Department of Economics

DISCUSSION PAPER SERIES IN ECONOMICS

DP No. 35

Language and Learning in Ethically Mixed Communities: A Study of School Children in an Indian Village

September 2020

Hemanshu Kumar
Rohini Somanathan
Mahima Vasishth

https://www.ashoka.edu.in/ecodp

Language and learning in ethnically mixed communities: A study of school children in an Indian village

Hemanshu Kumar* Rohini Somanathan† Mahima Vasishth‡ §

September 3, 2020

Abstract

We study the role of language on the learning of primary school children in a multi-ethnic village in India. Half the families have a mother tongue closely related to the medium of school instruction while the rest speak one of two tribal languages. Children solve mazes and take tests in reading and mathematics. Performance in language dependent tasks relative to mazes is lower for tribal children in segregated hamlets but not for those in ethnically mixed ones. We conclude that interventions that encourage language acquisition may help linguistic minorities more than mother-tongue instruction, which is the focus of current policy.

1 Introduction

Language has historically served as an important instrument for political consolidation. Colonizers governed in their home languages. Newly independent nations often reversed these policies and introduced indigenous languages into public schools and state administration to build national identity and make education widely accessible. As part of the expanding research on the economics of language, several studies have used changes in language policies as natural experiments to estimate the impact of the medium of school instruction on learning and labor market outcomes.

The findings from this literature are mixed. Indigenous language instruction sometimes improves literacy and cognitive skills in primary school, but often does not. Even a given policy reform or experiment may result in improvements along some dimensions and stagnancy or deterioration along others. For example, Ramachandran (2017) finds that the introduction of local language instruction led to sizeable literacy gains in Ethiopia in the mid-nineties, while Chicoine (2019) attributes these gains to increases in schooling that accompanied the elimination of school fees around the same time. Using cohort analysis and geographical and temporal variation in the implementation of these two policies, he finds that mother-tongue

*Kumar: Department of Economics, Ashoka University, Rajiv Gandhi Education City, Rai, Sonepat, Haryana 131029, India; hemanshu.kumar@ashoka.edu.in.

†Somanathan: Department of Economics, Delhi School of Economics, University of Delhi, Delhi 110007, India; rohini@econdse.org.

‡Vasishth: Department of Economics, University of California, Irvine, USA; vasishth@uci.edu.

§We are grateful to the Azim Premji University Research Funding Programme for a grant under project E224. We would like to thank Tista Bagchi for useful discussions and Pramod Kumar for invaluable assistance in the field.

instruction reduced schooling and had no impact on literacy. Laitin, Ramachandran, and Walter (2019) estimate achievement gains of local language instruction for students in grades 1, 3 and 5 in northwestern Cameroon and find gains for early grades in mathematics that dissipate by grade 5. Chakraborty and Bakshi (2016) find that the abolition of English teaching in public schools in an Indian state limited access to high wage occupations, while Coyne (2015) argues that the use of colonial languages in African countries increased inequality by impeding the progress of marginalized communites through school. Ginsburgh and Weber (2020) discuss several other historical examples from across the world.

It is difficult to systematically evaluate the policy of local language instruction for several reasons. The policies and their regional contexts are very varied. Local languages are sometimes supplemented by colonial languages and in other cases replace them, teachers may not be familiar with local scripts and children often have to transition back to another language, such as English or French in secondary school.[1] Second, in linguistically diverse societies, such as Nigeria and India, of the many hundreds of languages spoken, only a small set are officially recognized and used as a medium of instruction. A language policy that privileges some children may place others in the same school at a disadvantage. The demographic composition of the school would determine average changes in achievement.[2] A Global Education Monitoring Report by UNESCO uses data from across to world to compare learning levels for students who speak the language of instruction at home and those who do not. It finds that test scores are adversely affected when home and school languages differ and that there is no single local language within schools (UNESCO 2016). We cannot ensure that all children are taught in a language they understand by simply switching out of colonial languages.

Our paper focuses on the role of the medium of instruction in settings whether families speak multiple distinct languages within a narrow geography. We study the performance of all primary school aged children in an ethnically mixed village in India. Most of the roughly 300 children attend one of two public schools in the village. About half the households in the village speak an Indo-Aryan language closely related to Hindi, the medium of school instruction, while the other half are equally split between two tribal languages. We administer tests in reading and mathematics, in which we expect the language for instruction to influence learning. We also have the children solve a set of four mazes. We ask whether the performance on the language-dependent tasks relative to the mazes varies with the linguistic distance of a child's mother-tongue from the medium of instruction. Our main finding is that linguistic distance affects performance on these tasks relative to that on mazes for tribal children who live in hamlets with only other families of their tribe, but not for those children living near families speaking languages closely related to Hindi.

We view the paper as making two main contributions to the literature on language and learning. First, it provides an alternative measure of performance which may better capture the effect of linguistic distance on learning in traditional academic tasks. Mazes have been used by other studies to measure performance among population sub-groups (Gneezy, Niederle, and Rustichini 2003; Hoff and Pandey 2006). We use them as a non-language based measure to benchmark student ability. Cross-sectional data on performance confounds the effects of language with other factors, such as the quality of local schools and unobservable resources available to the child within the family and neighborhood. Measuring a student's perfor-

1. See Bamgbose (1984) for a discussion of these issues and of alternative programs in Nigeria.
2. See Odugu (2011), Chapter 6 for language policies in Nigeria and Benedikter (2011) for India.

mance on language-dependent tasks relative to her performance on mazes helps control for unobservables at the individual level.

Second, our results suggest that the disadvantages of being instructed in a foreign tongue can be overcome through policies that reduce the isolation of linguistic minorities. Mother-tongue instruction is simply not feasible if a subpopulation is linguistically diverse. Policies that make second language acquisition effective may be the most realistic option for equalizing opportunity across children from different ethnic groups.

The language question is particularly relevant in India today since the New Education Policy just released by the Government of India proposes using the local language as a medium of instruction in all public and private primary schools. The policy document uses the terms local language and mother-tongue interchangeably (Government of India 2020).[3] The two may in fact be very different for many disadvantaged populations in India. Equalizing access to education requires that we pay attention to linguistic diversity within schools, as well as to residential segregation, which affects the ease with which minority populations acquire facility in local languages.

The next section describes our experimental setting and places the languages spoken there in the broader context of India's language trees. Section 3 has details on the data and results which attempt to separately identify the role of linguistic distance of a mother-tongue from residential isolation of a group. We end with some remarks on the implications of our study for language policy.

2 The setting

India is one of the most linguistically diverse countries of the world. The Ethnologue counts 447 living languages in the country, classified into six families: Indo-European, Dravidian, Austro-Asiatic, Afro-Asiatic, Sino-Tibetan, and Kra-Dai.[4] Over three-quarters of the Indian population reports a mother tongue in the Indo-Aryan sub-family of the Indo-European group and most others speak Dravidian languages.[5] Tribal communities across the country often speak languages that are very different from the official language of their state. Officially recognized tribes are listed in schedules of the Indian constitution and are therefore called Scheduled Tribes (STs).[6] These tribes, which form 8% of the Indian population, lag far behind majority populations with school completion rates less than half of the population average.[7] The role of language in determining this gap remains an open question.

3. Page 12 of the document states:

> Wherever possible, the medium of instruction until at least Grade 5, but preferably till Grade 8 and beyond, will be the home language/mother tongue/local language/regional language. Thereafter, the home/local language shall continue to be taught as a language wherever possible. This will be followed by both public and private schools. High-quality textbooks, including in science, will be made available in home languages/mother tongue.

.

4. See Eberhard, Simons, and Fennig (2020).

5. The majority languages in the four south Indian states are all Dravidian.

6. See Government of India (2018) for census returns on mother tongue.

7. As of 2011, only 14% of the scheduled tribe population in the relevant age group completed secondary school while the population average was 31%. The primary school completion rates are 41% and 58% respectively (Government of India 2011a, Tables C-08 and C-08 ST). See Kumar and Somanathan (2016) for trends

Our experiment was designed to understand how the relationship between linguistic distance and learning is modified by physical proximity to households speaking the majority language. The setting is a village in Katihar district in the north Indian state of Bihar. Bihar is among the most linguistically diverse parts of the country, with 114 distinct mother tongues. Katihar is its most diverse district.[8] The village we study lies 15 kilometres from the district capital and has about 450 households. Of these, 45% are Scheduled Tribes.[9] Our village has two major tribes, the *Santhals* who speak *Santhali*, in the Austro-Asiatic language group, and the *Oraons*, whose speak *Kurux*, in the Dravidian group. The two tribes are among the largest in the country and together form about 12% of India's population of Scheduled Tribes.[10] The non-tribal population of the village speaks languages in the Indo-Aryan subgroup of the Indo-European family. Figure 1 illustrates how the languages spoken in our village fit into the language families found in India. Since both Santhali and Kurux are in separate language families, most existing measures of linguistic diversity would place them as equidistant from the Indo-Aryan languages spoken by the rest of the population in the village.[11]

3 Data and Results

We collected our primary data in December 2016. Through house-listing operations we found a total of 224 families in the village with children enrolled in grades 2 through 5, and a total of 331 children in these grades. We collected some coarse demographic information and asked about languages spoken at home and their self-reported social identity. This could be religion, tribe or caste. We found a total of 7 groups with more than 10 households, a large number for a village of this size. These are listed in Table 1 in descending order of the number of households in the sample. For each group, the table has the principal language spoken at home, completed years of education of the household head and the fraction whose main occupation is casual labor (a proxy for poverty).[12] The last two columns of the table show the total number of children in each groups and the fraction attending the two public schools in our village.

Most of the non-tribal households speak either Hindi or one of the Eastern languages shown

in educational mobility for Scheduled Tribes relative to the rest of the population between 1961-2001.

8. We use the fractionalization index now standard in the literature, $F = 1 - \sum_i s_i^2$, where s_i's are language shares. Based on the 114 groups enumerated, Katihar has an index of 0.81.

9. Census of India, 2011.

10. The Santhals are the third-largest Scheduled Tribe (ST) in the country and the Oraons are the fourth-largest, comprising 7% and 5% of the ST population respectively (Kumar and Somanathan 2016).

11. Many popular measures of linguistic distance between two languages are based on the number of shared nodes in a language tree. Since Santhali, Kurux, and the Indo-Aryan languages are in separate linguistic trees, they have no shared nodes, and the distance between any two takes the maximal value in these measures.

Another approach starts by creating a core vocabulary of concepts, then computes the edit distance to go from the word that expresses a concept in one language to the word in the other language (see Bakker et al. (2009)). Using the Automated Similarity Judgement Program developed by the Max Planck Institute of Evolutionary Anthropology (Holman 2014; Wichmann, Holman, and Brown 2016), the distance between Hindi and these two tribal languages is very similar.

See Ginsburgh and Weber (2020, pp. 365-370) for a discussion of the various methods of measuring linguistic distance.

12. We asked about the principal occupation of the head of the household. This was either cultivation, some kind of business (owning a small shop or kiosk, for example) and casual labor. The laborers are the poorest in this group and these families are usually without any income-generating assets.

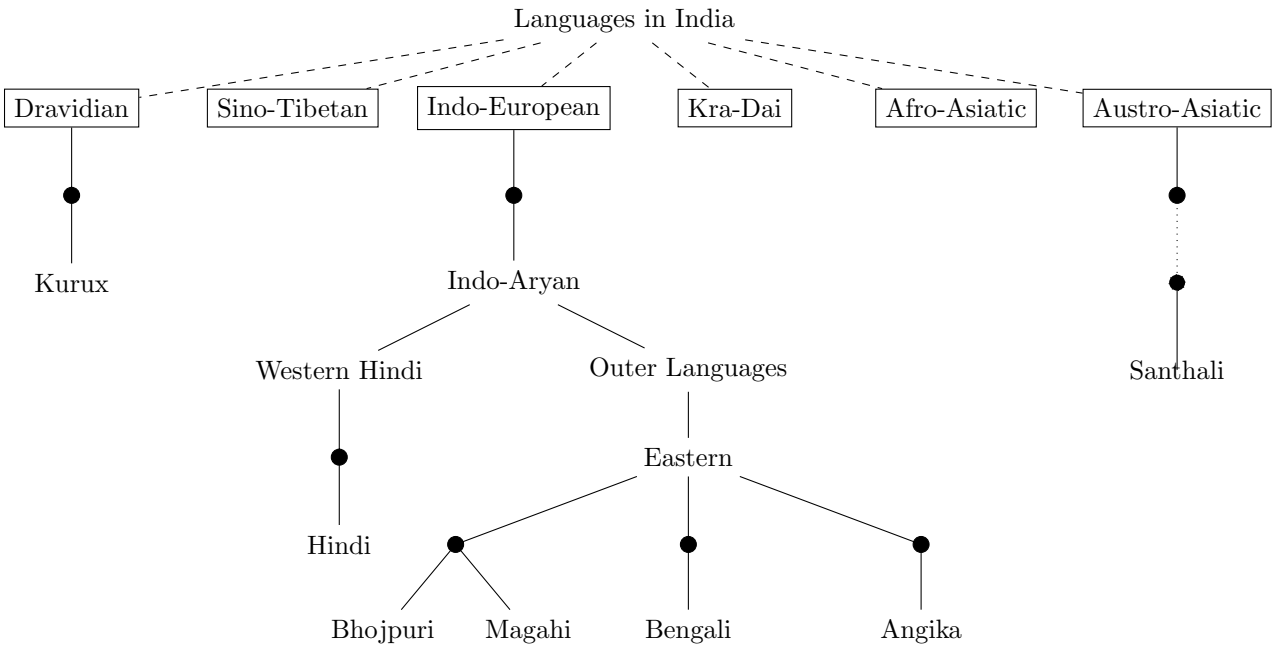

Figure 1: Languages in the sample represented in the Ethnologue linguistic tree

in Figure 1.[13] The Musahars and Turis have the lowest levels of education, followed by the Santhals and Oroans. This pattern is consistent with data for the state of Bihar from the Census of India.[14]

The distribution of these groups across the five hamlets in the village is shown in Table 2. The last column in the table also computes an isolation index proposed by Bell (1954), which captures the likelihood of within-group interactions. The asymmetry between the two tribes in terms of this index is striking. This arises because the Santhals almost exclusively inhabit two hamlets, while the Oraons all live in a single hamlet where about three-fifths of the population speaks an Indo-Aryan language closely related to the medium of school instruction. We argue that this difference could affect the familiarity of Oraon children with Hindi and this may affect their performance.

Over 90 per cent of primary schools in Bihar are managed by the Department of Education and almost all of them use Hindi as the medium of instruction.[15] This is true of the two schools in our village. We were able to test all but 30 of the 297 public school children. Most children were tested while at school. Those absent were tested at home. Our final tested

13. Of the households speaking an Eastern language, 90% speak Angika. There are 3 Hindu households who speak Kurux, the mother tongue of the Oraon group. They live in the same hamlet as the Oraons.

14. The 2011 Census for Bihar finds that only 2% of Musahars and 4% of Turis aged 15 years and above completed secondary school. Both these groups are part of the category of Scheduled Castes, which, along with the Scheduled Tribes, have been eligible for affirmative action since Independence. For both Oraons and Santhals, secondary school completion rates are about 9%. For the population of Bihar that does not fall into either of these scheduled categories, the average rate of secondary school completion is 25%. (Source: authors' calculations using Government of India (2011a, 2011b, 2011c))

15. Based on data from the 2013-14 wave of the District Information System for Education (DISE). 97.3% of the schools managed by Bihar's Department of Education use Hindi as their medium of instruction, with Urdu making up for most of the rest (2.5%).

Table 1: Group-wise distribution of characteristics

	Household Characteristics						Children	
	Households by primary language				Head of household			
	Indo-Aryan		Dravidian	Austro-Asiatic			Attend public school	
Caste	Hindi	Eastern	Kurux	Santhali	Casual labour	Education (yrs)		Number
Santhal	0	0	0	49	0.69	2.84	0.96	72
Oraon	0	0	45	0	0.60	2.18	0.89	70
Turi	7	19	0	0	0.31	1.96	0.95	41
Muslim	15	6	0	0	0.33	3.29	0.68	31
Teli	5	10	0	0	0.60	4.93	0.77	22
Mahaldar	2	9	0	0	0.27	4.18	1.00	15
Musahar	2	9	0	0	0.82	1.73	1.00	14
Others	16	27	3	0	0.30	5.48	0.92	66
Total	47	80	48	49	0.50	3.34	0.90	331

Castes with less than 10 households have been grouped together as Others. The mother is the head in 7 households where the father is deceased. The language distribution within the Eastern group is: Angika (72), Bhojpuri (5), Bengali (2), Magahi (1). The 3 non-Oraon households that speak Kurux belong to the Nai caste and inhabit the same hamlet as the Oraons.

Table 2: Hamlet-wise distribution of groups

	Hamlets						Isolation
Caste	A	B	C	D	E	Total	Index
Santhal	1	30	0	18	0	49	0.86
Oraon	0	0	0	0	45	45	0.29
Turi	0	0	0	1	25	26	0.13
Muslim	0	0	4	0	17	21	0.06
Teli	1	2	10	0	2	15	0.13
Musahar	11	0	0	0	0	11	0.35
Mahaldar	1	0	8	1	1	11	0.12
Others	15	1	16	0	14	46	0.05
Total	29	33	38	20	104	224	0.31

Castes with less than 10 households have been grouped together as Others. The isolation index is computed at the caste level. The values for Others and the Total sample are the arithmatic means over the respective constituent households.

sample is 267 children from 187 families.[16]

We administered tests in three areas: maze puzzles, elementary mathematics, and Hindi. All test questions were drawn from public school textbooks.[17] The maze test for all grades was identical and consisted of four puzzles of increasing difficulty. For mathematics, the lower

16. We found 231 children in school, the other 36 were tested at home. The remaining 30 could not be found at either location at the time of our visits.

17. Each state publishes and prescribes its own textbooks. We relied on textbooks from various states which use Hindi as a medium of instruction.

Table 3: Summary statistics (estimation sample)

	mean	std.dev.
Test scores ($N = 267$)		
Maze (%)	77.34	31.59
Math (%)	41.19	27.17
Hindi (%)	45.12	30.42
Child characteristics ($N = 267$)		
Female	0.51	0.50
Age (yrs)	8.84	1.61
Upper primary	0.48	0.50
Household characteristics ($N = 187$)		
Mother tongue		
Hindi	0.20	0.40
Eastern	0.37	0.48
Kurux	0.21	0.41
Santhali	0.22	0.42
Isolation index	0.34	0.33
Female head	0.03	0.18
Head's education (yrs)	2.86	3.50
Head's occupation		
Cultivation	0.14	0.35
Casual labour	0.51	0.50
Business or salaried	0.35	0.48

primary grades (2 and 3) were asked to count objects, order numbers into sequences and do simple addition and subtraction. Students in grades 4 and 5 had additional questions from grade 3 textbooks, including one each on multiplication and division. The Hindi test for lower grades involved naming objects and matching pictures to brief sentences describing them. For higher grades we added two comprehension exercises based on a short paragraph. Summary statistics on test scores, as well as child and household characteristics for our estimation sample are in Table 3.

Before discussing our empirical model and results, we would like to highlight sample selection issues inevitable in any study of this kind. It is not possible to identify the effects of language on learning using a cross-section of households because decisions to migrate out of a village, or attend private schools are systematically influenced by measured and unobserved child and household characteristics. The degree of selectivity in our sample is likely to vary by group. Richer families and those with better social networks outside the village are less likely to attend their local public schools. This implies that the tested children are not a random sample of children from these linguistic families or social groups. It is hard to imagine a design that generates such a sample, since we would like to estimate differences in achievement across linguistic groups for students exposed to the same school environments and families are free to move out that environment. To identify the effects of linguistic distance we therefore rely on comparing the performance of children in our sample across tasks that vary in their reliance on language for instruction. We do not claim that these estimates would be applicable to

children outside the public school system, but they do inform us about the role of language in learning within the system.

The children in our sample undertake three tasks which vary in their reliance on language: our mazes depend least on language, then mathematics problems and finally our language questions, which directly measure proficiency in Hindi. We denote by y_{ijkt}, the performance of child i with mother tongue j, belonging to social group k, on task t. This is measured in terms of the percentage of correct responses in the test. We begin with the model

$$y_{ijkt} = \alpha_i + \beta_{jt} L_j + \delta \mathbf{X}_{ik} + \epsilon_{ijkt}, \tag{1}$$

where α_i is an individual-level fixed effect on performance, L_j indicates the child's mother-tongue, $\mathbf{X}_{ik}$ is a vector of child, household and group characteristics that affect performance in the task and ϵ_{ijkt} is a random error term.

The difference in the scores on two tasks t and t' is then

$$\left(y_{ijkt} - y_{ijkt'}\right) = \left(\beta_{jt} - \beta_{jt'}\right) L_j + \left(\epsilon_{ijkt} - \epsilon_{ijkt'}\right) \tag{2}$$

The first three columns in Table 4 present least squares coefficient estimates of equation (2). We find that Santhali speakers systematically underperform in mathematics and Hindi *relative* to their performance in the mazes. There are no statistically significant differences for any of the other languages. These results are puzzling if we believe that linguistic distance affects performance in language dependent tasks, since both Kurux and Santhali are equally distant from the Indo-Aryan languages.

Table 4: Differences in percentage test scores

	(1) Math-Maze	(2) Hindi-Maze	(3) Hindi-Math	(4) Math-Maze	(5) Hindi-Maze	(6) Hindi-Math
Mother tongue						
Eastern	−3.57	−8.18	−4.61	−2.36	−6.49	−4.14
	(5.49)	(5.86)	(3.48)	(5.45)	(5.74)	(3.54)
Kurux	−4.86	−9.42	−4.56	−0.12	−2.81	−2.69
	(6.29)	(7.01)	(3.76)	(6.50)	(7.18)	(3.97)
Santhali	−16.31	−14.94	1.37	8.88	20.15	11.28
	(5.77)	(6.40)	(3.83)	(15.46)	(16.60)	(9.06)
Isolation index				−31.13	−43.37	−12.24
				(18.15)	(19.61)	(10.20)
Constant	−30.12	−23.89	6.23	−26.80	−19.27	7.53
	(4.47)	(5.03)	(2.76)	(4.94)	(5.52)	(2.97)
Observations	267	267	267	267	267	267

Heteroskedasticity-robust standard errors in parentheses.

The relative residential isolation of the Santhals compared to the Oraons, suggests that proximity to households speaking either Hindi or one of the Eastern languages may improve proficiency in this second language. Research on ethnicity and language has shown that second language acquisition is most efficient when an ethnic group has a favorable view of the community that uses this language (Trofimovich and Turuševa 2015). Residing in the same

neighborhood may do this through better interpersonal relations. To explore this hypothesis, we modify equation (1) to include the isolation index for the social group to which the child belongs. The values of this index for each group are from Table 2. We allow differential effects of the index on each of the three tasks since proximity will most affect tasks relying on language. The differenced model is now:

$$\left(y_{ijkt} - y_{ijkt'}\right) = \left(\beta_{jt} - \beta_{jt'}\right) L_j + \left(\gamma_{kt} - \gamma_{kt'}\right) I_k + \left(\epsilon_{ijkt} - \epsilon_{ijkt'}\right) \tag{3}$$

where y_{ijkt} and L_j are defined as before, I_k is the isolation index for child's social group. Coefficient estimates are in the last three columns of Table 4. Once we allow for the isolation of the social group, we find language plays no independent role in determining relative performance in the tasks.[18]

4 Implications for Policy

The National Education Policy released by the Government of India earlier this year emphasizes the importance of mother-tongue instruction. The results we present in this paper based on our study of ethnically mixed village and school environments point to the need for caution in implementing such policies. Given the enormous linguistic diversity in India, such policies would require multiple very different languages taught within the same schools, many of which struggle with both physical and human resources and often combine grades in a single classroom. Or, it could mean that existing schools that are reasonably funded get divided along ethnic lines and are no longer viable.

Studies in other countries with policies of mother tongue instruction have found that even minimally educated parents often prefer schools with instruction in widely spoken foreign languages, even in elementary school (Hungi et al. 2018; Araromi 2018). This is partly because many tribal languages do not have scripts and communities speaking these may find that bilingualism provides a more reliable route to greater mobility which may be blocked if language policies encourage isolation.

English remains the dominant language of university education in India, and if both school quality and facility in majority languages declines with the introduction of mother tongue instruction, we may exacerbate educational gaps between the tribal and other populations in India (Sridhar 1996). It is worth considering alternative interventions, formal and informal, that can bridge language gaps and encourage learning without dividing schools and students along ethnic lines.

18. It is plausible that the education and occupation of the head of the household also differentially affect tasks relying on language. When we allow for this in our specification, our results do not qualitatively change.

References

Araromi, Maxwell Olakunle. 2018. "Language policy implementation in multilingual Nigeria: French and mother-tongue experience." *LALIGENS: An International Journal of Language, Literature and Gender Studies* 7, no. 1 (February): 94–102.

Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant, and Eric W. Holman. 2009. "Adding typology to lexicostatistics: A combined approach to language classification." *Linguistic Typology* 13:167–179.

Bamgbose, Ayo. 1984. "Mother-tongue medium and scholastic attainment in Nigeria." *Prospects* 14, no. 1 (March): 87–93. ISSN: 0361-2333.

Bell, Wendell. 1954. "A probability model for the measurement of ecological segregation." *Social Forces* 32, no. 4 (May): 357–364.

Benedikter, Thomas. 2011. "The protection of linguistic rights in India: India's language policy towards linguistic minorities." *European Yearbook of Minority Issues, 2009* 8 (April): 453–482.

Chakraborty, Tanika, and Shilpi Kapur Bakshi. 2016. "English language premium: Evidence from a policy experiment in India." *Economics of Education Review* 50 (February): 1–16.

Chicoine, Luke. 2019. "Schooling with learning: The effect of free primary education and mother tongue instruction reforms in Ethiopia." *Economics of Education Review* 69 (April): 94–107.

Coyne, Gary. 2015. "Language education policies and inequality in Africa: Cross-national empirical evidence." *Comparative Education Review* 59, no. 4 (November): 619–637.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2020. *Ethnologue: Languages of the world.* 23rd ed. Online version: `http://www.ethnologue.com`. Dallas, Texas: SIL International.

Ginsburgh, Victor, and Shlomo Weber. 2020. "The economics of language." *Journal of Economic Literature* 58, no. 2 (June): 348–404.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics* 118, no. 3 (August): 1049–1074. https://doi.org/10.1162/00335530360698496.

Government of India. 2011a. "Census of India 2011: C-08 Education level by age and sex for population age 7 and above." Accessed August 22, 2020. https://censusindia.gov.in/2011census/C-series/C08.html.

———. 2011b. "Census of India 2011: SC-08 Education level by age and sex for population age 7 and above." Accessed August 22, 2020. https://censusindia.gov.in/2011census/SCST-Series/SC08.html.

———. 2011c. "Census of India 2011: ST-08 Education level by age and sex for population age 7 and above." Accessed August 22, 2020. https://censusindia.gov.in/2011census/SCST-Series/ST08.html.

Government of India. 2018. *Paper 1 of 2018: Language.* Census of India 2011. New Delhi: Office of the Registrar General, India.

———. 2020. *National Education Policy 2020.* New Delhi: Ministry of Human Resource Development, July.

Hoff, Karla, and Priyanka Pandey. 2006. "Discrimination, social identity, and durable inequalities." *The American Economic Review* 96, no. 2 (May): 206–211.

Holman, Eric W. 2014. *Programs for calculating ASJP distance matrices (version 2.2).* Available from `http://asjp.clld.org/software`.

Hungi, Njora, Joan Njagi, Patricia Wekulo, and Moses Ngware. 2018. "Effects of language of instruction on learning of literacy skills among pre-primary school children from low-income urban communities in Kenya." *Early Childhood Education Journal* 46, no. 2 (March): 187–199.

Kumar, Hemanshu, and Rohini Somanathan. 2016. "Affirmative action and long-run changes in group inequality in India." WIDER Working Paper 2016/85, June.

Laitin, David D., Rajesh Ramachandran, and Stephen L. Walter. 2019. "The legacy of colonial language policies and their impact on student learning: Evidence from an experimental program in Cameroon." *Economic Development and Cultural Change* 68, no. 1 (October): 239–272.

Odugu, Desmond Ikenna. 2011. "Education language policy process in multilingual societies: Global visions and local agendas in India, Nigeria and UNESCO." PhD diss., Loyola University Chicago, May.

Ramachandran, Rajesh. 2017. "Language use in education and human capital formation: Evidence from the Ethiopian educational reform." *World Development* 98 (October): 195–213.

Sridhar, Kamal K. 1996. "Language in education: Minorities and multilingualism in India." *International Review of Education* 42 (4): 327–347.

Trofimovich, Pavel, and Larisa Turuševa. 2015. "Ethnic identity and second language learning." *Annual Review of Applied Linguistics* 35:234–253.

UNESCO. 2016. *If you don't understand, how can you learn?* Global Education Monitoring Policy Paper 24. Paris, France: UNESCO, February.

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown, eds. 2016. *The ASJP Database (version 17).* https://doi.org/10.5281/zenodo.3835942.