basu (1974) and Ghosh (2000) for details on colonial Indian education.

[11] Unfortunately, the geographic coverage of the series varies across these decades because data on a few princely states are included in the earlier series. The 1901 and 1911 series from the Quinquennial Reviews include data on the princely states of Bombay (almost all of them), Madras (1), Bengal (1), and Central Provinces (several). However, the princely states represent only 5% of the population covered in these earlier series. Bengal includes data for Bihar and Orissa pre-1911.

[12] Calculations based on index of Indian prices reported in Appendix Table 11 A.1 in Kumar, ed. (1982).

human capital for most of the period relative to the Princely States and other countries at comparable levels of development.[13]

A significant portion of the up take in total spending was driven by an increase in provincial revenues for education that are also reported in table 1A along with educational expenditures from district board revenues and private sources of revenue such as fees and endowments.[14] However, these disaggregated revenue series do not capture total educational expenditures incurred by either provincial governments or district boards for two reasons. First, provincial governments gave grants to district boards and the historical references are unclear as to what extent provincial grants are included under board revenues versus provincial revenues.[15] In most provinces, provincial grants appear to be included under provincial revenues. Second, provincial governments and district boards directly managed public schools that charged fees but the revenues from all fees are reported under educational expenditures from fees. Table 1B reports district board educational expenditures in 1895, 1909 and 1920, which suggests that expenditures from board revenues reported in table 1A severely underestimates actual spending by the boards. In effect, revenues from provincial governments, district boards and fees reported in table 1A jointly contributed to district board expenditures.

Ideally, one would like to evaluate the effects of both provincial and district board expenditures on educational outcomes to make an accurate assessment of the efficacy of public investments. Unfortunately, the historical sources only report district board educational expenditures for a large set of districts and consequently the empirical analysis focuses on this category of spending. Nonetheless, these expenditures capture an important public investment that was especially targeted to the development of rural primary education and more importantly other public spending did not offset the variation in board expenditures. In fact, the variation in board spending generally

---

[13] See Davis and Huttenback (1986) and discussion in Chaudhary (2008).

[14] Revenues from urban boards are not separately reported under the disaggregated revenue series but they contributed less than 5% to total educational expenditures.

[15] In few provinces such as Central Provinces it appears that part of the grants were included under district board funds but official reports are vague about how this category is calculated from the revenue information.

6

parallels total public spending because provincial revenues determined the amount of grants provincial governments allocated to district boards for education.

The marked heterogeneity in board expenditures across provinces as seen in table 1B was a function of historical circumstances and differences in provincial revenues. Both tables 1A and 1B indicate that districts of Bombay enjoyed a relatively high level of public spending on education in each decade from 1881 to 1931 as compared to districts of Bengal and Bihar and Orissa. This inter-provincial heterogeneity dates back to the mid-19th century when the fiscal system was extremely centralized (1833-70) and provincial governments had no fiscal independence with all revenues accruing to the central government, which then allocated funds to them. Historians have argued that the distribution of money from the center to the provinces was extremely random without any regard to the underlying economic situation of the province.[16] The initial disparities from the centralization period were then systematically strengthened over the late 19th century as financial decentralization was gradually introduced because subsequent negotiations between the Imperial government and the provinces often used past expenditures as a guide for future spending.[17]

Differences in land revenues were another key factor for the variation in provincial revenues. Beginning in 1882 provinces began to receive a share of their land revenues that were previously credited to the central government. On account of the Permanent Settlement of 1793 in Bengal and

---

[16] See Chand (1930, 1931) for details. Misra (1960) makes the same point based on statistics reported in Chand (1930). Various contemporaries of the period have also emphasized these differences as random. For example, according to Sir Richard Strachey, a British administrator, "the distribution of public income degenerated into something like a scramble, in which the most violent had the advantage, with very little attention to reason" (Chand 1930, page 34). In testimony before the Welby Commission in 1896, Mr. Gokhale indicated that "The fact is that these inequalities are a legacy from the pre-decentralization period, when the expenditures of different Provinces was determined … not by the resources or requirements of these Provinces, but by the attention that their Governments succeeded in securing from the Central Government i.e. by the clamour that they made" (Misra 1960, page 37).

[17] In fact, the distribution of revenues often exacerbated existing differentials with Bombay receiving substantially larger per-capita grants as compared to Bengal. For example, as part of the 1887 quinquennial settlement, provinces were given a share of their land revenues with Bombay initially receiving 60% of its land revenues and Bengal 32%. Moreover, these proportions underestimate the extent of the differential because on account of the Permanent Settlement, land revenues in Bengal were significantly lower than in Bombay. In subsequent settlements, the proceeds from land revenue were equally shared between central and provincial governments. See discussion in Kumar, ed. (1982).

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. In comparison, land revenues were revised every thirty years in Temporary Settlement areas to account for changing agricultural conditions and consequently they received higher public revenues. 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.[18] Although land tenures and Land Settlements were different across provinces, they were generally the same within provinces.

While land revenues contributed to inter-provincial heterogeneity in provincial revenues and consequently district board grants, they were equally important to differences in board expenditures across districts within the same province. Beginning in the 1880's, additional surcharges on land revenues were introduced to fund the newly constituted district boards. These surcharges known as cesses were generally levied at the rate of 6.25% and represented more than 50% of the revenues of district boards that managed the provision of rural primary education as well as other local services.[19] 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 the cesses in Permanent Settlement areas were based on the notoriously inaccurate land revenue assessments of 1793 because the raw correlation between 1911 per-capita cesses and land revenues is 0.9 for Permanent Settlement districts.[20]

---

[18] See Baden-Powell (1907) and Banerjee and Iyer (2005) for details on different land tenure arrangements.

[19] 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 more details on local boards.

[20] See Chand (1947) for details on cesses in Bengal. Roy (2000) among others has noted the poor quality of data in Permanent Settlements districts. In particular, land revenue assessments in Permanent Settlement Bengal did not

Idiosyncratic variation in cesses thus created substantial variation in district board revenues within the Permanent Settlement region.

In Temporary Settlement districts, the cesses were directly assessed on land revenues that were calculated as a proportion (roughly 50%) of monetized net agricultural produce in Madras, a function of general conditions adjusted for different soil types in Bombay and roughly 50% of rental assets in the United Provinces and Punjab.[21] Revenue rates were very high in the early to mid-19th century but declined over the late 19th and 20th century and represented less than 5% of net agricultural produce by the 1930's (Roy 2000 and Kumar, ed. 1982). The 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."[22] Classification of soils and water conditions were the few objective measures, which factored into the calculations.

While the literature on Indian economic history has emphasized the high level of revenue assessments and its implications for 19th century Indian agricultural distress, the relative inaccuracy of the system has received little attention.[23] Settlement officers set the revenue demand at each reassessment and the "idiosyncrasies of the Settlement Officers" lead to substantial perhaps 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

---

involve any surveys or classification of soils, but were largely a function of historical revenues and negotiations between Indian landlords and the East India Company. See Baden-Powell (1907) and Kumar, ed. (1982) for details.

[21] The cess was assessed on double the land revenue in the United Provinces and Punjab. See Baden-Powell (1907) 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.

[22] 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) and Chand (1931, 1947).

[23] 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.

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). And, this statement refers to Bombay—one of the more advanced provinces vis-à-vis revenue assessments and surveys—which, indicates the problem was worse in other provinces.

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 "much depends upon the personal judgment of the officers in charge of the work of assessment or reassessment" (Chand 1931, page 60). 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). Different soil types such as black, red, etc. and rainfall appear to be the only objective variables, which factored into the calculation.[24] Controlling for differences in soil and rainfall, the subjective nature of the assessment can then be viewed as a source of quasi-random variation in revenues across districts. Land revenues thus offer a potential 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.

Cesses and grants jointly accounted for 75% of the income of district boards as reported in table 1B, while the rest came from tolls, school fees and other miscellaneous sources. While cesses accounted for 60% of revenues in the 19th century, provincial education grants to district boards increased in the beginning of the 20th century and by 1929-30 they contributed 43% to income as compared to 36% from cesses.[25] Although land revenues contributed to both inter- and intra-provincial heterogeneity in cesses, they only contributed to differences in grants across provinces because within provinces grants were allocated to districts based on their population size with a

---

[24] Baden-Powell (1907) describes the general process of assessment in each province and classification of soils are important for the assessment in most provinces other than UP and Punjab where rents were the basis of assessment.
[25] Calculations based on Chand (1947).

stated preference for favoring poorer districts or districts with larger minority populations.[26] The grants were primarily responsible for the temporal variation in educational spending within districts because cesses remained largely unchanged from year to year due to the infrequent reassessment of land revenues.

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, the unsystematic and idiosyncratic computation of land revenues within provinces created quasi-random variation in educational expenditures across districts in the same province. Controlling for province fixed effects and using per-capita land revenues as an instrument for expenditures, I find that greater colonial public investments lead to significant gains in cohort literacy especially for males.

3. Empirical Framework

Although public educational expenditures accounted for more than 50% of total educational spending in British India, no study has systematically evaluated whether higher public 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. However, this position may partially reflect official strategy to absolve the British Government from any blame for the relatively low levels of public investments on human capital over the colonial period. 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.

---

[26] 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, which also supports the targeting of poorer districts.

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. One has to effectively disentangle the role of demand side factors from public expenditures to understand whether expenditures independently influence outcomes or whether they are reflective of some underlying private demand for education.

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 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 primary education expenditures and historical data is readily available on total and age 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 perhaps a small contribution from school fees collected in board schools. The across-district differences in expenditures were due to both inter- and intra-provincial factors. Although some of the inter-provincial differences in expenditures were due to historical circumstances, Indian provinces differed along several dimensions and exploiting differences across provinces may be especially vulnerable to criticisms of omitted variables. Therefore, the empirical analysis focuses on within province variation by introducing province fixed effects. However, an OLS estimation that

focuses on differences in expenditures and outcomes within provinces is still subject to endogeneity concerns.

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 primary school children in 1911 would on average fall in the cohort aged 15 to 20 in 1921 for which we have literacy data. Consequently, we can examine the effect of 1911 expenditures on 1921 cohort literacy 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, more urbanized districts may have a higher private demand for education and more public resources to supply schools. A failure to control for urbanization 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 of 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 given data constraints, it is the best available measure. I also introduce variables to capture the share of the population supported by commerce and industry because the 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.[27]

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, and minorities such as

---

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

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. Finally, I control for urbanization, population density and income to capture differences in local economic conditions across districts. 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.[28] They are thus representative of the higher tail of the income distribution.

Despite the controls and province fixed effects, the OLS coefficient may nonetheless be reflective of some dimension of unobservable heterogeneity across districts, which affects both literacy and public spending. Fortunately, the Indian context 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 boards. Given land revenues determined cesses, we expect them to be strongly correlated with district board public educational expenditures.

Table 2 examines the determinants of 1911 per-capita district board educational expenditures and land revenues emerge as the most important 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 other observable differences across provinces, a 10% increase in per-

---

[28] Beginning in 1886, only incomes above Rs. 500 were taxed. The tax rate was less than 2% for incomes between Rs. 500 and 2,000, and only marginally higher for incomes above 2,000. See Kumar, ed. (1982) for details.

capita land revenues leads to a 3.3% increase in average educational expenditures (specification 5) with an F-statistic of 10.2 on land revenues. Although a dummy for Temporary Settlement districts is significant in the across province comparisons (specifications 2 and 3), neither this variable nor the land tenure variable (proportion of non-landlord districts) from Banerjee and Iyer (2005) has any statistically significant effect on across district expenditures within provinces (specification 6).

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 perhaps enjoyed a larger agricultural surplus were assessed higher revenues. Settlement Officers may also have been affected by the income, development and social structure of the district at the time of the settlement. Given the inherent subjectivity of the system, land revenues should be unrelated to the underlying demand for literacy after controlling for differences in geography, development and social structure within provinces. I focus on the within province variation because the method of assessing revenues varied across provinces and there may be some concern that unobservable and perhaps endogenous provincial characteristics influenced the choice of method. 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 observable differences across districts within provinces. As seen in table 2, the results on land revenues are strong even after including the additional controls.

The Indian context also offers an alternative strategy to address the potential endogeneity problem of public expenditures. 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 educational expenditures as an additional robustness test to identify the effects of expenditures on outcomes. The historical sources do not

15

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 development over this decade. This is a variation of the traditional first difference estimation and effectively controls for time invariant district characteristics. Although the grants were not randomly assigned across districts, official documents indicate a preference for rewarding poorer districts and those with larger minority populations received larger grants. 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 on literacy, population and development from the colonial censuses with data from the Indian district gazetteers. The dataset includes information on the districts of Assam, Bengal, Bihar and Orissa, Bombay (without Sind), Central Provinces, Madras and United Provinces.[29] The pure urban cities of Bombay, Calcutta and Madras are excluded from the analysis.

Although literacy was inconsistently enumerated in the pre-1911 censuses, a uniform definition of literacy for British India was adopted beginning with 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 generally believed the enumeration was relatively accurate in subsequent censuses.[30] Other than this test, there were no formal schooling requirements. In principle a child could gain literacy skills at home but this was perhaps restricted to

---

[29] The dataset includes districts of Bengal that became incorporated into East Bengal (a province of Pakistan) that became the independent country of Bangladesh in 1971. District boards were not constituted in some districts of Assam and Bengal, and they are excluded from the analysis because they have missing expenditures.

[30] 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).

a very small share of the population given literacy was extremely low in this period and a vast

majority of the literate population attended some school. Given the enumeration became consistent

beginning in 1911, I focus on 1911 and 1921 literacy outcomes such as overall literacy and cohort

literacy in the population aged 10 to 20 and 15 to 20 disaggregated by gender.

Unfortunately, the age enumeration in the colonial censuses was plagued with inaccuracies

because individuals often did not know their age.[31] This is perhaps to be expected for a country like

India in the early 20th century since numeracy and literacy generally go together. The inaccuracy

however introduces measurement error in the measures of cohort literacy. Since official discussions

suggest that uncertainty was the primary culprit for incorrect enumerations, the measurement error is

likely to be random or classical in nature. And, classical measurement error generally yields

consistent estimates.

In addition to literacy, I constructed measures of development namely urban population,

population density, and occupational structure from the 1911 and 1921 censuses. There are several

concerns regarding the accuracy of the smaller occupational categories and so I focused on broad

occupational types—the share of the population supported by agriculture, commerce, industry and

professionals—to minimize measurement error. I also extracted information on important caste and

religious groups namely Muslims, Christians, tribes, Brahmans and lower castes.[32] Moreover, the

census provides the necessary information to construct a measure of caste and religious

fragmentation, similar to the ethnic-fragmentation index used in the literature. This variable is

---

[31] 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.

[32] 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. If many castes were successful in enumerating themselves as higher castes, we would expect the population share of upper castes to perhaps substantially increase over time. However, Brahman and upper caste population shares are fairly stable over this period. In general, measurement error in the caste variables should attenuate the estimates toward zero.

constructed using data from the 1901 census but given the stability of population structures probably correlates well with 1911 patterns. To control for differences in soil type and geography, I used the World Bank Agricultural and Climate dataset to construct dummies for black, red and alluvial soil. [33] I also extracted data on normal rainfall from the 1911 census and created a dummy for coastal districts based on historical maps.

The district gazetteers are a unique source of data for this period. Each district series has two parts, A and B: 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 a large sample of districts, the expenditure statements of rural district boards report educational expenditures. Given rural boards managed the provision of primary education, this was an important category of public spending as discussed in section 2. I also extracted information on income tax and land tax revenues collected in each district. These variables were constructed for the 1901 and 1911 cross-section but not 1921 because some of the provinces stopped reporting this information after 1916. Table 3 presents summary statistics of the variables used in the analysis. Despite the increase in expenditures observed in table 1A, overall literacy rates did not increase dramatically from 1911 to 1921. High rates of dropout, wastage and relapses into illiteracy were frequently blamed for the lack of progress. Most of the population variables are fairly stable between 1911 and 1921 as are many of the development controls with urbanization barely increasing from 8.7% to 9.7% over the decade.

5. Results

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

5.1 Ordinary Least Squares Results

Table 4 presents OLS results of district board expenditures on overall literacy and cohort

---

[33] Data on soil type is from the post-independence period.

literacy among the population aged 10 to 20 and 15 to 20. The regressions include province fixed effects as well as controls for soil type, rainfall, coastal districts, income, social structure, urbanization and development. Panel A focuses on the 1911 cross-section and reports the effects of 1911 expenditures on 1911 outcomes, while panel B focuses on the 1921 cross-section but uses 1911 expenditures as the key measure of public investments. Although grants and surcharges on land revenues accounted for the bulk of district board income, the boards also received school fees from public schools as additional income. 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 may thus have more children attending schools and paying fees. Despite the extensive set of controls included in table 4 to address this concern, the OLS coefficients on expenditures appear to be biased.

1911 educational expenditures are positively correlated with overall and cohort literacy in both 1911 and 1921. According to official sources, a child became proficient in reading and writing (i.e. literate) after completing four years of primary school. If we assume that a child begins primary schooling at age 6, in principle 1911 educational expenditures should be uncorrelated with 1911 cohort literacy in the population aged 10 to 20 and especially those aged 15 to 20 unless expenditures are endogenous to literacy. The positive and statistically significant results on 1911 expenditures in panel A are troubling because they suggest that the OLS estimates are perhaps endogenous to unobservable factors. If expenditures were exogenous to literacy, 1911 board expenditures should be unrelated to 1911 cohort literacy in the population aged 15 to 20 that attended primary school at least 5 to 10 years earlier. To address this endogeneity problem, the next set of results focus on instrumental variables to identify the causal effect of expenditures on literacy.

5.2 Instrumental Variables Results

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

19

were an important determinant of educational expenditures controlling for provincial dummies as seen in table 2. Moreover, after controlling for objective difference in agricultural (i.e. soil type, rainfall, coastal) and economic (i.e. income and development) conditions, the variation in land revenues was largely due to idiosyncratic differences in computation that should be uncorrelated to the error term. Land revenues thus offer a potential instrument for educational expenditures.

Table 5 presents the 2SLS results of 1911 expenditures for both the 1911 and 1921 cross-sections controlling for province fixed effects. [34] The lack of statistical significance on the 1911 expenditures in panel A is reassuring because it suggests that after controlling for the endogeneity of expenditures via the instrument, 1911 expenditures are uncorrelated with 1911 literacy outcomes for the cohort aged 15 to 20, which is as we would expect. Moreover, 1911 expenditures lead to higher literacy for the 15 to 20 aged cohort that was in primary school 10 years earlier and would presumably have been most affected by higher expenditures in 1911. Although the OLS results found a positive effect of public expenditures on female literacy, the IV results indicate that higher board expenditures did not translate into better outcomes for women.

The IV estimates are generally smaller in magnitude as compared to the OLS estimates reported in table 4, which suggests that the endogeneity of expenditures generated an upward bias on the OLS estimates. Nonetheless, the magnitude of the IV estimates is striking—a 10% increase in 1911 per-capita district board spending translates into a 2.3 percentage point increase in literacy rates among the population aged 15 to 20 with substantially larger effects for male literacy in the population aged 15 to 20. In addition to public investments, tables 4 and 5 also report the coefficients on two key demand side factors namely the share of the urban population and professionals. Both these variables strongly influenced male and female literacy with similar magnitudes across the OLS and IV specifications. Differences in returns to literacy were thus also critical to outcomes.

---

[34] The IV estimates are similar in specifications without province fixed effects except the coefficients are larger in magnitude.

The findings suggest that colonial public investments on primary education lead to substantial gains in literacy even controlling for the endogeneity of expenditures. However, there may be some concern that land revenues were correlated with the unobservables affecting literacy even though the within province variation in revenues was largely a function of historical idiosyncrasies in calculation. The top panel of table 6 therefore presents additional robustness tests.

First, primary schools offered instruction in the vernacular medium as compared to secondary schools where English medium instruction was more common.[35] If the unobservable factors affecting the returns to English language literacy are similar to those for literacy in any other language (i.e. overall literacy), 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 unobservable factors that are positively associated with a general demand for education. Specifications 1 and 2 report the findings on English literacy among the population aged 10 to 20 and 15 to 20. As is evident from table 6, the IV estimates on 1911 expenditures are small in magnitude and statistically insignificant. Moreover, we can reasonably infer that the demand for English and non-English literacy was affected by similar forces because the same set of demand side factors namely urbanization and professionals are positively correlated with 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.

Second, I test for specific endogeneity problems in specifications 3, 4 and 5. One important concern relates to how district board expenditures on education correlated with expenditures on other public goods such as local infrastructure. 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

---

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

could substitute between different services. If areas with high public educational expenditures also spend more money on other services due to higher land revenues, then this would generate an upward bias on both the OLS and IV estimates because higher quality medical services may have an independent effect on literacy.[36] However, the OLS and IV estimates on educational spending are robust to the inclusion of 1911 medical and non-educational expenditures. Specifications 3 and 4 in table 6 only report the findings of the test on the IV estimates.

Another variable that may be correlated with land revenues, expenditures and literacy is 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*). Thus, it could be the case that non-landlord systems generated higher land revenues and a higher demand for basic literacy because cultivators in these districts directly negotiated 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 the post-independence period. However, they find limited effects when they control for state fixed effects because there was limited variation in tenures within states or provinces. Specification 5 includes the Banerjee and Iyer (2005) variable for non-landlord districts and the results on expenditures are unchanged.

Kumar (1982) argues that the assessment of land revenues was most pragmatic in Bombay because it accounted for changing agricultural conditions (albeit every thirty years!). This raises some concern that land revenues across Bombay districts are perhaps endogenous to literacy despite the extensive controls for objective differences across districts. To address this issue, specification 6 in the top panel of table 6 drops the 18 districts of Bombay where the assessment of land revenues could in principle be correlated with the error term. As is evident, the results on expenditures are

---

[36] 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.

robust to the exclusion of Bombay districts.

While the robustness tests focused on the IV estimates, the 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 the 160 districts used in the analysis. And, the temporal variation in expenditures over the 1901 and 1911 decade was largely due to increases in educational grants from provincial governments to district boards. The bottom panel of table 6 exploits this variation in expenditures and focuses on an alternate dependent variable namely the production of literates or changes in the literacy rate for the population aged 15 to 20 between 1921 and 1911. Specifications 1 and 2 focus on the level effects of expenditures on the production of literates and find a statistically significant effect of 1911 expenditures on the production of literates in the 15 to 20 aged cohort ten years later. Specifications 3 to 6 focus on changes in expenditures between 1901 and 1911 and find that this variable is positively correlated with the production of literates between 1921 and 1911 controlling for changes in urbanization, population density, social and occupational structure over this decade. Variables such as rainfall and soil type that do not vary between 1921 and 1911 are dropped from the regressions. If we had access to reliable literacy data for 1901 or expenditures for 1921, we could run a first difference regression. In the absence of that data, these regressions are a second best option and in a way they are essentially controlling for time-invariant district characteristics.

5.4 Discussion

The empirical analysis finds that colonial public investments lead to better educational outcomes such as higher literacy particularly for males. IV estimates of 1911 expenditures find a positive effect on 1921 literacy in the population aged 15 to 20, which is the cohort most likely to benefit from higher expenditures 10 years earlier. Moreover, within district changes in expenditures are also positively correlated with the production of literates aged 15 to 20 ten years later.

Although this exercise establishes that expenditures had a positive effect on cohort literacy, it

is unclear on the specific mechanism linking the two. Chaudhary (2008) finds that districts with higher land revenues per-capita had more publicly funded and managed primary board schools. And, qualitative discussions in the Quinquennial Reviews of Education suggest that these schools were often of higher quality with a larger proportion of trained teachers and often 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 them on to the students. This suggests that higher expenditures translated into better outcomes through higher quality public schools that offered a larger public subsidy to the rural population in the form of lower fees.

Even though higher public expenditures on primary education would have positively impacted literacy, the colonial government may have been reluctant to increase aggregate public spending perhaps because it was not cost-effective. The estimates in this paper provide a measure of the cost of literacy. A rough calculation suggests that a 10% increase in 1911 district board expenditures translates into an overall increase of Rs. 7260 using average 1911 district population size of 1,210,073 from the data. Moreover, a 10% increase in expenditures leads to a 2.4 percentage point increase in 1921 literacy among the population aged 15 to 20, which is equivalent to 2413 additional people. This back of the envelope calculation then suggests that it would have roughly cost the colonial government 3 rupees to make an additional person literate. Unfortunately, it is difficult to assess the returns to basic literacy in the colonial period because the historical sources do not report information on wages disaggregated by education or literacy. Nonetheless, the cost of 3 rupees does not appear prohibitively expensive especially in an institutional environment where more than 50% of total expenditures were directed to secondary schools and colleges. If public expenditures on secondary schools had been allocated to primary education, it would have increased overall literacy by 2.4 percentage points from say 6.1 % to 8.5 % over a decade.[37]

---

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

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 literacy using an instrumental variables strategy. A 10 percent increase in spending is associated with a 2.4 percentage point increase in literacy rates for the population aged 15 to 20 and the findings are robust to a variety of falsification tests. Qualitative evidence suggests that greater reliance on higher quality public schools with lower fees is perhaps a key mechanism from higher spending to greater literacy. While these findings are important for Indian economic history, they also have broader implications.

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% as compared to the national average of 5.6%) 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. The findings of this paper suggest that Indian districts inherited a certain educational endowment at Independence and differences in colonial investments were partially responsible for the 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 Bihar and Orissa that unfortunately have strong and persistent historical roots.

References

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.

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.

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

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.

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 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
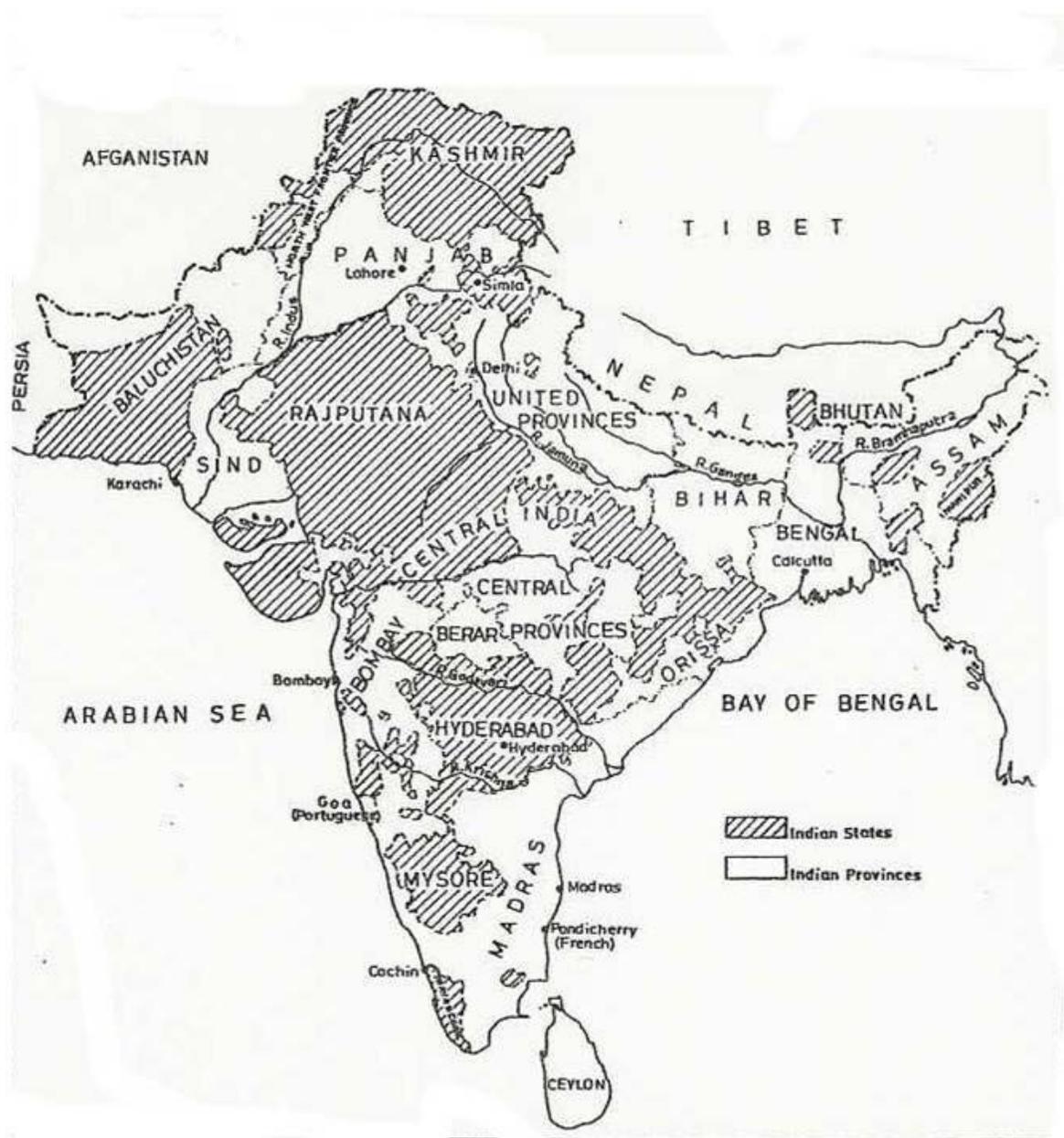
TABLE 1A: EDUCATIONAL EXPENDITURES PER 1000 OF THE POPULATION

	1881-82	1891-92	1901-02	1911-12	1921-22	1931-32
Total Expenditures:						
BRITISH INDIA	**95**	**139**	**167**	**211**	**744**	**1,000**
Bengal	96	131	160	384	715	844
Bihar & Orissa				170	339	462
Bombay	207	332	306	397	1,530	1,832
Central Provinces and Berar	79	97	108	147	584	669
Madras	112	165	208	228	803	1,214
Punjab	84	128	160	214	917	1,307
United Provinces	46	71	96	141	657	804
Expenditures from Provincial Government Revenues:						
BRITISH INDIA	**37**	**40**	**49**	**67**	**365**	**458**
Bengal	40	35	33	85	290	288
Bihar & Orissa				43	145	148
Bombay	68	104	131	190	878	869
Central Provinces and Berar	45	45	44	43	368	307
Madras	32	49	56	65	374	547
Punjab	33	42	45	59	420	699
United Provinces	21	12	22	46	344	450
Expenditures from District Board Revenues:						
BRITISH INDIA	**14**	**24**	**24**	**31**	**68**	**103**
Bengal	0.15	15	17	41	30	33
Bihar & Orissa				36	67	124
Bombay	49	46	38	30	60	104
Central Provinces and Berar	17	22	26	60	74	125
Madras	18	23	20	26	67	143
Punjab	24	29	27	46	123	115
United Provinces	14	31	31	36	77	76
Expenditures from Fees and Private Funds (Endowments, etc.):						
BRITISH INDIA	**42**	**68**	**87**	**104**	**279**	**381**
Bengal	56	80	109	254	390	493
Bihar & Orissa				88	122	179
Bombay	85	163	120	143	415	606
Central Provinces and Berar	14	24	30	30	101	176
Madras	59	87	125	131	345	472
Punjab	19	41	72	96	327	438
United Provinces	11	27	40	52	217	246

Source: Statistical Abstracts for 1881 and 1891 series. Progress of Education in India, Quinquennial Reviews (from 1897-1902 to 1927-1932) for 1901 to 1931 series. Bihar and Orissa are included under Bengal from 1881-1901. 1911 data for Bengal, Bihar and Orissa are for 1916-17 due to reporting issues with the 1911 province boundaries. The 1901 and 1911 series also include data on some princely states. All figures are nominal.

TABLE 1B: DISTRICT BOARD EDUCATIONAL EXPENDITURES AND REVENUES PER 1000 OF THE POPULATION

| | | | | | | | INCOME | | | | | | |
| | 1895-96 | 1909-10 | 1920-21 | 1895-96 | 1909-10 | 1920-21 | 1895-96 | 1909-10 | 1920-21 | 1895-96 | 1909-10 | 1920-21 |
	Educational Expenditures			Cess - Additional Surcharge on Land Revenues			Provincial Grants			Total Income		
Bengal	17	15	61	51	45	144	7	20	54	83	78	231
Bihar and Orissa			56			185			18			280
Bombay	82	121	360	152	173	174	56	131	313	265	343	524
Central Provinces	26	61	151	37	75	62	11	50	186	75	174	363
Madras	33	34	111	150	157	178	9	40	156	240	301	540
Punjab	31	62	215	111	139	212	4	68	212	147	267	568
United Provinces	27	60	138	40	77	157	24	46	119	81	172	327

Source: Gyan Chand (1947), *Local Finance in India* (Allahabad: Kitabistan, 1947), tables Appendix VI, Appendix X and Appendix XII. Population data was obtained from the Statistical Abstracts. Chand (1947) is unclear whether Bengal in 1909-10 excludes Eastern Bengal, which was a separate province merged with Assam over these years on account of the partition of Bengal in 1905. I have used the population figures for Bengal Proper, Bihar and Orissa but this could generate some mis-measurement in the Bengal per-capita numbers for 1909-10 if Chand (1947) is using figures for Bengal excluding Eastern Bengal.

TABLE 2: DETERMINANTS OF 1911 DISTRICT BOARD EDUCATIONAL EXPENDITURES

	(1)	(2)	(3)	(4)	(5)	(6)
Land Tax Revenues per-capita	0.0321	0.0286	0.0267	0.0204	0.0212	0.0214
	[0.0054]***	[0.0071]***	[0.0079]***	[0.0063]***	[0.0066]***	[0.0075]***
Temporary Settlement		0.019	0.0158		-0.002	-0.0065
		[0.0071]***	[0.0088]*		[0.0074]	[0.0113]
Income Tax Revenues per-capita		0.021	0.0289		0.1427	0.1501
		[0.0972]	[0.0978]		[0.0861]*	[0.0917]
Population Density		-0.0306	-0.0336		-0.0103	-0.0106
		[0.0104]***	[0.0130]**		[0.0082]	[0.0107]
Fraction Urban		0.1699	0.1726		-0.0116	-0.0178
		[0.0583]***	[0.0669]**		[0.0440]	[0.0520]
Proportion of Non-Landlord			0.0104			-0.0024
(Banerjee and Iyer 2005)			[0.0082]			[0.0091]
Geographic Controls	No	Yes	Yes	No	Yes	Yes
Social and Economic Controls	No	Yes	Yes	No	Yes	Yes
Province Fixed Effects	No	No	No	Yes	Yes	Yes
Constant	0.0153**	0.0628**	0.0463	0.0454***	0.0732	0.0828
	[0.0067]	[0.0290]	[0.0358]	[0.0132]	[0.0499]	[0.0639]
Observations	161	160	141	161	160	141
R-squared	0.56	0.73	0.75	0.80	0.85	0.85
F-test: Land Tax Revenues = 0	35.58	16.28	11.34	10.45	10.16	8.19
Prob > F	0.0000	0.0001	0.001	0.0015	0.0018	0.005

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, population share supported by professionals, share of Brahmans, lower castes (1901), Christians, Muslims, Tribes and a measure of caste and religious fragmentation (1901). Unless specified control variables are for the same year as the outcomes.

TABLE 3: SUMMARY STATISTICS

Variable	1911		1921	
	Mean	Std. Dev	Mean	Std. Dev
Census Variables				
Overall literacy rate	5.3%	2.7%	6.1%	3.0%
Overall male literacy rate	9.7%	4.8%	10.9%	5.2%
Overall female literacy rate	0.7%	0.6%	1.2%	1.0%
Literacy rate aged 10-20	6.6%	3.4%	7.7%	3.9%
Male literacy rate aged 10-20	11.2%	5.7%	12.5%	6.1%
Female literacy rate aged 10-20	1.3%	1.2%	2.1%	1.8%
Literacy rate aged 15-20	7.7%	3.7%	9.4%	4.4%
Male literacy rate aged 15-20	13.6%	6.4%	15.9%	7.2%
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	8.9%	7.2%	9.7%	7.9%
Population Density	409	260	407	266
Fraction Commerce	6.1%	2.4%	5.9%	2.4%
Fraction Industry	10.7%	5.0%	10.0%	4.9%
Fraction Professionals	1.4%	0.7%	1.4%	0.8%
Fraction Brahman	5.7%	4.4%	5.5%	4.4%
Fraction Muslim	16.3%	19.4%	16.6%	19.6%
Fraction Christian	0.9%	1.7%	1.0%	1.9%
Fraction Tribes	3.8%	9.0%	3.4%	8.2%
Fraction Lower Castes (1901)	14.2%	8.6%	14.2%	8.6%
Caste & Religious Fragmentation Index (1901)	0.79	0.14	0.79	0.14
District Gazetteer Variables				
1911 District Board Educ. Exp per-capita	0.06	0.04	.	.
1911 Land Tax Revenues per-capita	1.42	1.00	.	.
1911 Income Tax Revenues per-capita	0.05	0.05	.	.
No. of Observations	161	161	161	161

Sources: Census of India (1901 and 1911) and Imperial District Gazetteer Series (1914).

The dataset includes districts of Assam, Bengal (including those in present day Bangladesh), Bombay, Central Provinces and Berar, Madras and United Provinces with functioning district boards. Local boards were not constituted in a few districts of Bengal and the hill districts of Assam.

Population Density is constructed as Population/Area in 1921.

TABLE 4: OLS ESTIMATES OF PUBLIC EXPENDITURES ON LITERACY

	Overall Literacy			Cohort 10-20			Cohort 15-20		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Panel A: 1911 Cross Section									
District Board Educ Exp	0.2003	0.3326	0.0372	0.3587	0.5462	0.0995	0.3371	0.5142	0.0915
per-capita (1911)	[0.0631]***	[0.1185]***	[0.0193]*	[0.0898]***	[0.1493]***	[0.0379]***	[0.0856]***	[0.1446]***	[0.0410]**
Fraction Urban	0.1346	0.2219	0.0216	0.1389	0.2224	0.0405	0.167	0.278	0.0419
	[0.0396]***	[0.0656]***	[0.0127]*	[0.0488]***	[0.0778]***	[0.0213]*	[0.0473]***	[0.0787]***	[0.0207]**
Fraction Professionals	0.0078	0.0149	0.0012	0.0084	0.0152	0.0013	0.0108	0.0205	0.0017
	[0.0030]**	[0.0051]***	[0.0010]	[0.0041]**	[0.0065]**	[0.0018]	[0.0039]***	[0.0065]***	[0.0019]
Observations	160	160	160	160	160	160	160	160	160
R-squared	0.85	0.84	0.75	0.82	0.81	0.75	0.84	0.84	0.75
Panel B: 1921 Cross Section									
District Board Educ Exp	0.2463	0.4215	0.0431	0.4139	0.6438	0.1281	0.4554	0.7191	0.1231
per-capita (1911)	[0.0615]***	[0.1076]***	[0.0323]	[0.0983]***	[0.1435]***	[0.0641]**	[0.0977]***	[0.1501]***	[0.0644]*
Fraction Urban	0.1502	0.2447	0.0259	0.1387	0.2228	0.0339	0.1589	0.2524	0.0424
	[0.0315]***	[0.0519]***	[0.0140]*	[0.0398]***	[0.0604]***	[0.0247]	[0.0444]***	[0.0739]***	[0.0254]*
Fraction Professionals	0.0087	0.0134	0.0038	0.0108	0.0137	0.0062	0.0133	0.0175	0.0065
	[0.0033]***	[0.0059]**	[0.0013]***	[0.0042]**	[0.0068]**	[0.0023]***	[0.0045]***	[0.0081]**	[0.0024]***
Observations	160	160	160	160	160	160	160	160	160
R-squared	0.83	0.82	0.75	0.81	0.8	0.76	0.82	0.81	0.76

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions in panels A and B include province dummies, normal rainfall, dummies for coastal districts, alluvial soil, red soil and black soil, income tax revenues per capita (1911), the share of the urban population, population share supported by industry, population share supported by commerce, share of Brahmans, lower castes (1901), Christians, Muslims, Tribes and a measure of caste and religious fragmentation. Unless specified the control variables are for the same year as the outcomes.

TABLE 5: IV ESTIMATES OF PUBLIC EXPENDITURES ON LITERACY

	Overall Literacy			Cohort 10-20			Cohort 15-20		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Panel A: 1911 Cross Section									
District Board Educ Exp	0.0767	0.1085	0.0143	0.1804	0.2461	0.066	0.1803	0.2572	0.0592
per-capita (1911)	[0.0915]	[0.1712]	[0.0240]	[0.1237]	[0.2119]	[0.0464]	[0.1300]	[0.2121]	[0.0481]
Fraction Urban	0.1309	0.2152	0.021	0.1336	0.2134	0.0395	0.1073	0.1634	0.0379
	[0.0389]***	[0.0648]***	[0.0124]*	[0.0482]***	[0.0775]***	[0.0210]*	[0.0508]**	[0.0781]**	[0.0223]*
Fraction Professionals	0.0095	0.018	0.0015	0.0109	0.0194	0.0018	0.0089	0.0152	0.0015
	[0.0031]***	[0.0053]***	[0.0010]	[0.0043]**	[0.0067]***	[0.0018]	[0.0048]*	[0.0072]**	[0.0019]
Observations	160	160	160	160	160	160	160	160	160
R-squared	0.84	0.83	0.74	0.81	0.8	0.75	0.77	0.76	0.73
Panel B: 1921 Cross Section									
District Board Educ Exp	0.1186	0.1954	0.0154	0.2455	0.3687	0.0832	0.2389	0.3715	0.0746
per-capita (1911)	[0.0929]	[0.1672]	[0.0376]	[0.1164]**	[0.1873]*	[0.0687]	[0.1111]**	[0.1692]**	[0.0683]
Fraction Urban	0.1378	0.2227	0.0232	0.1224	0.196	0.0296	0.1034	0.1667	0.0217
	[0.0319]***	[0.0542]***	[0.0133]*	[0.0404]***	[0.0628]***	[0.0239]	[0.0385]***	[0.0563]***	[0.0238]
Fraction Professionals	0.0105	0.0165	0.0041	0.0131	0.0175	0.0068	0.0109	0.0141	0.0067
	[0.0035]***	[0.0060]***	[0.0015]***	[0.0045]***	[0.0070]**	[0.0026]***	[0.0044]**	[0.0064]**	[0.0027]**
Observations	160	160	160	160	160	160	160	160	160
R-squared	0.82	0.81	0.74	0.8	0.79	0.75	0.78	0.77	0.74

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions in panels A and B include the same set of controls as in table 4. District board educational expenditures per-capita in 1911 are instrumented using 1911 per-capita land revenues. See text for details. Unless specified the control variables are for the same year as the outcomes.

TABLE 6: ROBUSTNESS CHECKS

| | 1921 Eng. Literacy Rate | | 1921 Literacy Rate | | | |
| | 10-20 | 15-20 | Cohort 15-20 | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
District Board Educ Exp	-0.0581	-0.0600	0.2475	0.2509	0.3248	0.2591
per-capita (1911)	[0.0359]	[0.0448]	[0.1472]*	[0.1384]*	[0.1325]**	[0.1324]*
Fraction Urban	0.0445	0.0564	0.1353	0.1321	0.1698	0.1244
	[0.0095]***	[0.0121]***	[0.0457]***	[0.0446]***	[0.0472]***	[0.0482]**
Fraction Professionals	0.0035	0.0042	0.0159	0.0157	0.014	0.0179
	[0.0012]***	[0.0011]***	[0.0048]***	[0.0048]***	[0.0052]***	[0.0050]***
District Board Non-Educ Exp.			0.0159			
per-capita (1911)			[0.0339]			
District Board Medical Exp.				0.1217		
per-capita (1911)				[0.2125]		
Proportion of Non-Landlord					0.0019	
(Banerjee and Iyer 2005)					[0.0061]	
Observations	160	160	160	160	141	142
R-squared	0.77	0.79	0.81	0.82	0.83	0.82

| | Δ Literacy Rate (1921-1911) | | | | | |
	10-20	15-20	10-20	Male 10-20	15-20	Male 15-20
District Board Educ Exp	0.0819	0.1179				
per-capita (1911)	[0.0659]	[0.0609]*				
Δ Educational Expenditures (1911-1901)			0.1803	0.2509	0.2759	0.4651
			[0.0840]**	[0.1276]*	[0.0906]***	[0.1393]***
Δ Urban (1921-11)			0.1408	0.1781	0.1963	0.2334
			[0.0656]**	[0.0976]*	[0.0812]**	[0.1171]**
Δ Professionals (1921-11)			0.0047	0.0076	0.0064	0.0116
			[0.0016]***	[0.0027]***	[0.0023]***	[0.0040]***
Observations	160	160	154	154	154	154
R-squared	0.45	0.57	0.41	0.3	0.53	0.43

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. 1911 district board educational expenditures per-capita in the top panel and first two specifications of the bottom panel are instrumented using 1911 per-capita land revenues and they include the same set of controls as tables 4 and 5. Unless specified the control variables are for the same year as the outcomes. Specifications 3 to 6 in the bottom panel include the change in the share of Brahmans, Muslims, Christians, commercial population and population density between 1921 and 1911.