

# *Common Tongue: The Impact of Language on Educational Outcomes*

TARUN JAIN

This article investigates the impact of official language policies on education using state formation in India. Colonial provinces consisted of some districts where the official language matched the district's language and some where it did not. Linguistically mismatched districts have 18.0 percent lower literacy rates and 20.1 percent lower college graduation rates, driven by difficulty in acquiring education due to a different medium of instruction in schools. Educational achievement caught up in mismatched districts after the 1956 reorganization of Indian states on linguistic lines, suggesting that political reorganization can mitigate the impact of mismatched language policies.

Language policies are associated with a great deal of debate, policy action, and conflict. The use of uniform languages can lower the cost of communication, facilitate education, and expand economic growth. This lesson is quickly learned by new immigrants whose academic and economic success depends critically on mastering the *lingua franca* of their new home. At the same time, attempts to impose official languages can meet considerable resistance, both because language is an important component of identity, and because learning a new language is difficult.<sup>1</sup> This article examines the consequences of language mismatch in education, that is, when students are taught in a language, usually an official state language, that differs from the mother tongue. Specifically, I ask if language mismatch benefits or hinders long-term educational achievement. Second, if language mismatch hinders long-term educational achievement, does fixing this mismatch by changing the medium of instruction to align with students' mother tongue lead to catch-up in educational achievement?

*The Journal of Economic History*, Vol. 77, No. 2 (June 2017). © The Economic History Association. All rights reserved. doi: 10.1017/S0022050717000481

Tarun Jain is Assistant Professor of Economics and Public Policy, Indian School of Business, Hyderabad 500032 India. E-mail: tj9d@virginia.edu.

Urvashi Jain, Tilak Mukhopadhyay, Lakshmi Sripada, and Revathy Suryanarayana provided excellent research assistance. This paper benefited from feedback from Tanika Chakraborty, Sisir Debnath, Mukesh Eswaran, Leora Friedberg, Maitreesh Ghatak, Lakshmi Iyer, Krishna Kumar, John Leahy, Pushkar Maitra, Subha Mani, Prachi Mishra, Kaivan Munshi, Nathan Nunn, John Pepper, Rohini Somanathan, and Ashima Sood. I thank numerous seminar and conference audiences for helpful comments.

<sup>1</sup> For example, adopting Sinhalese as the official language led to civil war in Sri Lanka not only because of the affront to identity, but also because Tamils who did not speak Sinhalese lost government positions and power. Similarly, separatist movements in Quebec, Belgium, and Catalonia were also rooted in resistance to dominant languages.

If the language used in schools is the same as the learner's mother tongue, then educational achievement is greater as residents are more likely to enrol, understand instruction, and complete different levels of schooling. For instance, Hoyt Bleakley and Aimee Chin (2004) find that childhood immigrants to the United States from English-speaking countries report significantly greater educational achievement compared to those from non-English backgrounds. Language fluency might also be key for unlocking future economic opportunities. For example, the returns to post-secondary education declined by half when Morocco switched from French to Arabic as the medium of instruction in schools, primarily because most organized economic activity was conducted in French (Angrist and Lavy 1997).<sup>2</sup> David Clingingsmith (2014) argues that economic growth in sectors where communication is relatively important increases the incentive to learn new languages. Consequently, growth of the manufacturing sector increased bilingualism in mid-century India, especially among minority language speakers.

Language mismatch might also influence education through the provision of public schools, although the direction of this effect is not clear. If speakers of the majority language control public spending, they might construct more schools in the areas where the language is spoken because of political patronage or shared preferences, leading to higher educational achievement. Conversely, school construction and spending on educational infrastructure might be directed at economically weaker regions. If minority language regions are also economically weaker, then these areas would experience greater improvements in education achievement.<sup>3</sup> Finally, language mismatch might be associated with greater migration if individuals who do not speak the official language of a region relocate to a different region seeking education or economic opportunities in their own language.

The main contribution of this article is to examine the effect of language mismatch on educational achievement using large-scale historical events in India. Primary schooling in India is largely in the official language of a state or province, both in the colonial period (before 1947) as well as in the modern era. Colonial-era provinces, especially in South India, were formed without regard to language, which resulted in some areas where

<sup>2</sup> Similarly, using a policy change in West Bengal which revoked English as a medium of instruction in government-operated schools, Chakraborty and Kapur (2016) report that a 1 percent decrease in the probability of learning English decreases wages by 1.6 percent.

<sup>3</sup> Alesina and La Ferrara (2005) survey empirical studies on the link between ethno-linguistic diversity and public goods provision. In India, Banerjee, Iyer, and Somanathan (2005) examine multiple social cleavages and find that direct British control, landlord-based tax collection systems, and greater ethnic divisions are all associated with lower public good availability.

schools taught in the same language as the mother tongue of the majority of students, and other areas where they did not. If the cost of completing formal education is lower when instructed in one's mother tongue, then I expect long-term educational attainment to be lower in districts that have a different majority language from the official language. This illustrates the degree to which variation in language explains persistent underachievement in education within countries.

In the immediate post-Independence era, the 1956 reorganization of South Indian states on strict linguistic lines aligned the majority language of districts with the official language of states, so that almost all children were taught in their mother tongue. As a result, a convergence in education levels between the previously aligned and misaligned districts helps us to understand whether "fixing" language policies through political reorganization can address persistent educational shortfalls.

India is a particularly appropriate setting for studying the impact of language since more than 1,600 languages are used in the country, with 22 languages claiming both wide speakership as well as constitutional recognition (Registrar General 2001). The boundaries of modern Indian states correspond to the areas where these languages are used. Public schools offer instruction using the medium of the state language, and regulators, the judiciary and other arms of state governments use the state language for official correspondence with citizens. Thus, the setting permits comparison between multiple sets of major vernacular languages, each with tens of millions of speakers. As a result, the article circumvents the comparison of a colonial language, especially English, with a vernacular language which is troublesome both because then the returns from globalization cannot be isolated from pure language effects, and because populations who speak these languages are often systematically different.<sup>4</sup>

There are two major empirical challenges in generating reliable estimates of the relationship between language mismatch and education. The primary challenge is that bivariate comparisons between communities that either speak or do not speak the official language might not yield unbiased estimates of the relationship between language mismatch and education because of, at least, two potential sources of endogeneity. First, individuals who do not speak the official language may move to ethnic enclaves where they do not need to learn a new language. Second, communities that recognize the link between speaking a different language and poorer economic outcomes may form their own political units. For example,

<sup>4</sup> Azam, Chin, and Prakash (2013) and Jain, Maitra, and Mani (2016) discuss returns to spoken English skills in contemporary India.

linguistic minorities in countries such as Spain, Canada, and Cameroon launched separatist movements based on language. Thus, unbiased estimates require exogenous matching between languages and communities.

The pattern of British conquest and province formation in the eighteenth century helps generate such estimates. Provincial boundaries in British India were determined either by the sequence of British military conquest, with provinces cobbled together from various districts as imperial rule extended from the coasts into the hinterland, or when the British decided to leave native rulers in place. I argue that this process was largely exogenous to linguistic concerns, leading to the inclusion of some districts in provinces where the district's numerically dominant mother tongue language was the same as the official language of the province (henceforth, a "majority" district), and other districts in provinces where it was not (a "minority" district). If mother tongue instruction facilitates schooling, then historically majority districts should have had better educational outcomes compared to historically minority districts. Insofar that educational achievement persists over generations, minority districts could experience poorer educational outcomes even up to modern times, indicating the persistent effects of language mismatch.<sup>5</sup>

In 1955, the States Reorganization Commission recommended forming new states strictly on linguistic lines, a principle which the central government followed while redrawing state boundaries in 1956, 1960, 1966, and 1971 (Govt. of India 1955). After reorganization, new state boundaries consisted of those districts where the major language was the same as the official language of the state. A policy change in all the new states was extending the official language as the medium of instruction in schools and for all state business. Thus, a natural experiment is set up where each district was classified as either majority or minority before the reorganization by historical accident, and reassigned as majority after the reorganization. I expect that growth rates for each of the measures for educational achievement should be higher in historically minority districts after reorganization, as these districts "catch-up" after integrating into co-linguistic states.

The second major empirical challenge is separating the impact of language mismatch from other characteristics of ethnicity and culture that are correlated with language.<sup>6</sup> Figure 1 shows that minority districts are largely North and West of matched districts, which might suggest

<sup>5</sup> For example, Bleakley and Chin (2008) discuss the intergenerational transmission of language skills among immigrants to the United States.

<sup>6</sup> Indeed, a number of studies use linguistic variation as a measure of ethnic cleavages. See Easterly and Levine (1997) and the extensive literature cited in Alesina and La Ferrara (2005).

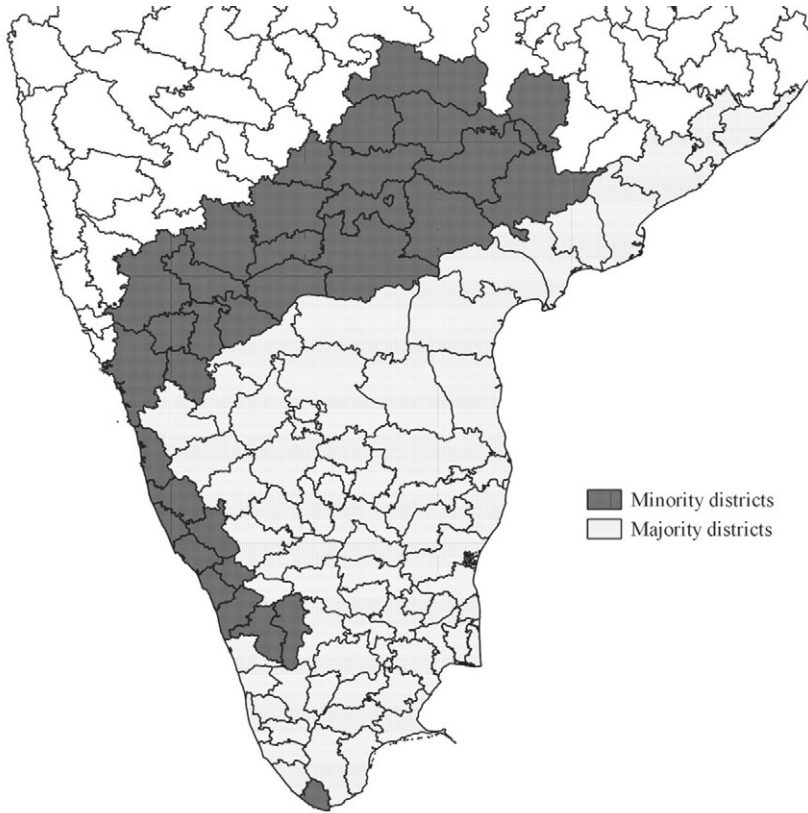


FIGURE 1  
MINORITY AND MAJORITY DISTRICTS

*Source:* Created by author.

other systematic economic, geopolitical, and cultural differences between these types of districts. I account for these differences using three strategies. I examine the historical record, which suggests that the British themselves were uncertain about the social and cultural characteristics of the regions they were conquering. There is no evidence that East India Company officials in the late eighteenth century anticipated mass education efforts that commenced nearly 50 years later. In addition, I try to account for the most obvious sources of differences between minority and majority districts, controlling for a number of geographical and social characteristics, as well as state fixed effects in the regressions. The model explains a large part of the variation in the outcome variables, which assuages concerns associated with potentially omitted variables. Finally, the empirical strategy relies on multiple, logically orthogonal tests that provide consistent evidence that mismatched language is the basis of worse educational outcomes.

Using a district level panel dataset based on the Census of India, I find that shared language potentially lowers communication costs and increases educational achievement rates, and that language mismatch for minority districts is associated with lower rates of educational achievement. The impact is greater on primary and secondary schooling, which is conducted in the vernacular. Specifically, the literacy rate in minority districts is estimated to be 18.0 percent lower than majority districts, and the middle school completion rate 25.2 percent lower. In contrast, although the fraction of college graduates in minority districts is 20.1 percent smaller, this coefficient is not robust across specifications. This finding is not surprising since the medium of instruction in most universities is English. While the precise channel that explains these findings is difficult to pinpoint, the analysis suggests that lower costs of comprehension when instruction is in the same language as the mother tongue is the main factor responsible for the results.

The reorganization of states in 1956 on linguistic lines reversed the impact of historical shortcomings in educational achievement. Minority districts experienced greater growth in educational achievement after reorganization as they caught up with majority districts. The basic test comparing minority and majority districts finds 46.8 percent higher matriculation growth rates in previously minority districts. However, minority districts achieved parity with the majority districts only by 1991, or about two generations later. So while realigning mother tongue with the medium of instruction can alleviate educational shortfalls, this process is long and points to the persistent effects of institutional arrangements.

## HISTORICAL BACKGROUND

### *British Conquest of India*

Commencing in 1757, when the East India Company gained control over the province of Bengal, British colonial rule in India lasted 190 years. From 1757 to 1857, the East India Company extended its control over the rest of India. The Company's administration ended with the mutiny of 1857, after which India was directly ruled as part of the British Empire. In South India, the Company had obtained feudal control over the Coromandel coast from the Nawab of Carnatic in 1640. The geographically contiguous areas around the trading post of Fort St. George (later called Madras and now Chennai) formed the Madras Presidency. In

Western India, seven islands acquired from Portugal as part of a royal dowry in 1661 became the city of Bombay (now Mumbai). This and subsequent territorial acquisitions in Western India, notably the Maratha territories obtained in 1817–1818, were integrated to form the Bombay Presidency.

In addition to direct rule by the British government, a number of regions were indirectly ruled through the agency of native kings and princes. The major princely states in peninsular India were Hyderabad, Mysore, and Travancore. Lakshmi Iyer (2010) shows that the British were selective about the regions annexed for direct rule.<sup>7</sup> Therefore, a key concern for the empirical analysis in subsequent sections is whether selection of regions administered directly was correlated with linguistic characteristics of those areas.

The main events that shaped the boundaries of colonial South India were the Anglo-Mysore Wars (1767 to 1799) between Tipu Sultan of Mysore and the British allied with the Marathas and the Nizam of Hyderabad. Mandar Oak and Anand Swamy (2012) describe the process of alliance formation and the pre-war commitments between the British, the Marathas and the Nizam on territorial division, arguing that the commitments were credible and not subject to post-war negotiations. After Tipu's defeat in the Third Anglo-Mysore war (1792), the Marathas regained Dharwad, the Gulbarga region was returned to the Nizam, and the British added Malabar, Salem, Bellary, and Anantapur to the Madras Presidency (see Figure 2).<sup>8</sup> Neither language nor social factors had a major role to play in determining colonial boundaries. Indeed, the report of Constitutional Reform Commission chaired by Secretary of State for India, Edwin Montagu, and the Viceroy of India, Lord Chelmsford, describing the process of both conquest and organization of the administrative structure of colonial India (Montagu and Chelmsford 1918, p. 28) stated that “[T]he present map of British India was shaped by the military, political or administrative exigencies or conveniences of the

<sup>7</sup> Iyer (2010) uses the annexation policy called “Doctrine of Lapse,” which specified that a territory was annexed if the ruler died without a natural male heir, as a source of exogenous variation to determine the impact of direct versus indirect British rule. This policy was relevant in North India which had a large number of small states, but less germane in the context of the present study which focuses on South India. I test whether my empirical conclusions are sensitive to direct versus indirect rule by the British government in the Online Appendix, and find that they are not.

<sup>8</sup> Tipu's father Hyder Ali had captured the fort of Dharwad in 1778 and therefore the Marathas had historical claims on the region. The British annexed Malabar to prevent Mysore's access to the sea.



FIGURE 2  
LOCATION MAP OF SOUTH INDIA

*Source:* Created by author.

moment, and with small regard to the natural affinities or wishes of the people.”

This sentiment was echoed 12 years later by the Simon Commission (Simon 1930, p. 54) established to review the constitutional structure of British India, “there were in India only a number of administrative areas [which had] grown up almost haphazard as the result of conquest, supersession of former rulers or administrative convenience.” The commission recommended reorganization of states to enable more coherent administration.

Despite the Simon Commission’s recommendations, the colonial government undertook no systematic reorganization of administrative units before the end of British rule in India in 1947. The provincial boundaries of independent India in 1947 reflected geographical continuity in the pattern of British military conquest in the eighteenth and nineteenth



centuries, with little consideration towards the cultural or social characteristics that united or divided the provinces.

*Education Provision in Colonial India*

Education in India before 1757 followed indigenous systems without standardization or significant state patronage and was restricted to the social and economic elites (Acharya 1978). As East India Company officials focused on administration of conquered territories in the eighteenth century, they introduced formal education both to train potential employees for clerical positions as well as to create acceptance of Western traditions and colonial rule (Evans 2002). A rich debate emerged on the language of instruction in government-aided schools between the Orientalists, who favored instruction in English, and the Vernaculars, who advocated instruction in local languages. Inspired by Thomas Babington Macaulay's (1835) famous *Minute on Indian Education*, Governor-General Lord William Bentinck decided to use English as the medium of instruction in mass education. However, instruction exclusively in English proved expensive with very few English language teachers or materials. Consequently, in 1854 Lord Dalhousie accepted Wood's recommendations (outlined in his *Despatch*, which Radhakrishnan (1948) called the "Magna Carta of English Education in India") to adopt vernacular languages for instruction in primary and secondary schools and use English for higher education (Windhausen 1964; Evans 2002).<sup>9</sup>

The British government established universities in Calcutta, Bombay, and Madras starting from 1857 (Radhakrishnan 1948) and the medium of instruction was English (Annamalai 2004). Growing demand for higher and professional education led to the establishment of the University of Allahabad in 1887, as well as 21 other universities in the twentieth century. With the exception of Osmania University in Hyderabad where undergraduate classes were taught in Urdu, the medium of instruction remained English (Radhakrishnan 1948).

<sup>9</sup> The officials of the Education Department at the province level (created specifically in response to Wood's *Despatch*) rather than central or local administrators, had significant influence on the development of schools, and educational outcomes (Nurullah and Naik 1951). Chaudhary (2010) reports that public financing was the backbone of the school system, accounting for nearly half the expenditures during colonial rule. This increased to 60 percent by 1947, with the rest from annual school fees levied on students. With these funds, the Bombay Presidency constructed a large network of public schools whereas Bengal, Bihar, and Orissa relied on private schools that were incorporated into the state system. In regions that experienced indirect colonial rule (native states), education policy was determined by local rulers. The Online Appendix contains more specific details on colonial education policy.

*Teacher Training*

Prior to 1854, formal teacher training was sparse.<sup>10</sup> Systematic teacher training was initiated in the vernacular in Bengal, Madras, and the Northwest provinces only after the *Despatch* of 1854 (Nurullah and Naik 1951). In Bombay, no separate teacher training institutes were established, instead, teacher training was conducted in the regular colleges of the Presidency for both Anglo-vernacular and vernacular schools. By 1881–1882, 9,314 teachers were employed in the Bombay Presidency, and 553 undergoing training. In Bombay, candidates already possessed general education, and were willing to undergo training for three years for better prospects. However, in Madras, candidates lacked education and were often unwilling to undergo long and rigorous training. Nonetheless, by 1881–1882, 500 teachers were trained annually at 32 training schools in Madras.

Following the report of the Indian Education Commission in 1882 (Hunter 1883), the Provincial governments accepted primary responsibility for teacher training. Consequently, teacher training facilities expanded considerably, with 926 training institutions for men (with enrollment of 22,774) and 146 for women operating in British India by 1921. Of these, only five were private, unaided institutions, indicating the extent of government control of teacher training.

After Independence in 1947, the school system expanded considerably, and teacher quality dropped concurrently with greater numbers. The report of the 2nd All India Educational Survey reports that 34.2 percent teachers had not completed matriculation, yet were employed in primary, middle, and secondary schools (NCERT 1967). Of the remaining, 38.4 percent passed matriculation while 15.9 percent had graduate or post-graduate qualifications. A significant fraction had “other” qualifications in music, craft, or physical education.

*Reorganization of Indian States*

Although the Indian National Congress, the main nationalist party, endorsed the principle of the linguistic provinces, India’s Independence accompanied by Partition on religious lines dampened enthusiasm for further division on an ethnic or cultural basis (Guha 2008). The death of an activist demanding a separate state for Telugu speakers, however, led

<sup>10</sup> Nurullah and Naik (1951) record that the Bombay Native Education Society gave three years’ training to 24 primary teachers, with Elphinstone Institution in Bombay also providing some training.

to the formation of Andhra Pradesh from the Telugu speaking districts of Hyderabad and Madras state along with the formation of the States Reorganization Commission (Govt. of India 1955). This commission recommended redrawing state boundaries entirely on linguistic principles, explicitly recognizing the role of shared language in reducing transaction costs stating that “Indian states, if linguistically constituted, will be able to achieve internal cohesiveness because language is a vehicle for communion of thoughts” (p. 35), especially through vernacular education in schools, leading to increasing administrative links within the state because “linguistic homogeneity as an important factor conducive to administrative convenience and efficiency” (p. 46). The modern states of Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu were formed in 1956, Maharashtra and Gujarat in 1960, and Haryana, Himachal Pradesh, and Nagaland in 1965 and 1970, and taught in the vernacular at the school level.

Using language data from the 1951 census, Table 1 shows that the commission followed the linguistic majority rule for every district. Telugu was the most common mother tongue in all districts assigned to Andhra Pradesh, Kannada in each district assigned to Karnataka, Malayalam in every district to Kerala and Tamil in districts to Tamil Nadu. This rules out both the possibility of bargaining or unobserved district characteristics influencing the reorganization exercise. The Online Appendix contains details on changes to medium of instruction in government-operated schools after states’ reorganization.

#### DATA DESCRIPTION

The primary source of data is the decadal Census of India conducted by the Ministry of Home Affairs of the Government of India. I use the 1951, 1961, 1971, 1981, and 1991 waves of the Census. Data at the district level from the last four waves has been compiled into a panel by Reeve Vanneman and Douglas Barnes (2000) and contains data on population characteristics such as literacy, educational achievement, and source of livelihood. Each variable is reported separately for all persons, men, rural residents, and rural male residents in the district. In addition, the 1981 and 1991 Census contain data on the number of schools and colleges at the district level.

The Vanneman and Barnes (2000) dataset is augmented with data on mother tongue, education, religion, and caste composition from the 1951 Census of India. This allows me to measure the baseline rates of educational achievement before 1956, and estimate the difference in

TABLE I  
VERIFICATION OF LANGUAGE-BASED REORGANIZATION

State	Major Languages			State	Major Languages		
Andhra Pradesh	Telugu	Urdu	Rajasthani	Tamil Nadu	Tamil	Telugu	Kannada
Srikakulam	1,895,114	11,966	5	Madras	961,743	234,379	15,866
Visakhapatnam	1,991,689	12,803	58	Chingleput	1,628,012	397,088	6,537
E. Godavari	2,270,664	20,766	249	N. Arcot	2,309,896	340,313	24,316
W. Godavari	1,647,308	22,979	634	S. Arcot	2,537,541	179,893	6,459
Krishna	1,679,530	76,655	8,706	Salem	2,589,194	488,885	189,387
Guntur	2,359,100	164,474	7,512	Coimbatore	2,138,996	653,070	267,196
Nellore	1,641,350	108,934	30	Nilgiri	120,831	26,469	107,492
Chittoor	1,225,248		6,760	Madurai	2,216,038	417,447	150,706
Cuddapah	1,006,250		2,692	Tiruchirapalli	2,587,107	249,696	14,621
Anantapur	1,145,948		19,791	Thanjavur	2,876,292	57,419	5,137
Kurnool	1,294,510		6,401	Ramanathapuram	1,819,079	195,750	44,241
Mahbubnagar	1,200,793	123,187	60,929	Tirunelveli	2,272,394	158,763	8,177
Hyderabad	1,041,773	571,422	32,037	Kanyakumari	667,528	1,463	242
Medak	882,177	123,715	20,243				
Nizamabad	634,981	80,777	24,873	Karnataka	Kannada	Telugu	Urdu
Adilabad	483,094	64,992	32,770	Bangalore	1,049,419	378,319	
Karimnagar	1,341,214	58,284	10,792	Tumkur	908,688	137,540	
Warangal	1,129,675	66,046	99,547	Chitradurga	628,600	138,465	
Khammam	636,394	39,707	51,896	Kolar	610,974	238,604	
Nalgonda	1,121,609	52,971	69,530	Bellary	554,499	96,147	
				Mysore	1,199,303	53,047	
Kerala	Malayalam	Tamil	Konkani	Mandya	667,528	11,657	
Cannore	1,139,740	3,404	8,498	S. Kanara	272,983	2,045	15,847
Kozhikode	1,782,135	5,322	9	Coorg	80,410	3,927	3,818
Palghat	1,269,160	102,827	1	Hassan	614,079	17,904	
Trichur	1,165,772	17,206	3,197	Shimoga	492,808	26,441	
Ernakulam	1,232,939	9,148	13,314	Chikmagalur	297,246	14,949	
Kottayam	1,046,815	112,308	461	Belgaum	1,085,195	14,974	120,155
Alleppey	1,288,937	4,380	5,317	Bijapur	1,177,809	13,428	128,455
Quilon	1,293,174	35,176	156	N. Kanara	283,024	3,313	36,986
Trivandrum	966,787	53,106	32	Dharwar	1,282,604	22,927	178,546
				Gulbarga	784,387	122,992	240,202
				Bidar	200,689	81,251	94,562
				Raichur	717,485	108,472	94,879

Notes: This table verifies that the allocation of districts to states in the 1956 reorganization was strictly on linguistic lines. "N.A." indicates data not available.

Source: Census of India 1951.

outcomes as a result of the language change. The sources of the data are the economic tables and the district census handbooks. While the economic tables report population size variables for all 321 districts in 1951, the district census handbooks report a more detailed set of variables, including educational achievement measures, for 140 districts.<sup>11</sup>

The 1951 and 1961 Census asked respondents about their *mother-tongue*, described in the census forms as the language “first spoken by the individual from the cradle.” In addition, Census forms in all waves defined *literacy* as the “ability to read and write in any language.” Since this definition is not specific to literacy in either the mother-tongue or the state’s official language, mechanical differences in literacy levels as a result of redefining literacy when a district is reassigned from one state to another are ruled out.

I include a number of district-level geographic controls that might impact educational attainment since India is primarily an agricultural economy. The Indian Meteorological Department (IMD) provides monthly rainfall readings at the subdivision level.<sup>12</sup> I calculate the mean and variance of the aggregate rainfall in the months of January and July for each census decade and include these four measures in the dataset. Also added are district level indicators for various soil types, especially the fraction of land which is either wasteland or under forests (Department of Land Resources 2011), along with the latitude, longitude, and elevation of the district headquarters.

The study is restricted to districts situated in the modern states of South India for four reasons. First, the initial wave of reorganization in 1956 only took place in South India. In subsequent waves, states bargained over districts and therefore the natural experiment is not as clean.<sup>13</sup> Second, documented changes in district boundaries between 1971 and 2001 were by and large in North India, whereas district boundaries in South India have remained relatively stable over time (Kumar and Somanathan 2009). Third, South India contained only three large princely states (Travancore, Hyderabad, and Mysore) where education policies are well documented in contrast to North India. Finally, the modern states of Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu exhibit significant diversity in major languages (unlike North India where Hindi is widespread), allowing for cleaner identification of the effects of language on economic outcomes.

<sup>11</sup> The complete dataset is available for download from Open ICPSR. Please see Jain (2017).

<sup>12</sup> Each district is matched to a subdivision.

<sup>13</sup> Guha (2008) recounts considerable political bargaining between the states of Maharashtra and Gujarat over the city of Bombay.

The language used by the majority of residents within a district identifies it as either a “majority” or a “minority” district (see Table 2). In a majority district, the district’s majority language is the same as the province’s official language (also used as the medium of instruction in schools) in the colonial era. In a minority district, the majority language of the district is different from the official medium of school instruction. After reorganization, minority districts were assigned to a state formed on the basis of its language, which was then uniformly used as the medium of instruction in schools within the state. Figure 1 shows how each modern district is classified.

The dataset yields 335 district-year observations (67 districts each observed decadally five times from 1951 to 1991), with only few missing observations for various dependent variables. Table 3 summarizes some of the variables of interest. Among time-invariant characteristics, 22 out of 67 districts are minority districts, 40 percent have some coastal boundary and nearly 63 percent of the districts were directly ruled by the British. The table also reports the mean altitude (1116.3 feet), latitude (13.6°N), and longitude (77.5°E) of districts’ headquarters. Time varying characteristics reported in Table 3 include the fraction of Scheduled Castes (14.3 percent) and Scheduled Tribes (3.0 percent) in the population, the cumulative literacy rate over time (37.0 percent) and the literacy rate by various cohorts.<sup>14</sup> Younger cohorts report higher literacy levels which is consistent with the expansion of basic education over time.<sup>15</sup> Completion rates decline for higher education levels, with 25.8 percent of the population completing primary school, 6 percent matriculating high school, and only 1.2 percent graduating from college. On average, 13.2 percent of a district’s land is wasteland and 5.6 percent under forest cover.

### EMPIRICAL ANALYSIS

The objective here is to estimate the impact of education in the mother tongue (or not) on educational outcomes. The classification of a district as a historically minority language or majority language district depends on the exogenous province formation by the British.<sup>16</sup> The subsequent assignment of districts to states in 1956 is on strict linguistic lines. Therefore,

<sup>14</sup> *Scheduled Castes* (SC) and *Scheduled Tribes* (STs) are official Constitutional terms for classifying various groups of historically disadvantaged indigenous people in India.

<sup>15</sup> Cohort-wise literacy available only for census waves from 1961 to 1991 and not for the 1951 census.

<sup>16</sup> The Online Appendix confirms that the results presented in this section are not sensitive to the influence of direct versus indirect British rule.

TABLE 2  
DISTRICT ALLOCATION

Region	Pre-1956 State	Post-1956 State	Districts (Modern Names)	Pre-Post 1956 Language	Classification
Andhra	Madras	Andhra Pradesh	Anantpur, Chittoor, East Godavari, Guntur, Kadapa, Krishna, Kurnool, Nellore, Prakasam, Srikakulam, Visakhapatnam, Vizianagaram, West Godavari	Telugu to Telugu	Majority
Mysore	Mysore	Karnataka	Chamrajnagar, Mysore, Kolar, Mandya, Tumkur, Bangalore Rural, Bangalore Urban, Bellary, Shivamogga, Chickamagaluru, Chitradurga, Hassan	Kannada to Kannada	Majority
Coorg	Coorg	Karnataka	Kodagu	Tulu to Kannada	Minority
Coastal Karnataka	Madras	Karnataka	Dakshin Kannada, Udupi	Tamil to Kannada	Minority
Gulbarga	Hyderabad	Karnataka	Bidar, Gulbarga, Koppal, Raichur	Urdu to Kannada	Minority
Belgaum	Bombay	Karnataka	Bagalkot, Belgaum, Bijapur, Dharwar, Gadag, Haveri, Uttar Kannada	Marathi to Kannada	Minority
Travancore	Travancore	Kerala	Alappuzha, Ernakulam, Idukki, Kollam, Kottayam, Pathanamthitta, Thiruvananthapuram, Thrissur	Malayalam to Malayalam	Majority
Malabar	Madras	Kerala	Kannur, Kasaragod, Kozhikode, Malappuram, Palakkad, Wayanad	Tamil to Malayalam	Minority
Kanyakumari	Travancore	Tamil Nadu	Kanyakumari	Malayalam to Tamil	Minority

*Notes:* This table shows the allocation of each 1951 district before and after 1956 and the classification as a minority or majority district.

*Source:* Census of India, 1951–1991 and various Acts of Parliament specifying district assignment in states’ reorganization.

TABLE 3  
SUMMARY STATISTICS

Variable	Observations	Mean	Standard Deviation
Minority (%)	335	32.8	
Andhra Pradesh (%)	335	32.8	
Kerala (%)	335	17.9	
Karnataka (%)	335	28.4	
Tamil Nadu (%)	335	20.9	
Coastal (%)	335	40.0	
Direct British rule (%)	335	62.7	
Wasteland (%)	335	13.2	0.089
Forest land (%)	335	5.6	0.055
Altitude (feet)	335	1116.3	1278.02
Latitude (feet)	335	13.6	3.413
Longitude (feet)	335	77.5	2.485
Scheduled caste (%)	294	14.3	0.053
Scheduled tribe (%)	291	3.0	0.042
Rainfall July (Mean)	335	2593.3	2402.63
Rainfall January (Mean)	335	89.2	80.30
Rainfall July (Standard Deviation)	335	831.4	675.45
Rainfall January (Standard Deviation)	335	120.8	99.57
Literates (5+ years, %)	296	37.0	0.182
Literates (5 to 9 years, %)	268	35.7	0.147
Literates (10 to 14 years, %)	268	61.2	0.216
Literates (15 to 19 years, %)	268	57.5	0.212
Literates (20 to 24 years, %)	268	52.8	0.211
Literates (25 to 34 years, %)	268	46.2	0.211
Literates (35+ years, %)	268	34.9	0.173
Primary school completion (%)	268	25.8	0.158
Middle school completion (%)	228	14.4	0.104
Matriculates (%)	295	6.0	0.053
Graduates (%)	228	1.2	0.012

*Notes:* This table shows the summary statistics of the final dataset. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves.

*Sources:* Census of India 1951–1991, Indian Meteorological Department and Ministry of Rural Development (Government of India).



the difference in outcomes between minority and majority districts, as well as the difference in growth rates after reorganization is used to establish the impact of language mismatch on educational outcomes.

If the mother tongue of a district is the same as the medium of instruction used in schools, then more students are likely to enrol in school and achieve basic literacy. I also expect greater rates of completion for subsequent levels of schooling (primary school, middle school, high school, and college) both, because the supply of students from earlier stages increases, and because the demand for schooling increases when students understand instruction better. Later, I estimate a first differences model of the impact of minority status on educational outcomes.

Jess Benhabib and Mark M. Spiegel (2005) develop a model where countries above a threshold level of human capital “catch-up” with technologically advanced countries once they are able to access technology as well. If language is a “social technology,” then districts should experience catch-up growth in educational achievement after reassignment corrects the language mismatch. I test for whether minority districts experience greater growth in educational achievement rates by estimating a difference-in-difference of the impact of minority versus majority districts on educational outcomes before and after reorganization.

### *Do Minority Districts Have Persistently Poorer Outcomes?*

First, I examine whether districts with historical minority language status have persistently poorer educational outcomes. In addition to testing for differences between minority and majority districts, I conduct three robustness checks. The first robustness exercise estimates a first differences model restricting the sample to the set of minority and majority districts that border each other. The second check uses the exogenously determined linguistic distance between the district’s majority language and the language used as the medium of instruction in schools. I expect that increasing linguistic distance within minority districts will lead to relatively poorer economic outcomes. Third, I expect that outcomes will be relatively worse in districts where the fraction of minority language speakers is large compared to districts where the number of speakers of each language is more evenly matched.

### *Test Using Minority Status*

Since data on district-level economic and social characteristics before province formation is not available, the key identifying assumption is

that the initial assignment of districts as minority or majority districts is uncorrelated with outcomes, as discussed earlier. The estimation equation is

$$y_{it} = \beta_0 + \beta_1 \text{minority}_i + \beta_2 \mathbf{Z}_i + \beta_3 \mathbf{X}_{it} + \text{decade}_t + \text{state}_i + \delta_{it} + \varepsilon_{it}, \quad (1)$$

where  $y_{it}$  represents log of outcomes. I expect systematic differences between minority and majority districts.  $\text{minority}_i$  is an indicator variable that is 1 if district  $i$  is a minority district, and 0 otherwise. Thus, the main coefficient of interest is  $\beta_1$ , which represents the marginal impact of being a linguistically mismatched district before the reorganization of states. The specification does not include district fixed effects because  $\text{minority}_i$  is constant over time. Hence, I introduce  $\mathbf{Z}_i$ , which is a vector of time-invariant district characteristics such as a coastal dummy and the altitude of the district headquarters, all of which can potentially impact  $y_{it}$ . Figure 1 shows that minority districts are geographically clustered in the North and West. To capture this aspect of geography, I include the longitude and latitude of the district headquarters. Also included in  $\mathbf{Z}_i$  is a dummy variable that indicates whether a district was under direct British rule or ruled indirectly through the agency of a princely state.<sup>17</sup> This dummy accounts for factors that are unobserved in the data, such as politics, that might have caused educational provision to be different across minority and majority districts.<sup>18</sup>

Equation (1) includes a vector of observed time-varying district characteristics  $\mathbf{X}_{it}$  that consists of the average and standard deviation in January and July rainfall over the decade, the fraction of terrain that is forested or wasteland, logs of the total and rural population of the district in the previous decade, and the fraction of residents who are from historically disadvantaged Scheduled Caste and Scheduled Tribe backgrounds.<sup>19</sup> Political factors that influence available resources for education are controlled by the fraction of time in the decade that the state government's party was the same as the central government party. Decade fixed effects ( $\text{decade}_t$ ) account for observed and unobserved decade characteristics that affect outcomes for all districts. State fixed effects ( $\text{state}_i$ ) control for state characteristics such as attitudes and cultural influences

<sup>17</sup> An alternative is to include a variable that represents the length of British rule. However, since the provincial boundaries of South India were determined well before Wood's *Despatch* in 1854, the length of exposure to British education does not vary.

<sup>18</sup> For example, the rulers of Hyderabad, Mysore, and Travancore might have had strong political incentives to expand education, partly to forestall direct British rule (see Iyer 2010).

<sup>19</sup> When the outcome variable examines only rural residents, this variable is the fraction of *rural* residents who are from SC or ST backgrounds.

that are constant over time. State-time trends ( $\delta_{it}$ ) control for linear time trends in the educational achievement for each state. Finally, unobserved characteristics are clustered by district.

I expect that majority language districts will have higher rates of literacy, middle school completion, and matriculation rates, that is,  $\beta_i < 0$  for these outcomes. While the qualitative impact on graduation rates predicted by the theory are the same, the coefficients for these outcomes might be less robust since English is commonly used as the medium of instruction in higher education, mitigating the impact of historical differences in language use.

Table 4a presents results from estimating equation (1) with total literacy rate, rural literacy rate, middle school completion rate, and rural middle school completion rate as outcome variables. Table 4b continues using the same specification to present estimates when matriculation rates, rural matriculation rates, college graduation rates, and rural graduation rates are the outcome variables. In Table 4a, the coefficients for literacy rates are negative ( $-0.198$  for total literacy and  $-0.225$  for literacy in rural areas,  $p < 0.01$ ). Since the dependent variable is log transformed, taking the exponential of these coefficients implies that the total and rural literacy rates in minority districts are 18.0 percent and 20.1 percent lower than the corresponding rates in majority districts. Similarly, the differences in middle school completion and matriculation rates are, after exponential transformation of the coefficients, even larger (25.2 percent and 28.3 percent lower in minority districts than majority districts). In college graduation rates, the fraction of graduates in minority districts is 28.3 percent smaller than majority districts. The increasing gap as the level of education increases is potentially due to sequentially lower supply of students at the next level. Figures 3a and 3b show log of literacy and college graduation rates separately for every decade from 1951 to 1991. These show visually that the difference between minority and majority districts persists for every year in the decade, although the difference diminishes over time.

Thus, the test of first differences between minority and majority language districts offers evidence that historical language mismatch has a persistent impact on modern educational outcomes. These results can be placed in context of other programs to boost education such as India's school subsidies for girls (National Programme for Education of Girls at the Elementary Level) which increased enrolment by 3 percentage points (Debnath 2013), the Indonesian INPRES school building program which increased schooling by 0.12 to 0.19 years (Duflo 2001), and Mexico's Progreso program which increased enrolment up to 11.1 percent (Schultz 2004).

TABLE 4A  
 RESULT FOR EDUCATION IN MINORITY VERSUS MAJORITY DISTRICTS (A)

	Total Literacy Rate	Rural Literacy Rate	Middle School Completion Rate	Rural Middle School Completion Rate
Minority district	-0.198*** (0.0576)	-0.225*** (0.0554)	-0.291** (0.107)	-0.383*** (0.116)
Total population	0.316*** (0.0743)	0.0511 (0.0655)	0.643*** (0.133)	-0.0190 (0.0914)
Rural population	-0.237*** (0.0842)	-0.0289 (0.0428)	-0.592*** (0.200)	-0.0247 (0.0474)
Fraction scheduled caste	-1.602*** (0.485)	-1.441*** (0.505)	-0.394 (0.724)	0.0255 (0.799)
Fraction scheduled tribe	-0.329 (0.622)	-0.790 (0.650)	-0.390 (0.677)	-1.072 (0.711)
Coastal district	0.250*** (0.0626)	0.328*** (0.0773)	0.180** (0.0772)	0.334*** (0.110)
Direct British rule	0.0108 (0.0379)	0.00572 (0.0513)	-0.0199 (0.0419)	-0.0133 (0.0666)
Fraction forest land	-0.708 (0.676)	0.0118 (0.847)	0.0427 (0.970)	2.270 (1.370)
Wasteland	0.0553 (0.388)	-0.340 (0.514)	-0.203 (0.511)	-1.522* (0.756)
Altitude	0.124*** (0.0311)	0.107*** (0.0376)	0.121** (0.0436)	0.123** (0.0538)
Latitude	-0.0187 (0.0120)	-0.0208* (0.0117)	-0.0330 (0.0195)	-0.0468* (0.0231)
Longitude	-0.0204 (0.0134)	-0.0273 (0.0167)	-0.0300 (0.0179)	-0.0380 (0.0252)
Rainfall July mean	0.0107 (0.0183)	0.000948 (0.0240)	-0.00503 (0.0252)	-0.0338 (0.0471)
Rainfall January mean	-1.700* (0.943)	-2.587** (1.130)	4.804* (2.321)	6.316** (2.395)
Rainfall July standard deviation	0.0366 (0.0344)	0.0612 (0.0394)	0.217*** (0.0626)	0.342*** (0.0921)
Rainfall January standard deviation	1.001** (0.463)	1.310** (0.561)	-1.652 (1.167)	-2.544* (1.284)
Aligned with center	0.127 (0.0793)	0.214** (0.0917)	-0.320 (0.248)	-0.350 (0.264)

TABLE 4A (CONTINUED)  
 RESULT FOR EDUCATION IN MINORITY VERSUS MAJORITY DISTRICTS (A)

	Total Literacy Rate	Rural Literacy Rate	Middle School Completion Rate	Rural Middle School Completion Rate
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.897	0.893	0.939	0.934
Observations	289	289	221	221

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

*Notes:* This table reports the coefficients obtained from estimating equation (1) with various measures of educational achievement as outcome variables. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

*Sources:* Census of India 1951–1991, Indian Meteorological Department and Ministry of Rural Development (Government of India).

### *Test Using Border Districts*

One concern is that these results are driven by omitted variables. To address this, I follow the strategy presented in Abhijit Banerjee and Iyer (2005) and estimate equation (1) on a sample of 37 districts that share a geographical border but differ in classification as minority or majority districts (Figure 4). Since these districts are arguably similar on unobserved characteristics compared to districts far from each other, this strategy helps mitigate omitted variable bias.

Table 5 shows larger and more precisely estimated differences between historically minority and majority districts in education achievement rates (28.5 percent,  $p < 0.01$  for literacy; 32 percent,  $p < 0.01$  for middle school completion; 29.5 percent,  $p < 0.05$  for matriculation; 25.7 percent,  $p > 0.10$  for graduation). Thus, the robustness test using a restricted sample of districts bordering each other suggests that omitted variables do not drive this result, and that the language status of the district directly affects economic outcomes.

### *Test Using Linguistic Distance Measure*

Another potential concern is that language may be systematically correlated with cultural or social differences, so the findings represent cultural rather than linguistic differences between minority and majority

TABLE 4B  
 RESULT FOR EDUCATION IN MINORITY VERSUS MAJORITY DISTRICTS (B)

	Matriculation Rate	Rural Matriculation Rate	College Graduation Rate	Rural College Graduation Rate
Minority district	-0.333** (0.132)	-0.389*** (0.125)	-0.334** (0.155)	-0.358** (0.142)
Total population	0.861*** (0.123)	0.0438 (0.0912)	1.253*** (0.212)	0.144 (0.177)
Rural population	-0.724*** (0.187)	-0.114 (0.0702)	-1.051*** (0.323)	-0.271 (0.197)
Fraction scheduled caste	-1.278 (0.877)	-1.018 (0.804)	-1.925* (1.094)	-1.406 (0.956)
Fraction scheduled tribe	-0.222 (0.802)	-1.456* (0.819)	0.271 (0.920)	-1.180 (0.986)
Coastal district	0.246** (0.0949)	0.396*** (0.122)	0.281** (0.102)	0.457*** (0.125)
Direct British rule	-0.0848* (0.0469)	-0.139** (0.0572)	-0.0954 (0.0706)	-0.132* (0.0675)
Fraction forest land	-0.0696 (0.979)	2.187* (1.069)	0.0326 (1.401)	2.909** (1.210)
Wasteland	0.0844 (0.574)	-0.948 (0.611)	-0.209 (0.862)	-1.392* (0.758)
Altitude	0.154*** (0.0478)	0.116** (0.0545)	0.152** (0.0702)	0.0793 (0.0694)
Latitude	-0.0377 (0.0224)	-0.0450* (0.0238)	-0.0409 (0.0245)	-0.0475** (0.0225)
Longitude	-0.0231 (0.0204)	-0.0314 (0.0241)	-0.0172 (0.0245)	-0.0390 (0.0296)
Rainfall July mean	-0.00427 (0.0304)	-0.0233 (0.0487)	-0.0553* (0.0319)	-0.0753* (0.0436)
Rainfall January mean	2.004 (1.248)	1.276 (1.360)	-0.102 (1.615)	-1.647 (1.660)
Rainfall July standard deviation	0.166** (0.0704)	0.234** (0.102)	0.449*** (0.0879)	0.546*** (0.122)
Rainfall January standard deviation	-0.407 (0.448)	-0.0551 (0.548)	1.196 (0.722)	1.623* (0.888)
Aligned with center	-0.105 (0.130)	-0.0454 (0.156)	0.212 (0.156)	0.474** (0.213)

TABLE 4B (CONTINUED)  
 RESULT FOR EDUCATION IN MINORITY VERSUS MAJORITY DISTRICTS (B)

	Matriculation Rate	Rural Matriculation Rate	College Graduation Rate	Rural College Graduation Rate
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.927	0.940	0.925	0.937
Observations	288	288	221	221

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

*Notes:* This table reports the coefficients obtained from estimating equation (1) with various measures of educational achievement as outcome variables. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

*Sources:* Census of India 1951–1991, Indian Meteorological Department and Ministry of Rural Development (Government of India).

districts. To address this concern, I propose using a measure of linguistic distance that is logically orthogonal to educational outcomes. This measure, developed by M. Paul Lewis, Gary F. Simons, and Charles D. Fennig (2009), is constructed by counting the number of nodes between each pair of languages on the family tree of Indo-European and Dravidian languages.<sup>20</sup> More nodes between a pair of languages imply greater difficulty for a speaker of one language to learn another language and vice versa, and translates into a higher score for linguistic distance. For example, in Figure 5, Tamil and Malayalam are close to each other on the family tree, implying that learning one language is relatively easy for speakers of the other language. This ease is captured by the pairwise linguistic distance of three between the two languages. On the other hand, Telugu speakers find it difficult to understand or learn Kannada and vice versa, which is represented by a pairwise linguistic distance of six.

Table 6 reports the linguistic distance between each pair of Scheduled Indian languages, where the average pairwise distance between the four

<sup>20</sup> Chiswick and Miller (2005) and Beenstock, Chiswick, and Repetto (2001) also develop and use measures of linguistic distance to report the pairwise distance between various languages.

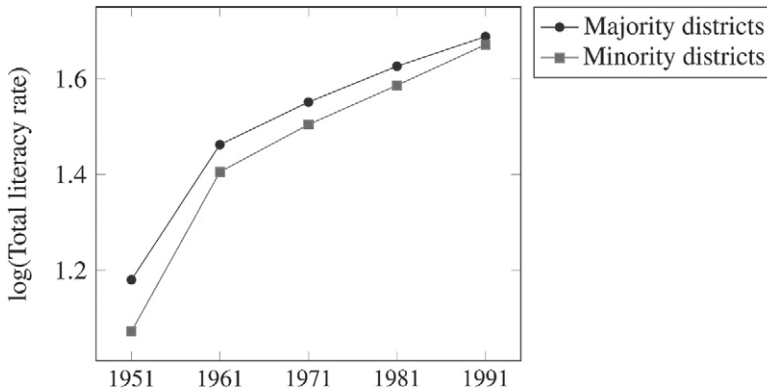


FIGURE 3A  
TOTAL LITERACY BY YEAR

Source: Created by author.

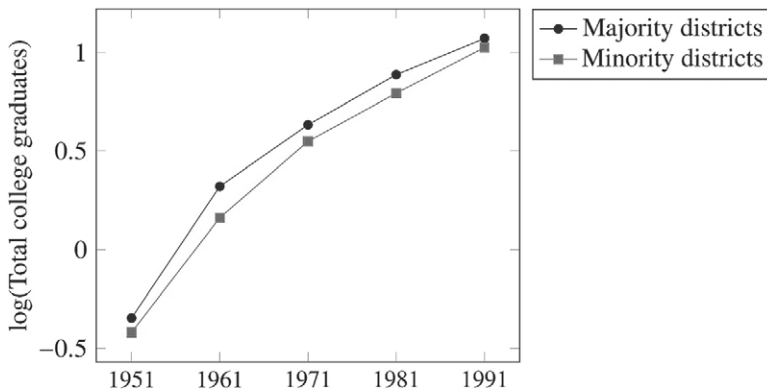


FIGURE 3B  
COLLEGE GRADUATION RATE BY YEAR

Source: Created by author.

relevant South Indian languages is 5.83.<sup>21</sup> Using the data from Table 6, I assign a linguistic distance measure to each minority district based on the pair of languages dominant in the state that the district was assigned to before and after reorganization (see Figure 6). I specify a model where

<sup>21</sup> Shastry (2012) reports that this measure is strongly correlated with two alternative and logically independent measures of linguistic distance. The first, developed by Shastry (2012) measures distance based on shared cognates, distance, and syntax. The second, based on the Comparative Indo-European Database developed by Dyen, Kruskal, and Black (1997) measures distance as the fraction of words from one language that are cognates of words from the second language. This study can use neither the Shastry (2012) measure since it reports the distance only between Hindi and other languages and not for every pairwise combination of languages, nor the Dyen, Kruskal, and Black (1997) measure since it does not include Dravidian languages.



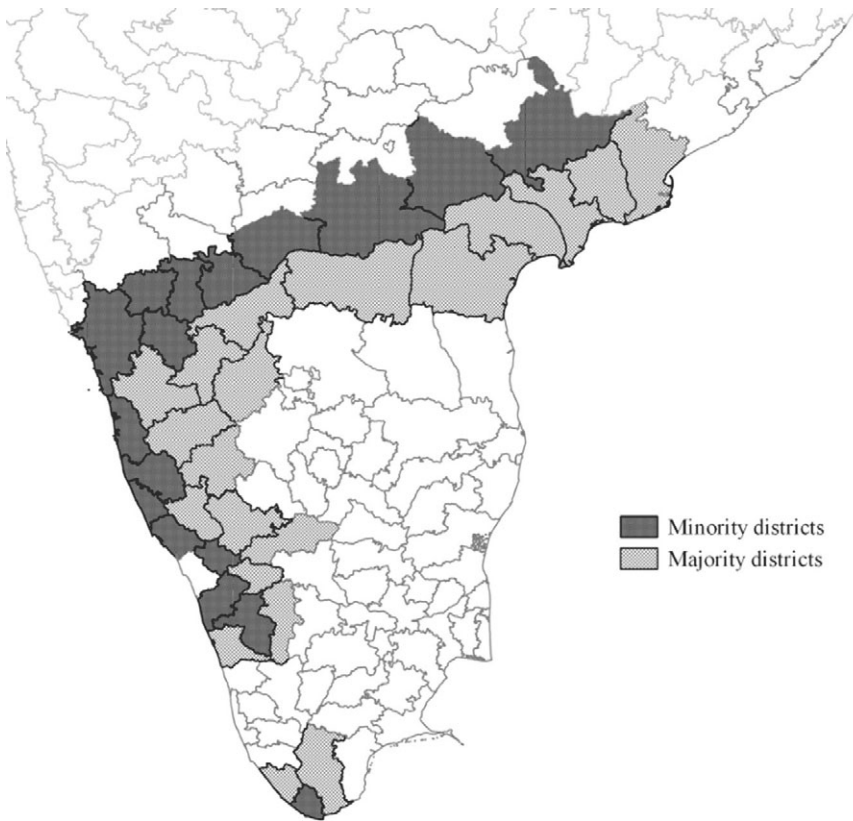


FIGURE 4  
SUB-SAMPLE OF NEIGHBORING MINORITY AND MAJORITY DISTRICTS

Source: Created by author.

the linguistic distance  $L_i$  is interacted with the  $minority_i$  dummy in equation (1). Note that equation (2) does not contain a separate levels term for linguistic distance since the variable is relevant only for minority language districts.

$$y_{it} = \beta_0 + \beta_1 minority_i + \beta_2 minority_i * L_i + \beta_3 \mathbf{Z}_i + \beta_4 \mathbf{X}_{it} \quad (2)$$

$$+ decade_t + state_i + \delta_{it} + \varepsilon_{it}.$$

In this specification,  $\beta_2$  is the marginal impact of a unit increase in linguistic distance on outcomes in a minority district. I expect  $\beta_2$  to be negative as linguistic distance between the mother tongue and the official language makes it more difficult to complete various educational levels.

Table 7 presents ordinary least square (OLS) estimates of  $\beta_2$  using the same set of educational outcomes as the previous section. Note that since

TABLE 5  
RESULT FOR EDUCATION IN MINORITY VERSUS MAJORITY DISTRICTS  
(BORDER DISTRICTS TEST)

	Total Literacy Rate	Middle School Completion	Matriculation Rate	College Graduation Rate
Minority district	-0.336*** (0.0857)	-0.386*** (0.114)	-0.350** (0.154)	-0.297 (0.196)
Control variables	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.924	0.951	0.952	0.962
Observations	126	97	126	97

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

Notes: This table reports OLS coefficients corresponding to  $\beta_1$  from equation (1) which estimates the impact of minority language status on various measures of educational achievement. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

Sources: Census of India 1951–1991, Indian Meteorological Department and Ministry of Rural Development (Government of India).

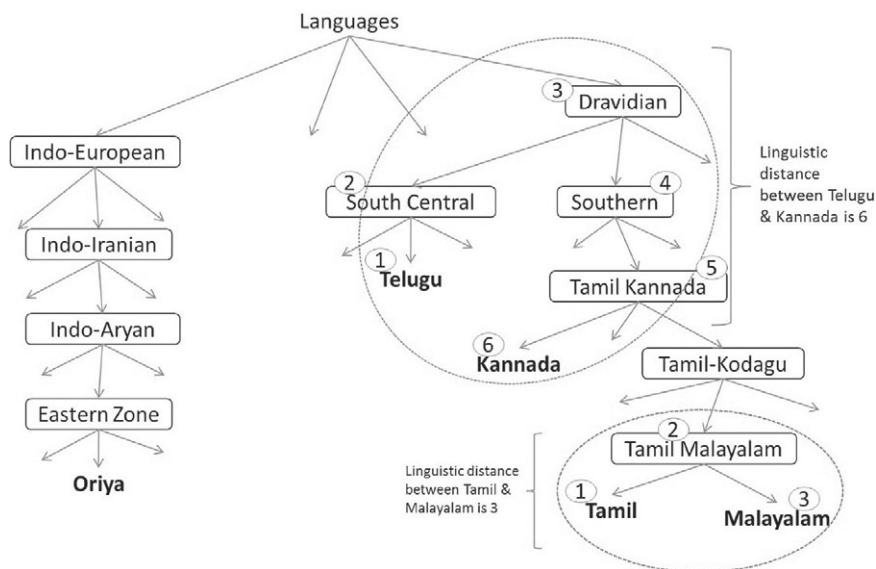


FIGURE 5  
CONSTRUCTION OF LINGUISTIC DISTANCE

Source: Created by author.

TABLE 6  
LINGUISTIC DISTANCE MEASURE

	Hindi	Urdu	Gujarati	Punjabi	Rajasthani	Konkani	Marathi	Assamese	Bengali	Bihari	Oriya	Kashmiri	Kannada	Malayalam	Tamil	Telugu
Hindi	0	1	4	4	5	6	5	6	6	6	6	7	11	13	13	10
Urdu	1	0	4	4	5	6	5	6	6	6	6	7	11	13	13	10
Gujarati	4	4	0	3	4	5	4	5	5	5	5	6	10	12	12	9
Punjabi	4	4	3	0	4	5	4	5	5	5	5	6	10	12	12	9
Rajasthani	5	5	4	4	0	6	5	6	6	6	6	7	11	13	13	10
Konkani	6	6	5	5	6	0	2	5	5	5	5	6	10	12	12	9
Marathi	5	5	4	4	5	2	0	4	4	4	4	5	9	11	11	8
Assamese	6	6	5	5	6	5	4	0	1	3	3	6	10	12	12	9
Bengali	6	6	5	5	6	5	4	1	0	3	3	6	10	12	12	9
Bihari	6	6	5	5	6	5	4	3	3	0	3	6	10	12	12	9
Oriya	6	6	5	5	6	5	4	3	3	3	0	6	10	12	12	9
Kashmiri	7	7	6	6	7	6	5	6	6	6	6	0	11	13	13	10
Kannada	11	11	10	10	11	10	9	10	10	10	10	11	0	5	5	6
Malayalam	13	13	12	12	13	12	11	12	12	12	12	13	5	0	3	8
Tamil	13	13	12	12	13	12	11	12	12	12	12	13	5	3	0	8
Telugu	10	10	9	9	10	9	8	9	9	9	9	10	6	8	8	0

Notes: This table shows the pairwise linguistic distance between major Indian languages.

Source: <http://www.ethnologue.com> and author's calculations.

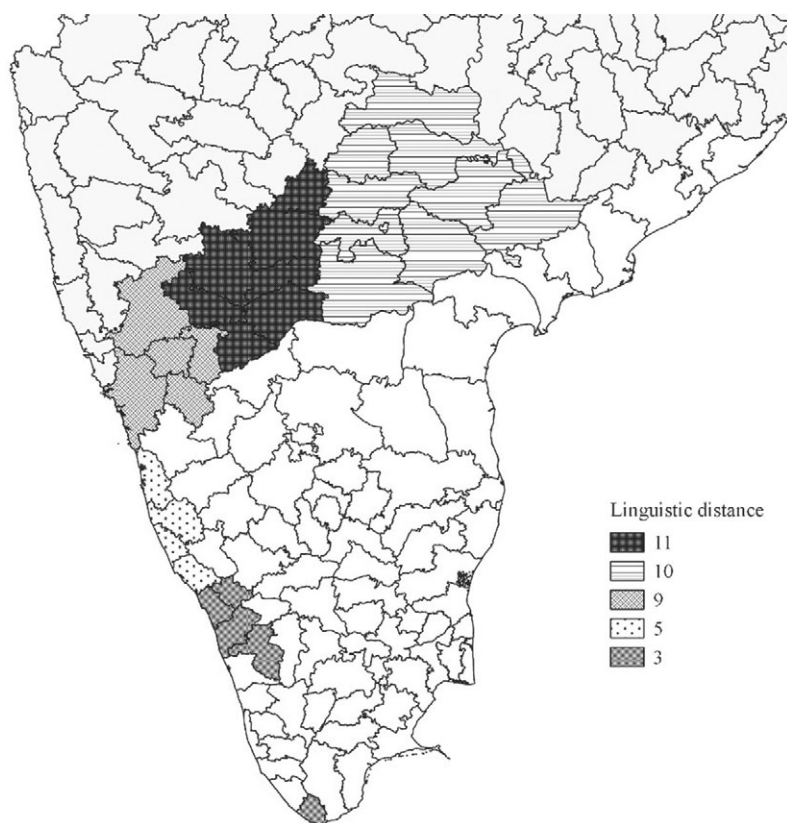


FIGURE 6  
LINGUISTIC DISTANCE BY DISTRICT

Source: Created by author.

linguistic distance is an ordinal measure constructed by counting nodes on the language tree,  $\beta_2$  does not have direct economic interpretation. The estimates suggest that economic outcomes are less with increase in linguistic distance between the district's mother tongue and historical official language. Table 7 reports that a marginal increase in linguistic distance within minority districts decreases literacy and middle school completion rates ( $-0.035$  and  $-0.062$ , respectively,  $p < 0.05$  for both). Matriculation rates are also lower in minority districts but the point estimate is small ( $-0.024$ ) and statistically indistinguishable from the null. Since English is the medium of instruction at the tertiary level, linguistic distance between vernacular languages should not have an impact on graduation rates, as evidenced by the small positive coefficient on college graduation ( $0.026$ ,  $p > 0.10$ ). Conversely, the larger and more precisely estimated impact of linguistic distance on primary and middle school

TABLE 7  
RESULTS FOR EDUCATIONAL OUTCOMES BY LINGUISTIC DISTANCE

	Total Literacy Rate	Middle School Completion	Matriculation Rate	College Graduation Rate
Minority district	0.0850 (0.136)	0.197 (0.283)	-0.136 (0.372)	-0.525 (0.475)
Minority district * Linguistic distance	-0.0350** (0.0157)	-0.0619** (0.0295)	-0.0238 (0.0366)	0.0257 (0.0470)
Control variables	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.900	0.941	0.927	0.925
Observations	285	218	284	218

\* = Significant at the 0.10 level.  
 \*\* = Significant at the 0.05 level.  
 \*\*\* = Significant at the 0.01 level.

Notes: This table reports OLS coefficients corresponding to  $\beta_2$  from equation (2) which estimates the impact of linguistic distance between the district’s mother tongue and the language used in schools among minority districts on various measures of educational achievement. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

Sources: Census of India 1951–1991, Indian Meteorological Department, Ministry of Rural Development (Government of India).

education compared to higher education suggests that language effects manifest themselves early in the education process when mother tongue teaching is more important.

*Test Using Minority Fraction Measure*

The binary *minority* variable as defined and used in previous sections does not capture the intra-district mix of languages. I expect that minority districts that are more polarized (i.e., fewer speakers of the official language) will experience poorer educational outcomes than otherwise. Thus, I construct a continuous measure of minority status, *MinorityFraction*, that helps differentiate polarized districts where the state’s minority language is spoken by a large fraction of residents from those districts where the number of minority and majority language speakers are more evenly matched.

$$MinorityFraction = \frac{MotherTongue - OfficialLang}{TotalPopulation} \tag{3}$$

In this definition, *OfficialLang* is the number of speakers of the official language of the pre-reorganization state where the district is located. *MotherTongue* is the number of speakers of the most popular language spoken in the district other than the state's official language. Hence, for minority districts,  $MotherTongue > OfficialLang$  and  $MinorityFraction > 0$ . Additionally, a large positive value for *MinorityFraction* indicates that a large fraction of the population speaks the minority language compared to the state language whereas a small positive value implies that the two languages are spoken by relatively same number of district residents.

The 1951 Census reports the top three languages spoken in each district from which I calculate *MinorityFraction* for each district and use in equation (4).

$$y_{it} = \beta_0 + \beta_1 \text{MinorityFraction}_i + \beta_2 \mathbf{Z}_i + \beta_3 \mathbf{X}_{it} + \text{decade}_t \quad (4) \\ + \text{state}_i + \delta_{it} + \varepsilon_{it}.$$

In equation (4),  $\beta_1$  indicates the marginal impact of increasing the share of minority language speakers on outcome variable  $y_{it}$ . Table 8 shows that districts with large minority language populations suffer from greater shortfalls in educational attainment. A unit increase in *MinorityFraction* is associated with 7.6 percent decrease in literacy rates, 12.6 percent lower middle school completion rates, and 12.9 percent lower matriculation rates ( $p < 0.05$  for all). The decrease in college graduation is 13.4 percent ( $p < 0.10$ ). Thus, districts that were more polarized in 1951 experienced significantly poorer educational outcomes in the post-Independence period, suggesting that language mis-alignment had a persistent impact on educational outcomes.

### *Plausible Channels*

In this section, I examine various mechanisms through which language might affect educational achievement. Although it is difficult to pinpoint the precise channel with the available data, I rule out the effects on occupational choice, provision of public schools, and inter-district migration on educational outcomes, suggesting the primacy of factors within the school, such as the medium of education, in determining educational achievement.

A mechanism through which language might influence educational attainment is through access to new occupations and business networks. If communication intensive occupations (such as those in the secondary

TABLE 8  
RESULTS FOR EDUCATIONAL OUTCOMES USING *MinorityFraction*

	Literacy Rate	Middle School Completion Rate	Matriculation Rate	College Graduation Rate
<i>MinorityFraction</i>	-0.0792*** (0.0279)	-0.135** (0.0533)	-0.138** (0.0639)	-0.144* (0.0728)
Control variables	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.892	0.940	0.931	0.936
Observations	262	200	261	200

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

*Notes:* This table reports OLS coefficients corresponding to  $\beta_1$  in equation (4) which estimates the effect of increasing fraction of speakers whose mother tongue is different than the official language on educational achievement. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

*Sources:* Census of India 1951–1991, Indian Meteorological Department, Ministry of Rural Development (Government of India).

and tertiary sectors) require knowing the majority language of the state or province, then returns to education will be greater for individuals who live in majority districts. However, Online Appendix Table 6 shows no significant differences in the occupational structure between the minority and majority districts. This suggests that districts’ historical language status did not persistently affect occupational structure. Hence, access to communication-intensive occupations or business networks is unlikely to be a major channel to motivate greater investments in education.

The second mechanism is the impact of public investments, especially school building, on educational attainment. State governments, which control primary and secondary education in India, might invest in minority language areas with relatively lower educational achievement to spur improvements in schooling completion rates.<sup>22</sup> Alternatively, if representatives from majority language districts are more likely to form the government, they may reward constituents with more educational infrastructure. Online Appendix Table 7 reports differences in the presence of

<sup>22</sup> Even in the colonial era, Chaudhary (2009) reports that school placement was influenced by caste and religious polarization in villages.

various school types in minority versus majority districts. Although all coefficients are negative, suggesting fewer schools in minority language districts, these cannot be statistically differentiated from the null. Thus, while I cannot conclusively rule out the role of public investments to explain differences in educational achievement, the empirical evidence for this channel is also weak.

Finally, I examine the impact of inter-district migration. Language policies might impact educational outcomes if individuals in minority districts who are more motivated to study migrate to majority districts, enrol in schools and complete increasing levels of education. To estimate the impact of this channel, I use the same specifications presented earlier. Along with total migration, I also include the male migration rate as the dependent variable since men are more likely to migrate in search of work than entire families. Online Appendix Table 8 shows that the coefficients for the minority variable are all small and statistically insignificant, which is consistent with other empirical studies that do not find significant inter-district migration in India over this period (Munshi and Rosenzweig 2009). This suggests that districts' historical linguistic status did not have much impact on migration and inter-district migration is an unlikely channel through which language influences educational outcomes.

#### *Catch up after Reorganization*

The panel structure of the dataset and the timing of the 1956 reorganization can be used to test for the catch-up hypothesis among minority districts after reassignment.<sup>23</sup> Only those districts for which 1951 census data is available are included. The estimation equation is

$$y_{it} = \beta_0 + \beta_1 \text{minority}_i + \beta_2 \text{Post}_t + \beta_3 \text{minority}_i * \text{Post}_t + \beta_4 \mathbf{X}_{it} \quad (5)$$

$$+ \beta_5 \mathbf{Z}_i + \text{decade}_t + \text{state}_i + \delta_{it} + \varepsilon_{it},$$

where  $y_{it}$  represents an educational outcome as measured in each census wave from 1951 to 1991.  $\text{Post}_t$  is an indicator variable that is 0 if the year is 1951 and 1 otherwise. Hence, the coefficient  $\beta_3$  represents the

<sup>23</sup> The Online Appendix describes an "internal replication" exercise that uses pooled individual level data from the National Family Health Survey's cross-sectional rounds from 1992 and 1998. With birth year identifiers, I examine the difference in educational achievement by cohort across minority and majority districts. The findings show an increase in educational achievement among cohorts born after 1945, which is consistent with the main findings.



TABLE 9  
 RESULT FOR EDUCATION BEFORE AND AFTER REORGANIZATION  
 IN MINORITY VERSUS MAJORITY DISTRICTS

	Literacy Rate	Middle School Completion Rate	Matriculation Rate	College Graduation Rate
Minority district	-0.461** (0.171)	-0.812*** (0.279)	-0.917*** (0.304)	-0.712* (0.346)
Minority * Post	0.285* (0.148)	0.571** (0.226)	0.632** (0.237)	0.414 (0.270)
Control variables	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.900	0.943	0.932	0.926
Observations	289	221	288	221

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

*Notes:* This table reports the impact of language on educational achievement before and after the 1956 reorganization of Indian states. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

*Sources:* Census of India 1951–1991, Indian Meteorological Department, Ministry of Rural Development (Government of India).

marginal impact of the 1956 reassignment on minority language districts. The key identifying assumption is that in the absence of the reorganization, there would be no systematic differences in the trend of  $y_{it}$  between minority and majority districts. I expect greater increase in enrolment and completion of formal education among minority language districts after reorganization.

The results in Table 9 show that reassignment had a large and significant impact on literacy, middle school completion, and matriculation rates. Minority districts experienced higher growth rates than majority districts. Specifically, the growth rate was 36.9 percent higher for literacy, 55.6 percent greater for middle school completion, 60 percent for matriculation, and 50.9 percent for college graduation. Figure 3a shows that the difference in literacy between minority and majority districts is large before reorganization and diminishing after 1956, suggesting that alignment of a district’s mother tongue with the state’s official language is

TABLE 10  
 RESULT FOR EDUCATION OUTCOMES BEFORE AND AFTER REORGANIZATION  
 BY LINGUISTIC DISTANCE

	Literacy Rate	Middle School Completion Rate	Matriculation Rate	College Graduation Rate
Minority * Linguistic distance	-0.0612** (0.0244)	-0.117*** (0.0339)	-0.0838* (0.0413)	-0.0184 (0.0523)
Minority * Linguistic distance * Post	0.0285* (0.0139)	0.0600** (0.0223)	0.0651*** (0.0228)	0.0483* (0.0262)
Control variables	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-decade trend	Yes	Yes	Yes	Yes
R-squared	0.903	0.946	0.932	0.927
Observations	285	218	284	218

\* = Significant at the 0.10 level.

\*\* = Significant at the 0.05 level.

\*\*\* = Significant at the 0.01 level.

*Notes:* This table reports the impact of language on educational achievement before and after the 1956 reorganization of Indian states. Each observation is a district-year, and the sample is pooled over the 1951, 1961, 1971, 1981, and 1991 census waves. Robust standard errors in parentheses are clustered at the district level.

*Sources:* Census of India 1951–1991, Indian Meteorological Department, Ministry of Rural Development (Government of India).

positively associated with enrolment in schooling. However, Figure 3b does not show much change in college graduation rates between 1951 and 1961, and minority and majority districts converge on this measure only in later decades. This is not surprising, since changing the medium of instruction in 1956 is unlikely to change the supply of potential college students by 1961. Instead, increases in schooling in the 1950s and 1960s as a result of the reorganization of states will impact university-level education by 1970 at the earliest.

In Table 10, I report coefficients associated with a specification where  $minority_i$  is interacted with  $Post_i$  and  $L_i$ . Consistent with previous results, all coefficients associated with literacy, middle school completion, matriculation, and graduation rates suggest that districts where the mother tongue was linguistically distant from the official language before 1956 benefitted more from reorganization, and higher education improved as a result of instruction in the mother tongue. Finally, the coefficient

associated with *MinorityFraction* interacted with *Post<sub>t</sub>*, is consistent with findings reported in Tables 9 and 10.<sup>24</sup>

## DISCUSSION

The historian Ramachandra Guha argued that the reorganization of Indian states was a transformative event in the life of a young republic (Guha 2008). It recognized and accommodated the development of a wide array of languages and associated cultural traditions while maintaining a federal and democratic polity. In this article, I not only demonstrate that colonial-era provinces which mixed together linguistic minority and majority areas led to differences in long-term educational outcomes, but also that language-based reorganization of state boundaries might have remedied these differences.

The magnitude of educational differences between districts that do and do not speak the official language is large, indicating that historical factors can dominate government policies or private remedies aimed at alleviating educational shortfalls. Nonetheless, catch up by minority districts after reorganization suggests that historical mismatches can be remedied through large scale political changes.

These results should be read with a number of caveats. First, just because I analyze the effects of language mismatch in this setting does not imply that these findings can be readily generalized for all situations. While I argue based on institutional characteristics that the results could be representative of other contexts, researchers should be cautious extending these findings to other languages, nations, or educational levels or systems. For instance, other educational systems might feature much more intense multilingual instruction that mitigates the impact of language mismatch. So the strength of language effects that I find might represent upper bounds compared to other educational systems.

Second, while I examine educational achievement in school completion rates, due to data limitations I do not estimate the impact of language mismatch on a wide range of additional outcome variables, including choice of subjects and test scores that the education literature deems important. Similarly, variables such as earnings are unexamined due to data limitations.

Third, the article does not outline a comprehensive model of the influence of language effects on educational outcomes. Absent a comprehensive

<sup>24</sup> Results are available from the author upon request.

mechanism or the ability to conduct experiments, I cannot evaluate the impact of specific policies (such as multiple language instruction or sorting students by mother tongue within a district) to improve student outcomes. Finally, this article does not account for bilingualism, multilingualism, and language shift, especially with increasing use of English in the last century.

Nonetheless, this article has implications on new state formation in India. After 1971, a number of Union Territories (areas administered by the central government) converted to formal statehood. More significantly, three new states—Chhattisgarh, Jharkhand, and Uttaranchal—were carved out in 2000 from larger states on the basis of distinct culture of these regions; a fourth state Telangana separated from Andhra Pradesh in 2014. The results presented in this article indicate that new states formed on the basis of shared language might experience better educational outcomes.

#### REFERENCES

- Acharya, Poromesh. "Indigenous Vernacular Education in Pre-British Era: Traditions and Problems." *Economic & Political Weekly* 13, no. 48 (1978): 1981–88.
- Alesina, Alberto, and Eliana La Ferrara. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43, no. 3 (2005): 762–800.
- Angrist, Joshua D., and Victor Lavy. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics* 15, no. 1 (1997): 48–76.
- Annamalai, E. "Medium of Power: The Question of English in Education in India." In *Medium of Instruction Policies: Which Agenda? Whose Agenda?*, edited by Amy B. M. Tsui and James W. Tollefson, 177–94. Mahwah, NJ: Lawrence Erlbaum Associates, 2004.
- Azam, Mehtabul, Aimee Chin, and Nishith Prakash. "The Returns to English-Language Skills in India." *Economic Development and Cultural Change* 61, no. 2 (2013): 335–67.
- 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): 1190–213.
- Banerjee, Abhijit, Lakshmi Iyer, and Rohini Somanathan. "History, Social Divisions, and Public Goods in Rural India." *Journal of the European Economic Association* 3, no. 2/3 (2005): 639–47.
- Beenstock, Michael, Barry R. Chiswick, and Gaston L. Repetto. "The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel." *International Migration* 39, no. 3 (2001): 33–60.
- Benhabib, Jess, and Mark M. Spiegel. "Human Capital and Technology Diffusion." *Handbook of Economic Growth* 1 (2005): 935–66.
- Bleakley, Hoyt, and Aimee Chin. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86, no. 2 (2004): 267–298.

- . “What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital Among Immigrants.” *Journal of Human Resources* 43, no. 2 (2008): 267–98.
- Chakraborty, Tanika, and Shilpi Kapur Bakshi. “English Language Premium: Evidence from a Policy Experiment in India.” *Economics of Education Review* 50 (2016): 1–16.
- Chaudhary, Latika. “Determinants of Primary Schooling in British India.” *Journal of Economic History* 69, no. 1 (2009): 269–302.
- . “Land Revenues, Schools and Literacy: A Historic Examination of Public and Private Funding of Education.” *Indian Economic & Social History Review* 47, no. 2 (2010): 179–204.
- Chiswick, Barry R., and Paul W. Miller. “Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages.” *Journal of Multilingual and Multicultural Development* 26, no. 1 (2005): 1–11.
- Clingingsmith, David. “Industrialization and Bilingualism in India.” *Journal of Human Resources* 49, no. 1 (2014): 73–109.
- Debnath, Sisir. “School Subsidy for Girls and Gender Gap in Enrolment.” Mimeo. Indian School of Business, 2013.
- Department of Land Resources. *Wasteland Atlas of India*. New Delhi: Ministry of Rural Development, Govt. of India, 2011.
- 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.
- Dyen, Isidore, Joseph B. Kruskal, and Paul Black. “Comparative Indo-European Database Collected by Isidore Dyen.” File IE-Rate 1, 1997.
- Easterly, William, and Ross Levine. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics* 112, no. 4 (1997): 1203–50.
- Evans, Stephen. “Macaulay’s Minute Revisited: Colonial Language Policy in Nineteenth-Century India.” *Journal of Multilingual and Multicultural Development* 23, no. 4 (2002): 260–81.
- Govt. of India. *Report of the States Reorganization Commission*. New Delhi: Manager of Publications, 1995.
- General, Registrar. “Census of India, 2001.” Available at [www.censusindia.gov.in](http://www.censusindia.gov.in).
- Guha, Ramachandra. *India after Gandhi: The History of the World’s Largest Democracy*. New York: Picador, 2008.
- Hunter, William. *Report of the Indian Education Commission 1881–82*. Calcutta: Manager of Publications, Govt. of India, 1883.
- Iyer, Lakshmi. “Direct Versus Indirect Colonial Rule in India: Long-Term Consequences.” *Review of Economics and Statistics* 92, no. 2 (2010): 693–713.
- Jain, Tarun. “Common tongue: The impact of language on educational outcomes [Data set]”. Ann Arbor, MI: Inter-university Consortium of Political and Social Research, 2017. Available at <http://doi.org/10.3886/E100356V1>.
- Jain, Tarun, Pushkar Maitra, and Subha Mani. “Barriers to Skill Acquisition: Evidence from English Training in India.” IZA Discussion Paper No. 10199, Bonn, Germany, 2016.
- Kumar, Hemanshu, and Rohini Somanathan. “Mapping Indian Districts across Census Years, 1971–2001.” *Economic & Political Weekly* 44, no. 41–42 (2009): 69–73.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. “Ethnologue: Languages of the world, Sixteenth edition.” Dallas, TX: SIL International, 2009. Available online at <http://www.ethnologue.com>.

- Macaulay, Thomas B. "Minute by the Hon'ble T. B. Macaulay, Dated the 2nd February 1835." In *Selections from Educational Records, Part I (1781–1839)*, edited by H. Sharp. Calcutta: Superintendent, Government Printing, 1835.
- Montagu, Edwin S., and Frederic J. Chelmsford. *Report on Indian Constitutional Reforms*. Calcutta: Superintendent Government Printing, 1918.
- Munshi, Kaivan, and Mark R. Rosenzweig. "Why Is Mobility in India so Low? Social Insurance, Inequality, and Growth." NBER Working Paper No. 14850, Cambridge, MA, 2009.
- NCERT. "Second All-India Educational Survey." New Delhi: Publication Unit, National Council for Educational Research and Training, Govt. of India, 1967.
- Nurullah, Syed, and Jayant P. Naik. "History of Education in India: During the British Period." Bombay: Macmillan and Company, 1951.
- Oak, Mandar, and Anand V. Swamy. "Myopia or Strategic Behavior? Indian Regimes and the East India Company in Late Eighteenth Century India." *Explorations in Economic History* 49, no. 3 (2012): 352–66.
- Radhakrishnan, Sarvepalli. *Report of the University Education Commission*. New Delhi: Government of India Press, 1948.
- Schultz, T. Paul. "School Subsidies for the Poor: Evaluating the Mexican Progresá Poverty Program." *Journal of Development Economics* 74, no. 1 (2004): 199–250.
- Shastri, G. Kartini. "Human Capital Response to Globalization: Education and Information Technology in India." *Journal of Human Resources* 47, no. 2 (2012): 287–330.
- Simon, John A. *Report of the Indian Statutory Commission*. New Delhi: Swati Publications, 1930.
- Vanneman, Reeve, and Douglas Barnes. "Indian District Data, 1961–1991: Machine-Readable Data File and Codebook." Center on Population, Gender, and Social Inequality. University of Maryland, 2000.
- Windhausen, John D. "The Vernaculars, 1835–1839: A Third Medium for Indian Education." *Sociology of Education* 37, no. 3 (1964): 254–70.