# IMPACT OF MOTHER TONGUE INSTRUCTION ON ACADEMIC PERFORMANCE: EVIDENCE FROM INDIA

Uttara Balakrishnan, American Institutes for Research

Chinmaya Holla, American Institutes for Research

Pooja Nakamura, American Institutes for Research

ABSTRACT

Although children in the developing world are enrolling in school at historically high levels, improvements in children's literacy skills are still lagging. Language plays a pivotal role in the perpetuation of this crisis. More than a third of children in low-and middle-income countries are taught in a language which is not familiar to them. In this context, we examine the impact of mother tongue instruction on academic performance from India, which has one of the largest learning deficits in the world. Using a large-scale dataset, we estimate the impacts of mother tongue instruction on reading and math outcomes for children aged 5 – 16 years in rural India using two different estimation strategies. To overcome endogeneity concerns, we use fraction of schools in a state using the dominant language as the medium of instruction as an instrument for mother tongue instruction. We show that mother tongue instruction significantly improves `both reading and math scores, with impacts being concentrated among younger children. We find that school attendance is the main mechanism explaining these impacts. Using a difference-in-differences estimation we also estimate the impact of a multi-lingual education pilot program in the state of Odisha and find similar improvements in reading and math scores. Our findings provide one of the first causal estimates of mother tongue instruction outside of the African context and have significant implications for policy in light of the 2020 National Education Policy which emphasizes mother tongue instruction up to grade five.

**JEL codes**: O12, O10, I25, I28

**Keywords**: India, mother tongue instruction, primary school education, academic performance

## 1 Introduction

The obvious disparities in learning outcomes in low-and middle-income country (LMICs) have been called a learning crisis (UNESCO Institute for Statistics, 2017) and learning poverty (Azevedo et al., 2021). Less than half of all children in these countries can read a simple story by the time they are 10 years old, with that number reaching 80 percent of all children in some contexts (Azevedo et al., 2021). With extended school closures and learning loss due to the COVID-19 pandemic, the current state of learning poverty is estimated to reach 70 percent of all students across LMICs, and this generation of students' risk losing $17 trillion in lifetime earnings in present value (World Bank, 2021). Language plays a pivotal role in the perpetuation of this crisis (Evans & Mendez Acosta, 2021; Nag et al., 2019). Unsurprisingly, a child will not learn in a language that they do not use and understand (Hoover & Tunmer, 2020); yet an estimated 37 percent of students in LMICs are still not taught in a language they best understand (World Bank, 2021). An even larger share of students is unprepared to transition to a second (often postcolonial) language during the primary or middle school years, leading to deep and broad structural inequities in learning in LMIC contexts. Research shows that mother tongue (mother tongue) instruction increases attendance (C. Benson, 2005); improves cost-effectiveness (Heugh, 2004); and increases the likelihood of girls and minorities staying in schools (C. Benson, 2005).[1] These is much less evidence on the intensive margin of schooling and the impact of mother tongue instruction on academic performance. Further, most of this limited causal evidence on learning outcomes comes from Africa (Evans & Mendez Acosta, 2021). Given this lacuna, we examine if mother tongue instruction impacts academic performance for school-aged children in the large, multi-lingual context of India.

The Indian context is important for two main reasons. First, India is one of the most linguistically diverse countries in the world. "Functional multilingualism", where people function in different social domains using different languages is widespread in India, which has unique implications for education policy and learning among students given the complex mismatches between language spoken at home and medium of instruction in school (Bhatia & Ritchie, 2008). There are over 200 mother tongues in India with over 10,000 speakers according to the 2011

[1] Mother tongue, a concept which dates back to the mid-nineteenth century, assumes that a child's linguistic skills are primarily developed by the mother, whose native language will be passed down to the child (Kumar Yadav, 2014; Ladousa & Davis, 2022). The term is now used to refer to a person's first or most comfortable language.

Census. Many modern Indian states were reorganized on linguistic lines after Independence and government schools in most states use the most spoken language of the state as the official medium of instruction. While India has made progress in developing strong mother tongue education programming (NCERT, 2011), the demand to transition to English instruction in earlier and earlier grades is strong, especially given the well- documented link between English and socioeconomic mobility pathways, unique to India (Azam et al., 2013; Coleman, 2011).[2]

The second reason to examine the Indian context, is India's large learning deficit. While India has been successful in getting children into school (national primary school enrollment rate in 2018 was 96 percent), large numbers of children lack basic skills. According to the 2005 Annual Status of Education Report (ASER) report, almost 44 percent of primary school-aged children enrolled in school could not read a basic paragraph and 50 percent could not do simple subtraction (Chakraborty & Jayaraman, 2019). Not much has changed in the almost two decades since then. According to the 2019 ASER report, nationwide productivity of education declined by 18 percent between 2008 – 2018. Therefore, it is important to analyze the extent to which a dissonance between home language and the language in which the child is taught in, is contributing to this decline by significantly weakening academic performance and the future capabilities of children.

Estimating the causal impact of mother tongue instruction on learning outcomes is challenging because of multiple sources of omitted variable bias. For example, areas where children are taught in their mother tongue (rural) could systematically differ from areas where they are not (urban) due to variables such as resource availability or teacher quality.[3] Differences could also exist in state-level policies regarding medium of instruction. Another omitted variable could be family background. Parents sending children to schools where medium of instruction matches the mother tongue are systematically different, particularly because there is an increased demand from parents for English as the medium of instruction (Chakraborty & Bakshi, 2016; Shastry, 2012).[4] Systematic differences may also exist between mother tongue and English teachers. There have been wide variations noted in the language proficiency of teachers in government and private

[2] For instance, in China, until 2003 English was taught mostly as a third language or a foreign language, since historically English was seen as a threat to China's political and economic integrity (Hu, 2005). The push for English as the medium of education has been much less widespread compared to India.
[3] For instance, implementation of multilingual education policies might look different in rural versus urban settings. Children in urban poor settlements might be exposed to many languages and might come from many different backgrounds due to migration to urban areas (A. Mohanty et al., 2010; Reddy, 2011). This complex language heterogeneity can be problematic for curriculum development and implementation of pedagogical practices. In contrast in rural areas, the issue might be that students might speak the same language, but might have different dialects, especially in tribal dominated states like Odisha and Jharkhand, among others.
[4] This could also be accompanied by variations in availability of print materials in any language at home (Reddy, 2011).

schools, especially in English proficiency where the medium of instruction is supposed to be English (Kurien, 2005). There is also wide variation across and within states on pedagogical practices around teaching letters, words, and grammar (Gupta, 2013).

In this context, to overcome this endogeneity problem, we estimate the impact of mother tongue instruction using an instrumental variable (IV) framework using data from all states from the 2018 ASER dataset. Specifically, we use two instruments: (a) the fraction of schools in a state where medium of instruction is the dominant language of the state, and (b) the fraction of government schools in a state. We argue that the higher the proportion of schools where instruction is in the dominant language of the state, the more likely it is that a child is taught in his/her mother tongue. This is simply because from a supply side perspective, if more schools offer mother tongue instruction that those that do not, then parents are more likely to send their children to those schools. The exogeneity condition is also likely to hold since the availability of schools teaching in the dominant language at the state level should not directly impact an individual child's academic performance. The second IV, the fraction of government schools in a state is also likely to meet the relevance and exogeneity condition since, government schools, especially those run by state governments are more likely to have the mother tongue as the medium of instruction. Using these two instruments, also allows us to provide a test for overidentification. In addition to providing all-India estimates using the 2018 ASER data in an IV framework, we present results from a case study in Odisha. Odisha implemented a pilot multilingual education (MLE) curriculum in a phased manner in 17 tribal districts from 2008 – 2013. We use this phased implementation of MLE in a difference-in-differences framework, to provide impacts of this pilot on academic performance using the 2007 and 2014 rounds of the ASER data.

This study has four main findings. First, with respect to all-India estimates from the IV framework, we show that both, reading and math scores, are impacted positively by mother tongue instruction. Being taught in one's mother tongue increases reading scores by 12 percent over the mean and by almost 20 percent for math scores. We find marginal heterogeneous impacts by gender, with girls witnessing larger improvements, suggesting that for the most part, there is no "gender penalty" for children who are not taught in their mother tongue. With respect to age, we find that most of the impacts are concentrated among younger children. Younger children witnessing a larger improvement is that mother tongue instruction in primary grades is in line with existing literature that shows that teaching younger children in their mother tongue lays the foundation for learning later on, including second language acquisition (World Bank, 2021). Second, the difference-in-differences results from the case study of Odisha paint a similar story. The impact of the pilot MLE program in Odisha resulted in an improvement in reading scores by 0.19 points, representing a 7 percent change from the baseline mean. For math scores, there is a 0.21-point improvement, or a 9 percent change over the baseline mean. We also find that the probability of being "on-track", the correct grade-for-age, is higher as a result of the MLE program in Odisha. Third, in terms of mechanisms, we examine if mother tongue instruction impacts enrollment and attendance. We find that enrollment is higher in villages with higher levels of mother tongue instruction. Attendance on a random unannounced day is also significantly higher. Finally, we present robustness using district level data from the National Achievement Survey (NAS) data from 2021 and find similar results, with proportion of students in a district being taught in their mother tongue being positively correlated with language and math scores.

Our study contributes to several strands of literature. First, we contribute to the existing literature on the impacts of mother tongue instruction on the extensive and intensive margin of schooling. Studies from Kenya, Uganda, Ethiopia, and Cameroon have shown that mother tongue instruction results in gains in reading outcomes for primary school children (Brunette et al., 2019; Kerwin & Thornton, 2021; Piper et al., 2016; Ramachandran, 2017). In Osun State, Nigeria, using a quasi-experimental design, a study found that students in schools teaching mother tongue had higher learning outcomes relative to schools teaching both mother tongue and English (Alimi et al., 2020). In another study in Ghana, researchers examined literacy outcomes for primary students who participated in a Complementary Basic Education program taught in their mother tongue and then transitioned to government schools, some of which continued to instruct students in their mother tongue, some of which instructed students in other local languages, and some of which instructed students in English (Carter et al., 2020). Data shows that the students who continued to receive instruction from government schools in their mother tongue had the best academic outcomes, while students who received instruction in local languages struggled but eventually caught up. Further, there is also evidence that initial mother tongue instruction is positively associated with the ability to learn a second language more easily in many African countries including Ethiopia, South Africa, and Cameroon (Laitin et al., 2019; Seid, 2019; Taylor & von Fintel, 2016). Our study contributes to this literature by providing one of the first estimates of the impacts of mother tongue instruction on the intensive margin of schooling outside of Africa, in the Asian context.

Second, most of the literature examining the impact of mother tongue instruction focuses on bilingual settings, with small sample sizes. For instance, studies from bilingual settings such as Uganda and Cameroon examining causal impacts of mother tongue instruction focus on between 12 – 24 schools with a maximum of 1,800 students (Brunette et al., 2019; Kerwin & Thornton, 2021; Laitin et al., 2019). The literature in multilingual settings is more limited, especially from India. For instance, there are only a few causal studies examining the impact of language in various settings in India which are also limited to a few states. (Jain, 2017) examines the impact of historically linguistically mismatched districts in India on the extensive margin of schooling and finds that mismatched districts had lower overall adult literacy rates. Using an IV framework, Muralidharan & Sundararaman, (2015) show that students switching from Telugu medium government schools to Telugu medium private schools did better academically than those who switched to English medium schools. By presenting all-India estimates on the impacts of mother tongue instruction on the intensive margin of schooling, our study fills an important gap.

The rest of this paper is organized as follows. **Section 2** provides more details about multilingualism in India and the Indian policy context. **Section 3** describes the data, while **Section 4** details the empirical strategy. **Section 5** presents the findings, robustness checks, and also explores some potential pathways. Finally, **Section 6** concludes.

# 2 Context

## 2.1 Multilingualism in India

India is a large multi-lingual country. There are 447 living languages in the countries (Kumar et al., 2020). India ranks fourth in in the world in linguistic diversity (Skutnabb-Kangas, 2013). According to the 2011 Census, 26 percent of the population of the country is bilingual and 7.1 percent of the population is trilingual. The rate of bilingualism in the 2011 Census is the highest recorded since 1961. There was a total of 271 mother tongues recorded in the 2011 Census with 10,000 or more speakers, with 123 mother tongues grouped under 22 official or “scheduled languages” and 147 mother tongues grouped under 99 non-scheduled languages (Chandras, 2022).[5] Of the scheduled languages, the three main languages based on total number of speakers according to the Census 2011 are Hindi (528 million speakers or 44 percent of the population), Bengali (97 million speakers or 8 percent of the population), and Marathi (83 million speakers or 7 percent of the population). The distribution of these languages across the country varies considerably with linguistically pluralistic communities in many districts in India. All states have different linguistic compositions, and most languages are minority languages in some states depending on the number of speakers. For example, Tamil is a majority language in Tamil Nadu and Puducherry, but a minority language in Odisha and West Bengal. In **Table 1** we present the 22 scheduled languages in the country and the states and union territories (UTs) where they are most widely spoken.[6] States are not mandated to adopt any official language as per the Constitution of India and there is no official national language for India.

Languages in India also vary in their linguistic origins and often have different scripts. Sin-Tibetan, Indo-Aryan, Afro-Asiatic, Kra-Dai, Austro-Asiatic, and Dravidian are the six main language families in India (Eberhard et al., 2020). The languages most predominant in the south of India including Tamil, Kannada, Telugu, and Malayalam have Dravidian roots, while languages most spoken in the northern and western regions have Aryan roots. In terms of scripts, several languages in northern India derive theirs from the Nagari script. Hindi, for instance uses the Devnagari script, while Punjabi, Gujarati, and Marathi are based on Nagari scripts or versions of the Devnagari script. Urdu, spoken by 71 million speakers in India derives from the Perso-Arabic script. However, Hindi and Urdu are grammatically identical and are often considered one language with different scripts (Eberhard et al., 2020; Kumar et al., 2020).

## 2.2 Linguistic Organization of States

India has 28 states and 8 UTs which are largely organized around linguistic lines. Shortly after Independence, the Madras Presidency was divided into Tamil Nadu and Andhra Pradesh to separate the Tamil and Telugu speaking regions in 1953. This was done after sustained protests by Telugu speakers for the demand of a separate state. The state reorganization movement which resulted in the creation of Andhra Pradesh, led to the formation of the Dhar commission to examine the feasibility of reorganizing additional states on linguistic lines. This culminated into the passing

[5] The actual number of mother tongues recorded were 1,369 rationalized mother tongues which were then classified into languages spoken by 10,000 or more speakers and fewer than 10,000 speakers.
[6] https://www.asianstudies.org/publications/eaa/archives/multilingualism-in-india/

of the States Reorganization Act of 1956 which systematized state boundaries in India on the basis of language.[7] The states of Kerala, Mysore, and Madras were created in 1956 (Mysore and Madras were later renamed to Karnataka and Tamil Nadu, respectively). At the same time, the princely state of Hyderabad was partitioned on linguistic lines with regions going to Andhra Pradesh, Madhya Pradesh, and Maharashtra. Similarly, Bombay province was divided on linguistic lines into the states of Maharashtra and Gujarat. In 1966, Punjab was divided into the states of Punjab, Haryana, and Himachal Pradesh. There was also a reorganization of the northeastern region of the country on linguistic lines after the passing of the North-Eastern Areas Reorganization Act in 1971. Finally, in 2014, Telangana was created from Andhra Pradesh, separating Telugu speaking inner districts from coastal districts.

## 2.3 Language of Instruction Context and Issues in India

As noted above, multilingualism is widespread is India, with several linguistic identities co-existing within individuals and micro-regions. A. K. Mohanty, (2006) argues that these multiple identities are defining features of Indian identities and that there's a constant negotiation of these identities. Depending on the context, people move with ease between these languages. Multilingualism acts, also, as a strategy for individuals and communities to maintain their mother tongue, not by rejecting the "local" language, but by linguistic accommodation (Mohanty, 1994; Mohanty 2003). With a view to accommodate these diversities, in 1968, the National Policy in Education instated an education policy called the Three-Language Formula (TLF), which mandated that all students should be learning at least three languages (Vaish, 2008). The policy recommended that students in Hindi-speaking states learn Hindi, English, and another modern language, preferable a South Indian language. In non-Hindi-speaking states, the policy envisioned the three languages to be the local state language, Hindi, and English (Mohanty, 2006). The intention behind the trilingual formula was to build competence in at least three languages by the time a child enters secondary education (Erling et al., 2016).

Over the years, administrators have built in greater flexibility for the implementation of the TLF policy, owing to accommodating diversity in classrooms and constraints of teacher capacity

[7] As noted in Graziosi, (2017), while India was similar to the former USSR in giving importance to linguistic and cultural homogeneity, democracy and operational efficiency were still the primary driving forces of the reorganization of states. The four main principles upon which the State Reorganization Act was based included: (a) no secession permitted, and repression of all secessionist movements; (b) denial of demands to create states based upon religion (Sikh) and not language; (c) to oppose the formation of multilingual states unless all major linguistic groups supported this solution; and (d) to deny all demands for a linguistic state, even if the linguistic criterion was clear, unless such demand proved to have popular support.

(Canagarajah & Ashraf, 2013). Within a state, the mix of languages in classroom is likely to be high and vary considerably from region-to-region intra-state. A Ministry of Education report says the following about languages in schools in India: "Although we get varying accounts, India's schools teach 58 to 69 different languages either as subjects or as media of instruction." (Lindsay & Ying Tan, 2003). In fieldwork done by Erling et al, (2016), they find that Hindi was "routinely and openly used to mediate and support the study of the English textbooks, and classroom codeswitching was viewed as a legitimate pedagogic strategy".

The last three decades have witnessed the build-up of a massive demand for English-proficiency for their children among parents, that has been driven by the perceived labor market returns to knowledge of English (Azam et al., 2013; Chakraborty & Bakshi, 2016). This has led to the mushrooming of low-cost private schools, most of which offer English as a medium of instruction from Grade 1 (Ashley et al., 2014). Following a shift of students from government schools to private schools, state governments have had to adapt by teaching English as a subject from Grade 1, rather than the recommended grades 5 or 6 (Ramanujam Meganathan, 2011). Some state schools are going further and introducing an English medium of instruction strand at early levels alongside the existing local language strand. The widespread proliferation of low-cost cost private schools, where English is often the purported medium of instruction represents a significant challenge to the state school system (James & Woodhead, 2014).

In 2019, the Indian government released a draft New Education Policy (NEP), which makes amendments to the 1968 policy in that it refers to the "mandatory" teaching of Hindi in states where Hindi is not spoken, as well as inclusion of Hindi from the primary levels. These new directions have led to a backlash and widespread protests in non-Hindi-speaking regions, reflecting past protests against Hindi dominance in India.[8,9,10] At the pre-primary level, the Draft National Early Childhood Care and Education (ECCE) Policy (Government of India, 2012[11]) argues that the "mother tongue or home language of the child will be the primary language of interaction in the ECCE programs. However, given the young child's ability at this age to learn many languages, exposure to the national language English in oral form, as required, will also be explored." These

[8] https://www.hindustantimes.com/analysis/why-the-three-language-formula-is-a-bad-idea/story-xkmnLInWyJGq6Pale1RdhJ.html#:~:text=three%20years'%20time.-,This%20formulation%20is%20developmentally%20inappropriate%20for%20young%20children%20and%20cannot,and%20script%20by%20Grade%203.

[9] https://www.thehindu.com/news/national/what-is-the-three-language-formula/article27698700.ece

[10] https://timesofindia.indiatimes.com/blogs/cash-flow/is-the-three-language-formula-really-implementable-now/

[11] https://shodhganga.inflibnet.ac.in:8443/jspui/handle/10603/234067

policies reflect the decades-long tensions around this issue of the need to promote multilingual education policies in India, as well as highlighting the necessity for more science and research to be undertaken to inform the construction of effective multilingual education policies in the country.

### 2.4 Theoretical Foundation of Learning in one's Mother Tongue

There is now increasing evidence from cognitive neuroscience that first generation learners tend to perform better when taught in a language that is familiar to them (Abadzi, 2008). According to this literature the way in which language is processed impacts reasoning and cognition. When children are educated through their mother tongue in primary school, for example, they have the cognitive and linguistic foundations to learn second languages quickly; conversely, if children switch from mother tongue instruction to second language instruction too quickly, they typically struggle to learn in both their first *and* second languages and perform poorly in school (Kumar Yadav, 2014). Studies on bilingual students using functional magnetic resonance imaging (fMRI) have shown that there are high costs to language switching which impact the ability to absorb learning which is not in one's mother tongue (Bernhofer & Tonin, 2022; Grabner et al., 2012; Venkatraman et al., 2006). Language switching is when someone has to use their cognitive resources to mentally extract information in another language which is different from the language, they are currently learning the concept in. This cost of language switching even for bilingual students results in inefficiencies and additional cognitive costs relative to when children can simply retrieve information in the same language as they are learning in. Studies also show that learning in an unfamiliar language result in lower intuition since a child has to reduce their speed of talking to think about their next word which also adds to the overall cognitive costs of non-mother tongue learning (Costa et al., 2017).

## 3 Data

In this study we use student and school level data from the 2018 round of the ASER, which is representative at the district level. ASER is an annual survey stated in 2005 and administered at the same time from September to November each year to reduce systemic seasonality bias. Children aged 5 – 16 years are surveyed irrespective of schooling status – so out of school children are also included in the assessments.[12] ASER is administered at home and on weekends. This allows us to measure effects on academic performance without confounding selection related to school attendance. Since ASER is only administered in rural areas, we are unable to use this dataset to get estimates of mother tongue instruction on academic performance in urban areas.

In addition to individual and household level characteristics, ASER also tests foundational reading skills and basic math ability. It includes a reading and math assessment administered to children in their native language in an oral format.[13,14] Each assessment takes about 10 minutes with the goal being to understand the skills that have been mastered by each child in reading and math. The reading and math assessments were developed considering state mandated curriculum in each state. The reading assessment shows whether the child can read a letter, a word, a paragraph, or a story, with the highest level of reading corresponding to grade 2 curriculum.[15] The math assessment shows whether the child can recognize numbers from 1 – 9, 10 – 99, can do subtraction, or do division, with the highest level of arithmetic corresponding to grade 3 or grade 4 curriculum, depending on the state.[16] Each assessment takes about 10 minutes and are orally administered.[17,18] We calculate aggregate scores for the reading and math assessments with each score ranging from 0 to 4.

In this study we use the 2018 ASER since that is the only round which collects information on school medium of instruction and the language in which a child is tested in, which is our proxy for a child's mother tongue. In **Table 2** we examine how well this proxy does with respect to capturing one's home language. In each state, column 1 notes the most dominant "test language" (defined as at least 50 percent or more of the sample testing in that language) and column 2 notes the "most spoken language of the state", per the 2011 Census. For most states the proxy variable

[12] Children of all ages (5 – 16 years) are administered the same tool since the goal of the ASER tool is to assess competency in early foundational skills in reading and math.
[13] The reading and math assessment tools can be accessed here: http://www.asercentre.org/p/50.html
[14] ASER administers the assessments orally to minimize the cognitive demands of reading and comprehension and to maintain a standard administration approach. Children are provided with a paper and pencil to help solve any subtraction or division problems in the math assessment.
[15] The content and vocabulary of the reading assessment including selection of words, sentences, and passages are aligned to the Grade 1 and Grade 2 textbooks in each state. Passages specific to Grades 1 and 2 are developed by considering orthography indicators including conjoint letters, secondary representation of letters, and simple letter usage.
[16] For instance, 3-digit by 1-digit division is expected to be mastered by Grade 3 children in some states and by Grade 4 in others.
[17] As noted in Vagh, (2012), the ASER tools have several advantages including being relatively simple and quick to administer to more than 700,000 children every year.
[18] Vagh, (2012) describes the validation of the ASER tools in detail. Inter-rater reliability estimates using Cohen's Kappa found 0.64 for the ASER reading assessment and 0.65 for the math assessment. Concurrent validity measured using Spearman's correlation coefficient ranges from 0.90 to 0.94 with the Fluency Battery test (an adapted version of the Early Grade Reading Assessment and the Dynamic Indicators of Basic Early Literacy Skills (https://shared.rti.org/content/early-grade-reading-assessment-egra-toolkit-second-edition)).

maps well. However, for some of the states in the north-eastern part of the country including Arunachal Pradesh, Manipur, Meghalaya, and Sikkim, and for Goa and Jammu and Kashmir, the testing language is the second or third most spoken language in the state. Thus, the proxy variable for home language has some measurement error, which we address in an instrumental variable (IV) framework below.

Our total sample size includes data from approximately 362,030 children across 592 districts. In the 2018 round, a total of 19 languages covering the main languages in most states in India were included as potential options for a child to choose from for both, the medium of instruction question and for the language to be tested in for the reading and math assessments. These include Hindi, Telugu, Bengali, Marathi, Tamil, Urdu, Gujarati, Kannada, Malayalam, Oriya, Punjabi, Assamese, Manipuri, Nepali, English, Garo, Khasi, Mizo, and Bodo. These languages are the same languages used by the NCERT as well in their administration of the Foundational Learning Study, 2022, barring one language of Konkani.[19] **Table 3** presents the proportion of the sample who list these 19 languages as their mother tongue (column 1) and those listing them as their medium of instruction (column 2). The greatest mismatch is with respect to English with only 12 percent of the sample listing it is as their mother tongue but more than a quarter list it as their medium of instruction.[20] **Table 4** presents the state wise breakdown of our main variable of interest (mother tongue being the same as the medium of instruction). On average 84 percent of the sample's medium of instruction is the same as their mother tongue, with Puducherry, Dadra and Nagar Haveli, and Punjab and Haryana being the lowest at 0.28, 0.23, 0.48, and 0.49, respectively.

Finally, **Table 5** presents summary statistics for the study sample for the 2018 ASER round.[21] We present summary statistics for children whose mother tongue is the same as their medium of instruction (column 1) and for those for whom it is different (column 2). As expected, the two groups are different on both outcome measures and individual and district level controls. The average reading score for children whose mother tongue is the same as the medium of instruction is statistically similar to those children for whom it is different. However, math scores are significantly different, with students for whom medium of instruction is different from their

[19] https://dsel.education.gov.in/fls_2022
[20] Studies have found that even in schools where the medium of instruction is supposed to be English, teachers might use English textbooks, but often teach in the dominant regional language to explain concepts (Lahoti & Mukhopadhyay, 2019).
[21] Our ASER sample includes a total of 592 districts which have data for all relevant variables.

mother tongue having higher math scores. This is unsurprising, since in India, there is a high demand for English -medium education (Azam et al., 2013) even in rural areas, and students who go to English-medium schools (and therefore have a medium of instruction different from their mother tongue), usually have higher parental education and more economic resources (Azam et al., 2013; Chakraborty & Bakshi, 2016).

# 4 Empirical Strategy

## 4.1 Ordinary Least Squares Estimation

We examine the impact of being taught in one's mother tongue on academic performance using the following ordinary least squares (OLS) specification:

$$Y_{ids} = \beta_0 + \beta_1 LOI_{ids} + \beta_2 X_{ids} + \beta_3 Z_{ds} + \eta_d + \pi_a + \varepsilon_{ids} \quad (1)$$

where $Y_{ids}$ is the outcome of interest (e.g., reading or math score) for student *i* in district *d* in state *s*. $LOI_{ids}$ is an indicator variable taking the value of 1 if the medium of instruction for a student is the same as his/her mother tongue (proxied by language in which test is taken) and 0 otherwise. We include district fixed effects, $\eta_d$, to control for time-invariant district-level heterogeneity. That is, we compare students who have the same medium of instruction as their mother tongue to those who do not, *within the same district*. We also control for child age fixed effects, $\pi_a$, to account for cohort-specific effects. Individual and village level controls, $X_{ids}$, include an indicator for male gender, household size, indicator variables for if a child's mother and father attended school, indicator variable for a house having an electricity connection, household having access to a newspaper, household having access to reading materials other than newspapers, household having a scooter, household having a car, household having a television, and household having a computer.[22] We also include village level controls from the ASER including indicator variables if a village has a pucca road, village has access to electricity, village has a bank, village has an internet café, village has a government primary school, village has a government middle school, village has a government secondary school, and village has a private school. We also include district-level controls from the 2018 – 2019 round of the Unified District Information System for Education Plus (UDISE+), $Z_{ds}$ , including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools,

[22] Since the ASER data does not collect information on household income, we use mother and father's education and household size as proxies similar to other studies (see e.g., (Balakrishnan & Tsaneva, 2021).

average proportion of male teachers, average proportion of female teachers, and average proportion of teachers with a graduate degree.[23] Finally, standard errors are clustered at the district level to allow for correlation of the error term within a district.

As was seen in Table 5 above, students who are taught in their mother tongue are different statistically on many observable characteristics than those who are not. To account for this, we control for various individual, village, and district specific characteristics that might vary between children, as noted above. However, there are still possible sources of omitted variable bias. The most important source of omitted variable bias is household income. In India, given the large economic returns to English-language skills, most households, poor and rich aim to send their children to English medium schools (Bhattacharya, 2017; Faust & Nagar, 2001). This push for English language education has also resulted in some states such as Jammu & Kashmir and Nagaland, making English as the official medium of instruction in schools. We are able to control for household income to a certain extent by including proxies for household income including household size and parental education which is the standard approach in the literature (Azam et al., 2013; Card, 1999).

As noted in Azam et al., (2013), another important source of omitted variable bias is geography. Regions where mother tongue instruction is more prevalent will likely be different from places where it is not. Since education is a state subject in India, states deciding to enforce mother tongue instruction in lower grades will be systematically different from those that do not. It is also possible that states and regions where English medium of instruction is more prevalent as opposed to mother tongue instruction, are also better of economically, which can directly impact academic performance (Shastry, 2012) To account for this, we control for a child's district of residence by including district specific fixed effects. The final important source of omitted variable bias is a student's ability. A student's ability could be correlated to a family's economic background or aptitude and their decision to send their child to a school teaching in the mother tongue versus some other language. We control for this by including control variables for parental education. Several studies use parental education as proxies for an individual's ability (Card, 1999). However, it is still possible that some ability bias remains. Finally, in addition to omitted

[23] UDISE+ is a management information system (MIS) covering almost 1.4 million schools on an annual basis. UDISE+ collects school level information from all recognized schools imparting formal education from pre-primary through Grade 12 and includes information on school profile, physical school infrastructure, teachers, and enrollment data. See further details here: https://udiseplus.gov.in/#/home. UDISE+ is a school level dataset which we aggregate up to the district level and then merge to the ASER data.

variable bias, there is also possibility of measurement error in our main endogenous regressor itself as noted previously. Specifically, we proxy for mother tongue using the test language of the child. Since the number of language options for both mother tongue and the test language of the child are the same, it is reasonable to believe that the child will choose to be assessed in his/her mother tongue. However, it is still possible that there is some measurement error due to the use of this proxy. This is especially true in the states of Jammu and Kashmir and Arunachal Pradesh, where the most spoken language according to the 2011 Census, are Kashmiri and Nissi, respectively, but most children still chose to be tested in English in both states. To address the multiple sources of omitted variable bias and possible measurement error, we use an IV strategy outlined below.

### 4.2 Instrumental Variables (IV) Framework

As noted above, the OLS estimates potentially suffer from multiple sources of omitted variable bias. To address this endogeneity problem, we instrument for a child's medium of instruction being the same as their mother tongue using two separate instrumental variables. Our first IV is the proportion of schools at the state level where the medium of instruction is the dominant language of the state according to the 2011 Census Language Atlas.[24] Specifically, the Census 2011 records the number of speakers for each major language in each state. We merge this information to the UDISE+ data which records for each school in the state the medium of instruction of that school. We aggregate the UDISE+ data to the state level and calculate the percentage of schools using the dominant language of the state as the medium of instruction as per the 2011 Census. We restrict the data to only schools in rural areas to match with the ASER data. For our instrument to be valid it must meet both, the relevance and exogeneity conditions. We argue that the higher the proportion of schools where instruction is in the most spoken language of the state, it more likely it is that a child's mother tongue is the same as his/her medium of instruction. This is simply because from a supply side perspective, if more schools offer mother tongue instruction that those that do not, then parents are more likely to send their children to those schools. The exogeneity condition is also likely to hold since the availability of schools teaching in the most spoken language at the state level should not directly impact an individual child's academic performance.

[24] https://censusindia.gov.in/nada/index.php/catalog/42561#metadata-themes_topics

**Figure 1** presents variation in our instrument by state. States in the northeastern region of the country like Sikkim and Arunachal Pradesh have the lowest proportion of schools where the medium of instruction is the most spoken language of the state. For instance, while Nissi is the most spoken language in Arunachal Pradesh, almost 97 percent of schools have English as their main medium of instruction. In southern states including Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, and Telangana, about three quarters of schools have the most spoken language of the state as the medium of instruction, while a quarter of schools have other languages, including English as the main medium of instruction. Similar instruments have been used in other literature examining the impact of electrification on household welfare and empowerment (see e.g., Sedai et al., 2021).

The second IV that we use is the proportion of government schools at the state level. Government schools, especially those run by state governments are more likely to have the mother tongue as the medium of instruction and thus, would also fulfill the relevance condition (Jha, 2021).[25] Schools funded by state governments in India largely have the main language of the state as the medium of instruction. Like the previous IV, we also aggregate this up to the state level thereby ensuring that the IV will not directly be associated with an individual student's outcome variable.

Using these two instruments, we estimate the impact of mother tongue instruction on academic performance using the following two stage least squares (2SLS) specification:

$$LOI_{ids} = \alpha_0 + \alpha_1 PropLanguageSchools_s + \alpha_2 PropGovtSchools_s + \alpha_3 X_{ids} + \alpha_4 Z_s + \eta_a + u_{ids} \quad (2)$$

$$Y_{ids} = \beta_0 + \beta_1 \widehat{LOI}_{ids} + \beta_2 X_{ids} + \beta_3 Z_s + \pi_a + \varepsilon_{ids} \quad (3)$$

where $PropLanguageSchools_s$ is equal to the proportion of schools in state *s* with the same medium of instruction as the most spoken language in the state; and $PropGovtSchools_s$ is the proportion of government schools in state *s.* All other variables are the same as before. The main identifying assumption of the IV is that *conditional* on the individual, household, and state level controls, the two instruments impact academic performance only through their impact on the likelihood that the student's medium of instruction is the same as their mother tongue. The identifying assumption could be violated if households choose to send children to a school with a

[25] As noted above, India's three language formula (TLF) recommended the teaching of three languages in schools in India, with the first language being the most common regional language or mother tongue (Jha, 2021).

specific medium of instruction *based on the* proportion of government schools or proportions of schools with the same medium of instruction as the most spoken language *at the state level*. However, an individual household is unlikely to make schooling decisions based on state level factors. It is also possible that migration to areas with a higher supply of schools with a household's 'preferred' medium of instruction could bias our estimates as well. Yet, selective migration is not a concern in our sample since several studies have documented that India witnesses very low levels of inter-district rural migration (Imbert & Papp, 2015; Munshi & Rosenzweig, 2009; Topalova, 2010).

### 4.3 Case Study from Odisha

In addition to examining results at an all-India level, we also examine the impact of mother tongue instruction by exploiting plausible exogenous variation in mother tongue programming in one specific state, Odisha.[26] Odisha started a Multilingual Education (MLE) program in 2006, which was implemented in a phased manner in the 17 tribal districts from 2008 – 2013. Ethnographically, Odisha offers a unique case study to understand the impacts of mother tongue education on academic performance since it has the largest number of tribal communities in the country (62 communities as per the Census 2011, including 13 Particularly Vulnerable Tribal Groups).[27,28] While overall literacy rates have seen an improvement, per the 2011 Census literacy rates for scheduled castes (STs) in Odisha is only 38 percent compared to the state average of 63 percent (Ota et al., 2020). For most ST children, their mother tongue is different from the state language of Odisha – Oriya, which is the most common medium of instruction in state government schools in Odisha. Textbooks are primarily in Oriya making it extremely difficult for these children to understand what is being taught in class, impacting learning outcomes and retention of students. On the teaching side, there is inadequate training and lack of pedagogical tools to deal with diverse classroom environments. Dropout rates are highest in the 10 districts with a high proportion of ST population (NCERT, 2011). In a 2003 Government of Odisha vision document, inappropriate language of instruction and unsuitable curricula and textbooks, lack of community participation due to language issues, emerge as primary barriers for improving learning outcomes for tribal

[26] Odisha is in the eastern part of India and has a population of 41 million (Census 2011), with Scheduled Tribes (STs) comprising 23 percent of the population.
[27] https://repository.tribal.gov.in/bitstream/123456789/73776/1/SCST_2018_book_0016.pdf
[28] Overall, on an all-India level, STs constitute 8 percent of the population and have much lower secondary school completion rates than the average (14 percent vs 31 percent of all India). Primary school completion rates are also lower at 41 percent compared to the all-India average of 58 percent.

children and reducing dropout rates (NCERT, 2011). In this context, in 2006, based on recommendations from the State Tribal Advisory Committee, the state government decided to adopt MLE in tribal districts across Odisha.

Based on the guiding principles of the National Curriculum Framework, 2005, MLE was adopted in a phased manner, with 10 tribal languages being selected in the beginning as the medium of instruction in government schools across 8 tribal districts.[29] These languages included Santhali, Munda, Koya, Kuvi, Kishan, Oram, Soura, Kui, Bonda and Juang across the districts of Gajapati, Keonjhar, Mayurbhanj, Malkangiri, Sambalpur, Sundergarh, Rayagada, and Kandhamal. Implementation occurred in these 8 districts from 2008 – 2010 across 545 schools. From 2011 – 2013, the MLE program was expanded to 1485 in 21 tribal languages schools across all the remaining tribal districts including Anugul, Bargarh, Balasore, Dhenkanal, Ganjam, Kalahandi, Koraput, Nawarangapur, and Nuapada. MLE was implemented through multiple approaches. First, a package of instructional materials including syllabus, books, teaching learning materials (TLM), and thematic curriculum, were developed in the tribal languages and the main state language Oriya. The approach to developing the materials was based on two constructs: basic interpersonal communication skill (BICS) and the cognitive academic language proficiency skill (CALPS) (NCERT, 2011) with the goal of combining cultural aspects of the tribal communities with the curricular knowledge of the school. This was done with active engagement of teachers from the tribal communities. Second, teachers were training first on the theories and methods of the MLE and then were trained using participatory approaches on the teaching methods to be employed including daily lesson plans and weekly lesson plans, among others. Mostly tribal teachers were engaged as MLE teachers, or active recruitment was conducted for tribal teachers in locations which had low existing teacher capacity. MLE teachers were also provided hands on training in one MLE school. In addition to teachers, principals, cluster, and block level coordinators were also trained on the MLE approaches. In addition to the 15-day trainings, refresher trainings were also held, in addition to monthly meetings to discuss ongoing challenges to MLE implementation and to share best practices and successes. Block and cluster level officers were also instructed to monitor the day-to-day implementation of MLE in classrooms.

[29] The 2005 National Curriculum Framework recommended that language teaching needs to be multilingual not only in terms of the number of languages offered to children but also in terms of evolving strategies that would use the multilingual classroom as a resource.

We use the phased implementation of MLE across Odisha in a difference-in-differences framework to estimate intent-to-treat (ITT) effects of the MLE program on academic performance. We will compare children in the 17 tribal districts ('treatment') to children in the other 13 districts in Odisha ('comparison'), before the beginning of the MLE program to after the program had been rolled-out in all tribal districts. We use the 2007 ASER round as the "pre-period" and the 2014 ASER round as the "post-period" since MLE was rolled out in all 17 districts by 2013. In other words, the first differences will compare outcomes in the pre- and post-periods and the second difference will difference out outcomes between treatment and comparison districts. The advantage of using this estimation approach is that we use plausibly exogenous variation in program implementation and mitigate omitted variable bias from the OLS estimation noted in Section 4.1.

We will estimate ITT impacts using the following difference-in-differences equation:

$$Y_{idt} = \beta_0 + \beta_1 MLE_d + \beta_2 Post_t + \beta_3 Post * MLE_d + \beta_4 X_{idt} + \pi_a + \eta_d + \varepsilon_{idt} \qquad (4)$$

where $Y_{idt}$ is the outcome of interest (e.g., reading or math score) for student *i* in district *d* at time *t*. $MLE_d$ is an indicator variable taking the value of 1 if the district is a tribal district and 0 otherwise. $Post_t$ is an indicator which is 1 if the observation is from 2014 and 0 in 2007. $X_{idt}$ includes limited individual controls since the 2007 ASER round had minimal background information. Controls include an indicator for male gender, household size, and indicator variable for a child's mother having attended school. All other variables are similar to those described for Equation 1 above. Standard errors are clustered at the district level to allow for correlation of the error term within a district. The coefficient on the interaction term, $\beta_3$, is our main difference-in differences estimate. The identifying assumption is that trends in treatment and comparison districts in our main outcomes would have remained the same in the absence of the MLE program. While we are unable to directly test for this since the first round of the ASER data that is publicly available is 2007, we examine baseline differences in the treatment and comparison districts in **Table 6** and find no significant differences.

## 5 Findings

### 5.1. Results from ASER

We first present the OLS estimates from **Equation 1** in columns 1 – 6 in **Table 7**. Every coefficient represents a different regression. Columns 1 and 4 show results in reading and math scores, respectively without any controls and only district fixed effects; columns 2 and 5 include household and village level controls from the ASER; and columns 3 and 6 include district level controls from the UDISE+. Without any controls, mother tongue instruction is associated with positive changes in reading scores, and a negative change in math outcomes (columns 1 and 4). This is not surprising since OLS results without any controls suffer from multiple sources of omitted variable bias as noted previously. Adding household and village controls in columns 2 and 5, results in positive improvements in both reading and math scores. Overall, in the specification with all controls in columns 3 and 6, mother tongue instruction is associated with a 0.31-point improvement in reading scores (12 percent improvement over the mean) and 0.13-point improvement in math scores (5.2 percent improvement over the mean), both statistically significant at the 1 percent level. Reading and math scores for both girls and boys are positively correlated with mother tongue instruction, with reading and math scores for girls being associated with a marginally larger improvement than for boys. As noted above, OLS estimates are potentially downward biased because of omitted variable bias and reverse causality. Consequently, we present results from IV estimates next.

We first test the first stage of the IV specification from **Equation 2.** There is a positive and statistically significant (at the 1 percent level) impact of the two instruments (i.e., proportion of schools in the state with medium of instruction as the most spoken language and proportion of government schools) on mother tongue instruction. The first-stage joint F-statistic is 120.04. Since we have two instruments, we are able to conduct the Sargen-Hansen test of overidentification. The p-value from the chi-square test is

In **Table 8**, Panel A, we present the second stage IV results for the impact of mother tongue instruction on reading and math scores. We find that, results from the IV specification are similar to the OLS results. In the specification without any controls (columns 1 and 4), reading and math scores increase by 0.58 points and 0.50 points, respectively. In our preferred specification with all controls (columns 3 and 6), mother tongue instruction causes a 0.32-point improvement in reading scores (12 percent over the mean or 0.21 standard deviations) and a 0.47-point improvement in math scores (20 percent improvement over the mean or 0.31 standard deviations), statistically significant at the 10 and 5 percent levels, respectively. Our overall findings are consistent with Piper et al. (2016), who find that a mother tongue literacy instruction experiment in Kenya improved literacy outcomes by 0.3 – 0.6 standard deviations. Similarly, Seid, 2019) find that mother tongue instruction in Ethiopia in grades 1 – 4, improved math scores by 0.16 standard deviations.

In Panels B and C of **Table 8**, we examine heterogeneous impacts for girls and boys. In general, mother tongue instruction leads to higher reading and math scores for both boys and girls, with improvements in math scores being marginally higher for girls. This is in line with some nascent studies from Africa which investigate the gendered impact of mother tongue instruction that show that girls perform better. For instance, a study from Namibia (Van Wyk & Mostert, 2016), finds that second language acquisition, measured by vocabulary and oral communication tests, is stronger for girls taught in their mother tongue versus those who are not, relative to boys. This is most likely because girls have more intersecting layers of disadvantage and lower baseline values of learning outcomes. Studies from Africa and Latin America (C. Benson, 2002; Hovens, 2002) examining gendered differences in bilingual education have found that girls learning in mother tongues or other familiar languages have a higher probability of staying in school and doing better on learning tests. They are also less likely to repeat grades suggesting that mother tongue instruction is correlated with larger improvements for girls relative to boys (C. J. Benson, 2002). Small case studies from Niger, Mozambique, and Guinea-Bissau (C. Benson, 2002; C. J. Benson, 2002; Caral. Benson, 2005) have shown that the potential mechanism behind girls benefitting more from mother tongue instruction have included higher probability of girls enrolling if they are taught in a familiar language and by association a familiar culture and set of values. mother tongue based curriculum might make it easier for parents to communicate and participate in school activities, helping increase the perceived relevance of schooling for girls by their parents. Finally, teachers from similar linguistic communities who interact with parents are more likely to be deemed trustworthy by parents and less likely to exploit girls (Hovens, 2002).

In **Table 9** we also examine heterogeneous impacts by age. Results are presented separately for children of primary school age (5 – 10 years) and older children (11 – 16 years), as they might respond differently to mother tongue instruction. Only younger children witness impacts of mother tongue instruction. The impacts for primary school children are large and statistically significant at the 1 percent level for both reading and math scores. There are no statistically significant impacts for older children for either reading or math scores. This is not unsurprising since studies show that mother tongue instruction in early grades impacts outcomes at both the extensive and intensive margin of schooling including second language acquisition in later grades (Seid, 2019; Taylor &

von Fintel, 2016). For instance, Nakamura et al., 2019) show that there is a certain level of mastery that is required in one language (most commonly the mother tongue) before mastery is achieved in another language. Most of the literature examining impacts of mother tongue instruction focus on primary school students during which time the returns to mother tongue education are largest. Studies from Ethiopia also show that mother tongue instruction in primary school improves later life outcomes including higher school completion, employment, and wage rates (Ramachandran, 2017; Seid, 2022).

### 5.2. Results from Odisha Case Study

ITT results from the difference-in-differences estimation for the impact of the MLE program in Odisha are presented in columns 1 - 6 **Table 10**. Every coefficient is from the interaction term in Equation 4 and represents a different regression. In general, the difference-in-difference results are similar the OLS and IV results presented above. On average, in the specification with all controls, reading scores improve by 0.19 points (statistically significant at the 1 percent level). This represents a 7 percent change from the baseline mean. For math scores, there is a 0.21-point improvement (significant at the 5 percent level), representing a 9 percent change over the baseline mean. In Panels B and C, we examine impacts for girls and boys separately and find that, similar to the OLS and IV results, girls and boys have similar ITT effects in Odisha as well. In addition to reading and math outcomes, we also examine impacts on dropout rates and an indicator for being “on-track”. (Shah and Steinberg 2017) define “on-track” children as those who are in the correct grade for their age. Approximately, 88 percent of children in the Odisha sample are in the correct grade-for-age. In columns 7 and 8 in **Table 10**, we show the MLE program does not have a statistically significant impact on the probability of children dropping out of school. On the other hand, the MLE program significantly increases the probability of children being on-track by 2.4 percentage points (statistically significant at the 5 percent level). Our findings are also consistent with Seid (2016) who finds that mother tongue instruction in Ethiopia improves probability of enrollment and being in the right grade for age. Similarly, Jain (2017) finds that historically linguistically mismatched districts have lower literacy rates and college graduation rates in India.

### 5.3. Robustness using National Achievement Survey (NAS) Data

We conducted robustness tests using district-level data from the National Achievement Survey (NAS) that was conducted in 2021.[30] NAS is a nationally representative large-scale survey of students' learning undertaken by the Ministry of Education, Government of India. We use data from Grades 3 and 5 in State Govt. schools, Govt. Aided schools, Private Unaided recognized schools, and Central Government schools. The test contains psychometrically reliable and valid questions on Language and Mathematics. NAS is based on grade-specific competencies and learning outcomes designed by the National Council of Educational Research and Training (NCERT) for Language and Mathematics. For each grade, NCERT developed four comparable test booklets, each of which contained questions on all subjects that were tested. There were 47 items in the Grade 3 assessment and 53 items in the Grade 5 assessment. Language and math scores are calculated on a 0 – 100 scale.

Using the NAS, we assess the impact of being taught in one's mother tongue on academic performance using an OLS model with the following specification, with all children from Grades 3 and 5 being pooled together:

$$Y_{ds} = \beta_0 + \beta_1 LOI_{ds} + \beta_2 X_{ds} + \eta_s + \varepsilon_{ds} \quad (5)$$

where $Y_{ids}$ is the outcome of interest (e.g., average Language or Math score at the district level) for district $d$ in state $s$. $LOI_{ds}$ is a continuous variable indicating the percentage of students at the district level whose mother tongue matches the medium of instruction for district $d$ in state $s$. We also include state fixed effects, $\eta_s$. District level controls, $X_{ds}$, include a battery of survey responses by students, teachers, and school administrators, aggregated to the district-level. These include responses related to school infrastructure, work practices, parental engagement, interest in and understanding content, and support for learning. We also include district-level shares of households which own land, have a monthly income greater than 10,000 INR, share of individuals who have completed primary education, and the share of scheduled caste (SC) and scheduled tribe (ST) households in the district. We obtain these data from the Socio-Economic and Caste Census (SECC) of 2012, compiled by the Socioeconomic High-resolution Rural-Urban Geographic Platform (SHRUG) team.[31] The combined NAS – SECC sample includes data from 484 districts.

[30] https://nas.gov.in/report-card/2021
[31] https://www.devdatalab.org/shrug

Finally, standard errors are clustered at the district level to allow for correlation of the error term within a district. **Appendix Table 1** presents the summary statistics for the scores and the controls.

We present the OLS estimates from **Equation 5** in columns 1 – 6 in Panel A of **Appendix Table 2**. Columns 1 and 4 show results in language and math scores, respectively without any controls and only state fixed effects; columns 2 and 5 include district level controls from the NAS survey data; and columns 3 and 6 include district level controls from the SECC data. Without any controls, mother tongue instruction is associated with positive changes in both reading and math scores (columns 1 and 4). After adding student, teacher, and school administrator controls in columns 2 and 5, the magnitude of the association decreases but the results are still positive and statistically significant for both language and math scores at the 1 percent level. Overall, in the specification with all controls in columns 3 and 6, mother tongue instruction is associated with a 0.11 percentage point increase in language scores and 0.21 points improvement in math scores, both statistically significant at the 1 percent level. Heterogeneous analysis for Grades 3 and 5 separately are reported in Panels B and C, but are estimated imprecisely due to much smaller sample sizes.

### 5.4. Mechanisms

Previous studies examining the impact of mother tongue instruction on learning have identified several potential mechanisms. Students learning in a familiar language might be more likely to adjust better in school (Seid, 2019; Trudell, 2005). Instruction in one's own language is also more likely to increase classroom participation thereby helping them develop higher level cognitive skills quickly (Sonaiya, 2002). Further, as noted previously teachers teaching in a language familiar to the student are also potentially deemed as more trustworthy and less likely to sexually abuse students thereby increasing attendance and attentiveness (Caral. Benson, 2005). All of these pathways are mediated through higher school attendance and attentiveness. While we do not have measures for attentiveness, we provide suggestive evidence below on the relationship between mother tongue instruction and school level attendance.

Since the student-level ASER data does not have information on attendance, we test if mother tongue instruction leads to lower student absences using the school level data collected by ASER. School attendance in each grade our sample is measured by the number of children in attendance on a random day divided by the total enrollment in that grade. We estimate a similar IV specification as before, except at the school/village level. The independent variable is the percentage of children in a village getting mother tongue instruction. Following (Adukia, 2017) we present estimates for the average attendance rate at the school/village level. Results are presented in columns 1 – 2 of **Appendix Table 3**. On average attendance rates are higher for both younger and older children. Grades 1 to 5 see an increase in attendance of 47 percentage points and Grades 6 – 8 also see an increase of 44 percentage points (both statistically significant at the 1 percent level). It is not surprising that older children have impacts on the extensive margin of school (i.e., attendance) and not in the intensive margin of schooling (i.e., academic performance), since mother tongue instruction might make older children more regularly attend school, but the impacts on learning once they are in school might be more limited. Since the attendance data is at the school level, we are unable to explore the gendered nature of attendance impacts. We also examine teacher attendance in column 3 of Appendix Table 3. Teachers who are more comfortable teaching in a language they know and are familiar with, are more likely to come to school to teach in that language. Teacher attendance at the school level is measured in the same way as student attendance i.e., attendance of regular teachers on a random day divided by the total number of regular teachers. We find that there are no statistically significant impacts of mother tongue instruction on teacher attendance in our sample.

## 6 Conclusion

Bilingual learning theory clarifies that the cognitive and linguistic skills acquired in the mother tongue can serve as a critical foundation for learning new language(s) through the transfer of specific skills across (August, Diane Shanahan, 2006; Chung et al., 2019; P. R. Nakamura et al., 2019). While curriculums around the world are often taught in English or other post-colonial languages due to their global demand and connection to social mobility, these education systems fail to meet the needs of linguistically diverse students, many of whom do not receive educational instruction in their mother tongue (Nag et al., 2019; P. Nakamura et al., 2023). Being taught in languages that are not spoken in the home limits development of reading and writing skills, isolates students who often can’t turn to parents for educational support, lowers confidence and participation in the classroom, and negatively impacts students’ education (UNESCO, 2016). Furthermore, when curriculums are only taught in hegemonic languages, already limited opportunities for ethnolinguistic minorities decrease further (UNESCO, 2016). India is a multilingual country with relatively strong mother tongue education programming (NCERT, 2011); however, the country also faces pressure to introduce English in earlier grades (Azam et al., 2013), perpetuating the learning crisis.

In this study, we examine the impact of mother tongue instruction in India using two different estimation strategies. We find that mother tongue instruction leads to better reading and math outcomes for boys and girls of primary school age. The main mechanism driving these effects is school attendance. We find that student attendance is higher when instruction is in a familiar language. Additionally, we find evidence of a positive impact on being in the correct grade-for-age. We also find suggestive evidence that teacher absences are lower suggesting that mother tongue instruction is doubly beneficial, children are directly impacted by lower school absences, and indirectly impacted by their teachers' school attendance.

India's recent NEP 2019 recommends mother tongue teaching in government and private schools until Grade 5, and preferably until Grade 8. However, this is only likely to happen if there is sufficient demand for mother tongue education from parents, as opposed to English-medium education as noted in Azam et al., (2013). In order to generate demand for mother tongue education, rigorous causal evidence on the impacts of mother tongue education in India are essential and our study starts to fill this gap. Future research is needed to examine more closely and causally the pathways through which mother tongue educations impact learning outcomes in India. Additional research is also essential to understand longer term impacts on school completion and employment outcomes.

## References


Abadzi, H. (2008). Efficient learning for the poor: New insights into literacy acquisition for children. *International Review of Education*, *54*(5–6). https://doi.org/10.1007/s11159-008-9102-3

Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics*, *9*(2), 23–59.

Alimi, F. O., Tella, A., Adeyemo, G. O., & Oyeweso, M. O. (2020). IMPACT OF MOTHER TONGUE ON PRIMARY PUPILS' LITERACY AND NUMERACY SKILLS IN OSUN STATE. *International Online Journal of Primary Education*, *9*(2).

Ashley, L. D., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., & Rose, P. (2014). The role and impact of private schools in developing countries–bibliography and literature reviews. In *London: DfID*.

August, Diane Shanahan, T. (2006). Developing Literacy in Second-Language Learners : Report of the National Literacy Panel on Language-Minority Children and Youth Edited by. *Center for Applied Linguistics*.

Azam, M., Chin, A., & Prakash, N. (2013). The returns to english-language skills in India. *Economic Development and Cultural Change*, *61*(2). https://doi.org/10.1086/668277

Azevedo, J. P., Goldemberg, D., Montoya, S., Nayar, R., Rogers, H., Saavedra, J., & Stacy, B. W. (2021). Will Every Child Be Able to Read by 2030? Defining Learning Poverty and Mapping the Dimensions of the Challenge. In *World Bank Group Education Global Practice*.

Balakrishnan, U., & Tsaneva, M. (2021). Air pollution and academic performance: Evidence from India. *World Development*, *146*. https://doi.org/10.1016/j.worlddev.2021.105553

Benson, C. (2002). Bilingual education in Africa: An exploration of encouraging connections between language and girls' schooling. *Education—A Way out of Poverty*.

Benson, C. (2005). The importance of mother tongue- based schooling for educational quality Commissioned study for EFA Global Monitoring Report. *Commissioned Study for EFA Global Monitoring Report*.

Benson, C. J. (2002). Real and Potential Benefits of Bilingual Programmes in Developing Countries. *International Journal of Bilingual Education and Bilingualism*, *5*(6). https://doi.org/10.1080/13670050208667764

Benson, Caral. (2005). Girls, Educational Equity and Mother Tongue-Based Teaching. In *UNESCO Bangkok. Asia and Pacific Regional Bureau for Education, PO Box 967, Prakhanong Post Office, Bangkok 10110*.

Bernhofer, J., & Tonin, M. (2022). The effect of the language of instruction on academic performance. *Labour Economics*, *78*. https://doi.org/10.1016/j.labeco.2022.102218

Bhatia, T. K., & Ritchie, W. C. (2008). Bilingualism in South Asia. In *The Handbook of Bilingualism*. https://doi.org/10.1002/9780470756997.ch29

Bhattacharya, U. (2017). Colonization and English ideologies in India: a language policy perspective. *Language Policy*, *16*(1). https://doi.org/10.1007/s10993-015-9399-2

Brunette, T., Piper, B., Jordan, R., King, S., & Nabacwa, R. (2019). The impact of mother tongue reading instruction in twelve Ugandan languages and the role of language complexity, socioeconomic factors, and program implementation. *Comparative Education Review*, *63*(4). https://doi.org/10.1086/705426

Canagarajah, S., & Ashraf, H. (2013). Multilingualism and education in South Asia: Resolving policy/practice dilemmas. In *Annual Review of Applied Linguistics* (Vol. 33). https://doi.org/10.1017/S0267190513000068

Card, D. (1999). Chapter 30 The causal effect of education on earnings. In *Handbook of Labor Economics: Vol. 3 PART* (Issue 1). https://doi.org/10.1016/S1573-4463(99)03011-4

Carter, E., Sabates, R., Rose, P., & Akyeampong, K. (2020). Sustaining literacy from mother tongue instruction in complementary education into official language of instruction in government schools in Ghana. *International Journal of Educational Development*, *76*. https://doi.org/10.1016/j.ijedudev.2020.102195

Chakraborty, T., & Bakshi, S. K. (2016). English language premium: Evidence from a policy experiment in India. *Economics of Education Review*, *50*. https://doi.org/10.1016/j.econedurev.2015.10.004

Chakraborty, T., & Jayaraman, R. (2019). School feeding and learning achievement: Evidence from India’s midday meal program. *Journal of Development Economics*, *139*, 249–265.

Chandras, J. (2022). Hypothetical reported speech as pedagogical practice in multilingual classrooms in India. *Language and Education*, *36*(4). https://doi.org/10.1080/09500782.2021.1981368

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. In *Journal of Neurolinguistics* (Vol. 50). https://doi.org/10.1016/j.jneuroling.2018.01.003

Coleman, H. (2011). Developing Countries and the English Language: Rhetoric, risks, roles and recommendations. In *Dreams and Realities: Developing Countries and the English Language*.

Costa, A., Vives, M., & Corey, J. D. (2017). On Language Processing Shaping Decision Making. *Current Directions in Psychological Science*, *26*(2). https://doi.org/10.1177/0963721416680263

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2020). *Ethnologue: Languages of the World 23rd edition*. Dallas, Texas: SIL International.

Erling, E. J., Adinolfi, L., Hultgren, A. K., Buckler, A., & Mukorera, M. (2016). Medium of instruction policies in Ghanaian and Indian primary schools: an overview of key issues and recommendations. *Comparative Education*, *52*(3). https://doi.org/10.1080/03050068.2016.1185254

Evans, D. K., & Mendez Acosta, A. (2021). Education in Africa: What Are We Learning? *Journal of African Economies, 30*(1). https://doi.org/10.1093/jae/ejaa009

Faust, D., & Nagar, R. (2001). Politics of Development in Postcolonial India: English-Medium Education and Social Fracturing. *Economic and Political Weekly*, *36*(30).

Grabner, R. H., Saalbach, H., & Eckstein, D. (2012). Language-Switching Costs in Bilingual Mathematics Learning. *Mind, Brain, and Education*, *6*(3). https://doi.org/10.1111/j.1751-228X.2012.01150.x

Graziosi, A. (2017). India and the Soviet Model: The Linguistic State Reorganization and the Problem of Hindi. *Harvard Ukrainian Studies, 35(1/4), 443-471*.

Gupta, R. (2013). More than ABC: Instructional Practices and Children's Understanding of Literacy through English. *Contemporary Education Dialogue*, *10*(1). https://doi.org/10.1177/0973184912465160

Heugh, K. (2004). *Is multilingual education really more expensive. Making multilingual education a reality for all*.

Hoover, W. A., & Tunmer, W. E. (2020). *Using the Cognitive Foundations Framework to Create Coherence in Reading Practices*. https://doi.org/10.1007/978-3-030-44195-1_13

Hovens, M. (2002). Bilingual Education in West Africa: Does It Work? *International Journal of Bilingual Education and Bilingualism*, *5*(5). https://doi.org/10.1080/13670050208667760

Hu, G. (2005). English language education in China: Policies, progress, and problems. In *Language Policy* (Vol. 4, Issue 1). https://doi.org/10.1007/s10993-004-6561-7

Imbert, C., & Papp, J. (2015). Labor Market Effects of Social Programs - Evidence from India's Employment Guarantee. *American Economic Journal: Applied Economics*, *7*(2), 233–263. https://doi.org/10.1257/app.20130401

Jain, T. (2017). Common Tongue: The Impact of Language on Educational Outcomes. In *Journal of Economic History* (Vol. 77, Issue 2). https://doi.org/10.1017/S0022050717000481

James, Z., & Woodhead, M. (2014). Choosing and changing schools in India's private and government sectors: Young Lives evidence from Andhra Pradesh. *Oxford Review of Education, 40*(1). https://doi.org/10.1080/03054985.2013.873527

Jha, J. (2021). Public Policy Dilemma of Choosing the Medium of Instruction for School Education: A Case for Questioning Fallacies and Connecting Objectives. *Journal of Asian Development Research*. https://doi.org/10.1177/2633190x211034697

Kerwin, J. T., & Thornton, R. L. (2021). Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures. *Review of Economics and Statistics*, *103*(2). https://doi.org/10.1162/rest_a_00911

Kumar, H., Somanathan, R., & Vasishth, M. (2020). *Language and Learning in Ethically Mixed Communities: A Study of School Children in an Indian Village*.

Kumar Yadav, M. (2014). Role of Mother Tongue in Second Language Learning. *International Journal of Research (IJR)*, *1*.

Kurien, J. (2005). Notes for the meeting of the national focus group on teaching of English, and note on introduction of English at the primary stage. In *Unpublished manuscript. (New Delhi: National Council of Educational Research and Training.)*.

Ladousa, C., & Davis, C. P. (2022). South Asian Language Practices: Mother Tongue, Medium, and Media. In *Annual Review of Anthropology* (Vol. 51). https://doi.org/10.1146/annurev-anthro-041420-110048

Lahoti, R., & Mukhopadhyay, R. (2019). School choice in rural India perceptions and realities in four states. *Economic and Political Weekly*, *54*(49).

Laitin, D. D., Ramachandran, R., & Walter, S. L. (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*(1). https://doi.org/10.1086/700617

Lindsay, J., & Ying Ying Tan. (2003). *Babel or behemoth: language trends in Asia: Vol. Vol. 1. NUS Press*.

Mohanty, A. K. (2006). Multilingualism of the Unequals and Predicaments of Education in India: Mother Tongue or Other Tongue? In *Imagining Multilingual Schools: Languages in Education and Glocalization*. https://doi.org/10.21832/9781853598968-014

Mohanty, A., Panda, M., & Pal, R. (2010). Language policy in education and classroom practices in India: Is the teacher a cog in the policy wheel? In *Negotiating language policies in schools:Educators as Policymakers*.

Munshi, K., & Rosenzweig, M. (2009). Why is Mobility in India so Low? Social Insurance, Inequality, and Growth. *National Bureau of Economic Research*. https://doi.org/10.3386/w14850

Muralidharan, K., & Sundararaman, V. (2015). The aggregate effectof school choice: Evidence from a two-stage experiment in India. *Quarterly Journal of Economics, 130*(3). https://doi.org/10.1093/qje/qjv013

Nag, S., Vagh, S. B., Dulay, K. M., & Snowling, M. J. (2019). Home language, school language and children's literacy attainments: A systematic review of evidence from low-and middle-income countries. *Review of Education*, *7*(1). https://doi.org/10.1002/REV3.3130

Nakamura, P., Leyew, Z., Molotsky, A., Ranjit, V., & Kamto, K. (2023). PROTOCOL: Language of instruction in schools in low- and middle-income countries: A systematic review. *Campbell Systematic Reviews*, *19*(2). https://doi.org/10.1002/cl2.1319

Nakamura, P. R., de Hoop, T., & Holla, C. U. (2019). Language and the Learning Crisis: Evidence of Transfer Threshold Mechanisms in Multilingual Reading in South India. *Journal of Development Studies*, *55*(11). https://doi.org/10.1080/00220388.2018.1493192

NCERT. (2011). *Programme evaluation report: Multilingual education in Orissa*.

Ota, A. B., Mohanty, S. C., & Mohanty, B. N. (2020). *Demographic Profile Of Scheduled Tribes Of Odisha (1961-2011)*.

Piper, B., Zuilkowski, S. S., & Ong'Ele, S. (2016). Implementing mother tongue instruction in the real world: Results from a medium-scale randomized controlled trial in Kenya. *Comparative Education Review*, *60*(4). https://doi.org/10.1086/688493

Ramachandran, R. (2017). Language use in education and human capital formation: Evidence from the Ethiopian educational reform. *World Development*, *98*. https://doi.org/10.1016/j.worlddev.2017.04.029

Ramanujam Meganathan. (2011). Language Policy in Education and the Role of English in India: From library language to language of Empowerment. *Developing Countries and the English Language*.

Reddy, P. (2011). Biliteracy development in middle-school students from the slums of India. In *Unpublished doctoral dissertation*.

Sedai, A. K., Vasudevan, R., Pena, A. A., & Miller, R. (2021). Does reliable electrification reduce gender differences? Evidence from India. *Journal of Economic Behavior and Organization*, *185*. https://doi.org/10.1016/j.jebo.2021.03.015

Seid, Y. (2019). The impact of learning first in mother tongue: evidence from a natural experiment in Ethiopia. *Applied Economics*, *51*(6). https://doi.org/10.1080/00036846.2018.1497852

Seid, Y. (2022). Mother-tongue instruction and later labour market outcomes: evidence from a natural experiment in Ethiopia. *Applied Economics*. https://doi.org/10.1080/00036846.2022.2083067

Shah, M., & Steinberg, B. M. (2017). Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. *Journal of Political Economy, 125*(2), 527–561.

Shastry, G. K. (2012). Human capital response to globalization: Education and information technology in India. *Journal of Human Resources*, *47*(2). https://doi.org/10.3368/jhr.47.2.287

Skutnabb-Kangas, T. (2013). Linguistic genocide in education- or worldwide diversity and human rights? In *Linguistic Genocide in Education- or Worldwide Diversity and Human Rights?* https://doi.org/10.4324/9781410605191

Sonaiya, R. (2002). Autonomous language learning in africa: A mismatch of cultural assumptions. *Language, Culture and Curriculum*, *15*(2). https://doi.org/10.1080/07908310208666637

Taylor, S., & von Fintel, M. (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review*, *50*. https://doi.org/10.1016/j.econedurev.2016.01.003

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.2.4.1

Trudell, B. (2005). Language choice, education and community identity. *International Journal of Educational Development*, *25*(3). https://doi.org/10.1016/j.ijedudev.2004.08.004

UNESCO. (2016). If you don't understand, how can you learn? Global Monitoring Report. In *Policy Paper 24* (Issue February).

UNESCO Institute for Statistics. (2017). More Than One-Half of Children and Adolescents Are Not Learning Worldwide. In *UIS Fact Sheet No. 46*.

Vagh, S. B. (2012). Validating the ASER testing tools: Comparisons with reading fluency measures and the Read India measures. In *Unpublished report*.

Vaish, V. (2008). Biliteracy and globalization: English language education in India. In *Biliteracy and Globalization: English Language Education in India*. https://doi.org/10.1080/07908310902789677

Van Wyk, J., & Mostert, M. L. (2016). The influence of mother tongue and gender on the acquisition of english (L2). The case of afrikaans in windhoek schools, Namibia. *Cogent Education*, *3*(1). https://doi.org/10.1080/2331186X.2016.1210997

Venkatraman, V., Soon, C. S., Chee, M. W. L., & Ansari, D. (2006). Effect of language switching on arithmetic: A bilingual fMRI study. *Journal of Cognitive Neuroscience*, *18*(1). https://doi.org/10.1162/089892906775250030

World Bank. (2021). *Loud and Clear: Effective Language of Instruction Policies for Learning*.

**Tables and Figures**

**Tables and Figures**

**Figure 1. Proportion of Schools Teaching in Dominant Language of State**

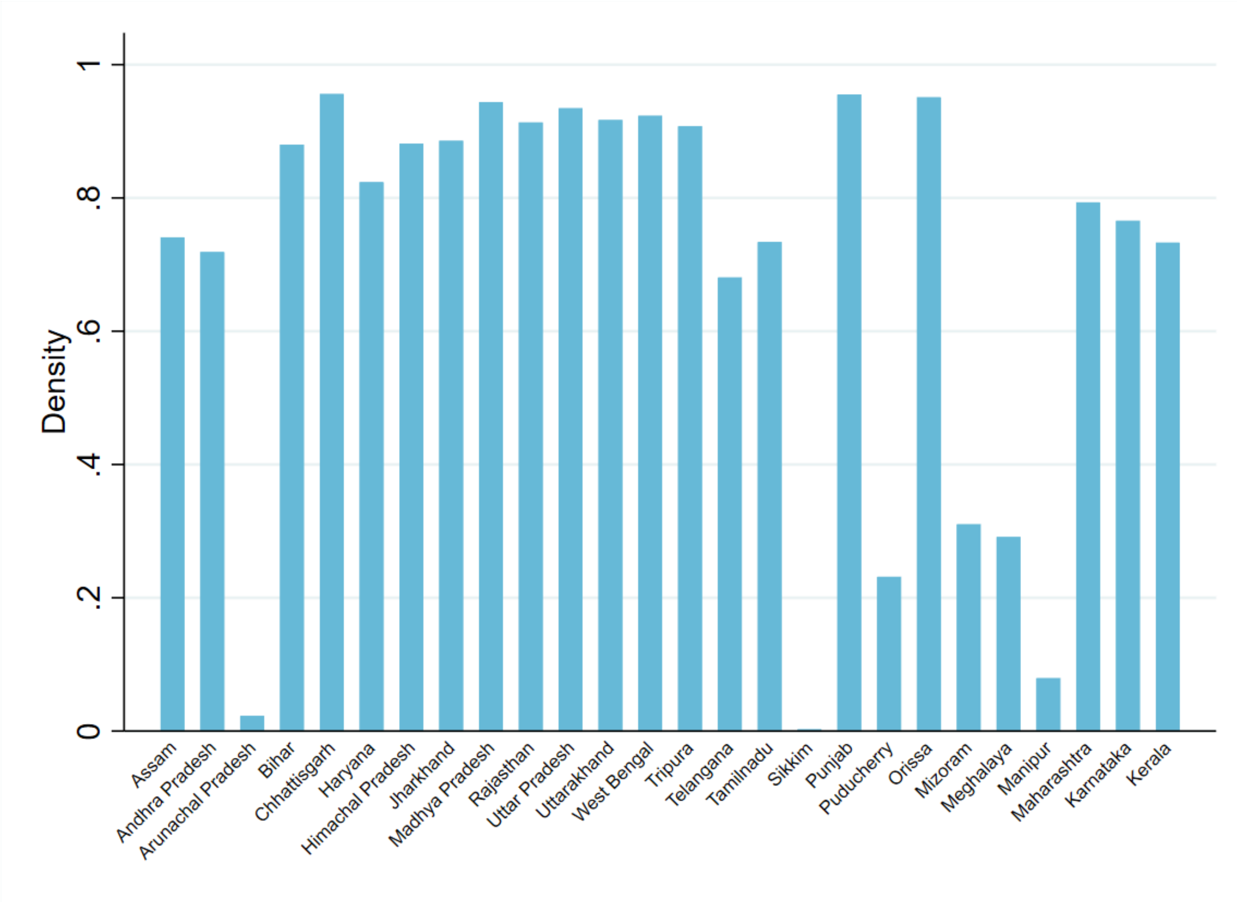

**Table 1. Scheduled Languages of India**

| Language | Family | Speakers in Millions (2011 Census) | Official Recognition by State |
|---|---|---|---|
| Assamese | Indo-Aryan, Eastern | 15.3 | Assam |
| Bengali | Indo-Aryan, Eastern | 97.2 | West Bengal, Tripura, Assam, and Jharkhand |
| Bodo | Tibeto-Burman | 1.48 | Assam |
| Dogri | Indo-Aryan, Northwestern | 2.6 | Jammu and Kashmir |
| Gujarati | Indo-Aryan, Western | 55.5 | Dadra and Nagar Haveli, Daman and Diu, Gujarat |
| Hindi | Indo-Aryan | 528 | Andaman and Nicobar Islands, Bihar, Dadra and Nagar Haveli, Daman and Diu, Chhattisgarh, Delhi, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Mizoram, Rajasthan, Uttar Pradesh, Uttarakhand, West Bengal, Jammu and Kashmir, Ladakh |
| Kannada | Dravidian | 43.7 | Karnataka |
| Kashmiri | Indo-Aryan, Dardic | 6.8 | Jammu and Kashmir |
| Konkani | Indo-Aryan, Southern | 2.25 | Dadra and Nagar Haveli, Daman and Diu, Goa |
| Maithili | Indo-Aryan, Eastern | 13.6 | Jharkhand |
| Malayalam | Dravidian | 34.8 | Kerala, Lakshadweep, and Puducherry |
| Manipuri | Tibeto-Burman | 1.8 | Manipur |
| Marathi | Indo-Aryan, Southern | 83 | Maharashtra |
| Nepali | Indo-Aryan, Northern | 2.9 | Sikkim, West Bengal |
| Odia | Indo-Aryan, Eastern | 37.5 | Odisha, Jharkhand, and West Bengal |
| Punjabi | Indo-Aryan, Northwestern | 33.1 | Delhi, Haryana, Punjab, and West Bengal |
| Sanskrit | Indo-Aryan | 0.02 | Himachal Pradesh, Uttarakhand |
| Santhali | Austroasiatic | 7.3 | Jharkhand |
| Sindhi | Indo-Aryan, Northwestern | 2.7 | Not official in any state |
| Tamil | Dravidian | 69 | Tamilnadu, Puducherry |
| Telugu | Dravidian | 81.1 | Andhra Pradesh, Telangana, Puducherry |
| Urdu | Indo-Aryan, Central | 50.7 | Jammu and Kashmir, Telangana, Jharkhand, Delhi, Bihar, Uttar Pradesh, and West Bengal |

Source: https://www.asianstudies.org/publications/eaa/archives/multilingualism-in-india/

**Table 2. Dominant Testing Language and Most Spoken Language**

| State/Union Territory | Most Dominant Testing Language per 2018 ASER | Most Spoken Language per Census 2011 |
|---|---|---|
| Andhra Pradesh | Telugu | Telugu |
| Arunachal Pradesh | English | Nissi |
| Assam | Assamese | Assamese |
| Bihar | Hindi | Hindi |
| Chhattisgarh | Hindi | Hindi |
| Dadra and Nagar Haveli | Marathi | Gujarati |
| Daman and Diu | Gujarati | Gujarati |
| Goa | English | Konkani |
| Gujarat | Gujarati | Gujarati |
| Haryana | Hindi | Hindi |
| Himachal Pradesh | Hindi | Hindi |
| Jammu and Kashmir | English | Kashmiri |
| Jharkhand | Hindi | Hindi |
| Karnataka | Kannada | Kannada |
| Kerala | Malayalam | Malayalam |
| Madhya Pradesh | Hindi | Hindi |
| Maharashtra | Marathi | Marathi |
| Manipur | English | Manipuri |
| Meghalaya | English | Khasi |
| Mizoram | Mizo | Mizo |
| Nagaland | English | Naga/English |
| Orissa | Oriya | Odia |
| Puducherry | Tamil | Tamil |
| Punjab | Punjabi | Punjabi |
| Rajasthan | Hindi | Hindi |
| Sikkim | English | Nepali |
| Tamilnadu | Tamil | Tamil |
| Telangana | English | Telugu |
| Tripura | Bengali | Bengali |
| Uttar Pradesh | Hindi | Hindi |
| Uttarakhand | Hindi | Hindi |
| West Bengal | Bengali | Bengali |

Notes: Most dominant testing language is defined as at least 50 percent of the 2018 ASER sample tests in that language.

**Table 3. Distribution of Languages as Mother Tongue and Medium of Instruction (ASER 2018)**

| Language | % Mother Tongue | % Medium of Instruction |
|---|---|---|
| Hindi | 53.27 | 44.63 |
| Telugu | 1.22 | 1.1 |
| Bengali | 3.04 | 2.97 |
| Marathi | 4.77 | 4.26 |
| Tamil | 4.46 | 2.76 |
| Urdu | 0.04 | 0.48 |
| Gujarati | 3.62 | 3.56 |
| Kannada | 6.34 | 4.8 |
| Malayalam | 0.60 | 0.34 |
| Oriya | 3.99 | 3.93 |
| Punjabi | 2.44 | 1.17 |
| Assamese | 2.62 | 2.68 |
| Manipuri | 0.00 | 0.01 |
| Nepali | 0.00 | 0.02 |
| English | 11.88 | 25.63 |
| Garo | 0.28 | 0.28 |
| Khasi | 0.34 | 0.31 |
| Mizo | 0.98 | 0.96 |
| Bodo | 0.09 | 0.12 |

Source: ASER 2018

**Table 4. Proportion of Sample with Medium of Instruction = Mother Tongue (ASER 2018)**

| State/Union Territory | Medium of Instruction = Mother Tongue |
|---|---|
| Andhra Pradesh | 0.90 |
| Arunachal Pradesh | 0.97 |
| Assam | 0.97 |
| Bihar | 0.85 |
| Chhattisgarh | 0.98 |
| Dadra and Nagar Haveli | 0.23 |
| Daman and Diu | 0.73 |
| Goa | 0.90 |
| Gujarat | 0.97 |
| Haryana | 0.49 |
| Himachal Pradesh | 0.59 |
| Jammu and Kashmir | 0.98 |
| Jharkhand | 0.88 |
| Karnataka | 0.75 |
| Kerala | 0.70 |
| Madhya Pradesh | 0.98 |
| Maharashtra | 0.88 |
| Manipur | 0.99 |
| Meghalaya | 0.79 |
| Mizoram | 0.92 |
| Nagaland | 1.00 |
| Orissa | 0.98 |
| Puducherry | 0.28 |
| Punjab | 0.47 |
| Rajasthan | 0.93 |
| Sikkim | 0.96 |
| Tamilnadu | 0.63 |
| Telangana | 0.92 |
| Tripura | 0.88 |
| Uttar Pradesh | 0.80 |
| Uttarakhand | 0.74 |
| West Bengal | 0.98 |
| Overall | 0.84 |
| **Number of Observations** | **362,030** |

Source: ASER 2018

**Table 5. Descriptive Statistics for 2018 ASER Sample**

| | Medium of instruction = Mother tongue | Medium of instruction ≠ Mother tongue | p-value |
|---|---|---|---|
| | [1] | [2] | [3] |
| ***Academic Performance*** | | | |
| Reading score (out of 4) | 2.59 | 2.70 | 0.22 |
| Math score (out of 4) | 2.26 | 2.55 | 0.00 |
| ***Household and Individual Controls*** | | | |
| Male (0/1) | 0.49 | 0.55 | 0.05 |
| Household size | 5.77 | 5.73 | 0.67 |
| Mother attended school (0/1) | 0.55 | 0.81 | 0.00 |
| Father attended school (0/1) | 0.70 | 0.85 | 0.00 |
| Child takes paid tuition (0/1) | | | |
| Household has electricity connection (0/1) | 0.89 | 0.96 | 0.00 |
| Household gets newspaper (0/1) | 0.06 | 0.19 | 0.00 |
| Household has other reading material (0/1) | 0.07 | 0.10 | 0.02 |
| Household owns bike/moped (0/1) | 0.37 | 0.66 | 0.00 |
| Household owns car (0/1) | 0.07 | 0.20 | 0.00 |
| Household owns television (0/1) | 0.57 | 0.83 | 0.00 |
| Household owns computer (0/1) | 0.16 | 0.37 | 0.00 |
| Village has a pucca road (0/1) | 0.84 | 0.92 | 0.00 |
| Village has electricity (0/1) | 0.98 | 0.99 | 0.05 |
| Village has bank (0/1) | 0.28 | 0.40 | 0.00 |
| Village has internet café (0/1) | 0.21 | 0.28 | 0.21 |
| Village has government primary school (0/1) | 0.93 | 0.93 | 0.98 |
| Village has government middle school (0/1) | 0.65 | 0.68 | 0.41 |
| Village has government secondary school (0/1) | 0.17 | 0.19 | 0.30 |
| Village has private school (0/1) | 0.41 | 0.52 | 0.00 |
| Village has anganwadi (0/1) | 0.93 | 0.93 | 0.77 |
| ***District Controls*** | | | |
| Proportion of schools approachable by all-weather road | 0.84 | 0.81 | 0.10 |
| Average working hours for teachers | 4.96 | 4.99 | 0.61 |
| Proportion of minority managed schools | 0.02 | 0.03 | 0.01 |
| Average proportion of male teachers | 0.61 | 0.52 | 0.00 |
| Average proportion of female teachers | 0.39 | 0.48 | 0.00 |
| Average proportion of teachers with a graduate degree | 0.41 | 0.39 | 0.00 |
| **Number of Observations** | **306709** | **55321** | |

Notes:
[1] Sample is restricted to children 5 - 16 years
[2] Mother tongue is proxied is proxied by test language on the day of the survey
[3] Source: Annual Status of Education Report 2018
[4] For the student's t-test in column 3 standard errors are computed assuming correlation of individual observations over time within each district. Number of districts = 592

**Table 6. Baseline Balance Test for Odisha Case Study**

| | Treatment districts | Comparison districts | p-value |
|---|---|---|---|
| | [1] | [2] | [3] |
| ***Academic Performance*** | | | |
| Reading score (out of 4) | 2.34 | 2.79 | 0.41 |
| Math score (out of 4) | 2.12 | 2.55 | 0.42 |
| ***Household and Individual Controls*** | | | |
| Male (0/1) | 0.53 | 0.52 | 0.94 |
| Household size | 5.72 | 5.84 | 0.85 |
| Mother attended school (0/1) | 0.55 | 0.81 | 0.52 |
| Child age (years) | 9.93 | 10.15 | 0.86 |
| **Number of Observations** | **15338** | **13545** | |

Notes:

[1] Sample is restricted to children 5- 16 years in Odisha in the 2007 round of the ASER.

[2] For the student's t-test in column 3 standard errors are computed assuming correlation of individual observations over time within each district. Number of districts = 30

**Table 7. OLS Results Reading and Math Outcomes by Gender**

| | Reading Score | | | Math Score | | |
|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] |
| *Panel A: Total* | | | | | | |
| **Medium of instruction same as mother tongue** | 0.062*** | 0.299*** | 0.306*** | -0.110*** | 0.117*** | 0.133*** |
| | (0.020) | (0.020) | (0.021) | (0.015) | (0.015) | (0.015) |
| Number of observations | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 |
| *Panel B: Females* | | | | | | |
| **Medium of Instruction same as mother tongue** | 0.065*** | 0.312*** | 0.318*** | -0.116*** | 0.125*** | 0.138*** |
| | (0.022) | (0.024) | (0.024) | (0.017) | (0.018) | (0.018) |
| Number of observations | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 |
| *Panel C: Males* | | | | | | |
| **Medium of Instruction same as mother tongue** | 0.045** | 0.284*** | 0.291*** | -0.103*** | 0.106*** | 0.125*** |
| | (0.021) | (0.020) | (0.022) | (0.016) | (0.016) | (0.016) |
| Number of observations | 181,284 | 181,284 | 181,284 | 181,284 | 181,284 | 181,284 |
| No controls | X | | | X | | |
| ASER household and village controls | | X | X | | X | X |
| UDISE+ district controls | | | X | | | X |
| District fixed effects | X | X | X | X | X | X |

Notes: The independent variable is an indicator if the medium of instruction is the same as the mother tongue. Columns 1 and 4 do not include any controls, except district fixed effects. Columns 2 and 5 include household and village controls including district fixed effects, indicator for male gender, household size, indicator variables for a child's mother and father attended school, indicator variable for a house having an electricity connection, household having access to a newspaper, household having access to reading materials other than newspapers, household has a scooter, household has a car, household has a television, and household has a computer. Village level controls include indicator variables if a village has a pucca road, village has access to electricity, village has a bank, village has an internet café, village has a government primary school, village has a government middle school, village has a government secondary school, and village has a private school. Finally, columns 3 and 6 include district level controls including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools, average proportion of male teachers, average proportion of female teachers, and average proportion of teachers with a graduate degree. Standard errors are in parentheses, are clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

**Table 8. IV Results Reading and Math Outcomes by Gender**

| | Reading Score | | | Math Score | | |
|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] |
| *Panel A: Total* | | | | | | |
| **Medium of instruction same as mother tongue** | 0.577** | 0.386** | 0.310* | 0.499* | 0.646*** | 0.469** |
| | (0.285) | (0.195) | (0.163) | (0.259) | (0.217) | (0.188) |
| Number of observations | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 |
| *Panel B: Females* | | | | | | |
| **Medium of Instruction same as mother tongue** | 0.864*** | 0.468** | 0.327* | 1.045*** | 1.075*** | 0.859*** |
| | (0.360) | (0.216) | (0.188) | (0.355) | (0.244) | (0.243) |
| Number of observations | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 |
| *Panel C: Males* | | | | | | |
| **Medium of Instruction same as mother tongue** | 0.336 | 0.314* | 0.350** | 0.084 | 0.288 | 0.340* |
| | (0.242) | (0.190) | (0.160) | (0.207) | (0.212) | (0.193) |
| Number of observations | 181,284 | 181,284 | 181,284 | 181,284 | 181,284 | 181,284 |
| No controls | X | | | X | | |
| ASER household and village controls | | X | X | | X | X |
| UDISE+ district controls | | | X | | | X |

Notes: The independent variable is an indicator if the medium of instruction is the same as the mother tongue. Columns 1 and 4 do not include any controls, except district fixed effects. Columns 2 and 5 include household and village controls including district fixed effects, indicator for male gender, household size, indicator variables for a child's mother and father attended school, indicator variable for a house having an electricity connection, household having access to a newspaper, household having access to reading materials other than newspapers, household has a scooter, household has a car, household has a television, and household has a computer. Village level controls include indicator variables if a village has a pucca road, village has access to electricity, village has a bank, village has an internet café, village has a government primary school, village has a government middle school, village has a government secondary school, and village has a private school. Finally, columns 3 and 6 include district level controls including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools, average proportion of male teachers, average proportion of female teachers, and average proportion of teachers with a graduate degree. Standard errors are in parentheses, are clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

**Table 9. IV Results Reading and Math Outcomes by Age**

| | Reading Score | | | Math Score | | |
|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] |
| *Panel A: 5 - 10 years* | | | | | | |
| **Medium of instruction same as mother tongue** | 0.850*** | 0.711*** | 0.658*** | 0.804*** | 0.971*** | 0.866*** |
| | (0.284) | (0.202) | (0.182) | (0.251) | (0.201) | (0.193) |
| Number of observations | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 | 362,030 |
| *Panel B: 11 - 16 years* | | | | | | |
| **Medium of Instruction same as mother tongue** | -0.019 | -0.340 | -0.311 | -0.122 | -0.045 | 0.011 |
| | (0.231) | (0.252) | (0.217) | (0.261) | (0.301) | (0.269) |
| Number of observations | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 | 180,746 |
| No controls | X | | | X | | |
| ASER household and village controls | | X | X | | X | X |
| UDISE+ district controls | | | X | | | X |

Notes: The independent variable is an indicator if the medium of instruction is the same as the mother tongue. Columns 1 and 4 do not include any controls, except district fixed effects. Columns 2 and 5 include household and village controls including district fixed effects, indicator for male gender, household size, indicator variables for a child's mother and father attended school, indicator variable for a house having an electricity connection, household having access to a newspaper, household having access to reading materials other than newspapers, household has a scooter, household has a car, household has a television, and household has a computer. Village level controls include indicator variables if a village has a pucca road, village has access to electricity, village has a bank, village has an internet café, village has a government primary school, village has a government middle school, village has a government secondary school, and village has a private school. Finally, columns 3 and 6 include district level controls including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools, average proportion of male teachers, average proportion of female teachers, average proportion of regular teachers, average proportion of teachers with a graduate degree and average proportion of teachers with a bachelor in education degree. Standard errors are in parentheses, are clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

**Table 10. Difference-in-Differences Results from Odisha**

| | **Reading Score** | | **Math Score** | | **Probability of drop out** | **Probability of on track** |
|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] |
| *Panel A: Total* | | | | | | |
| **Post X Treatment** | 0.157** | 0.193*** | 0.184* | 0.211** | 0.001 | 0.024** |
| | (0.078) | (0.073) | (0.103) | (0.099) | (0.011) | (0.010) |
| Number of observations | 45,196 | 45,196 | 45,196 | 45,196 | 45,196 | 45,196 |
| *Panel B: Females* | | | | | | |
| **Post X Treatment** | 0.148 | 0.185** | 0.192* | 0.222** | -0.008 | 0.026*** |
| | (0.098 | (0.094) | (0.116) | (0.113) | (0.012) | (0.010) |
| Number of observations | 21,738 | 21,738 | 21,738 | 21,738 | 21,738 | 21,738 |
| *Panel C: Males* | | | | | | |
| **Post X Treatment** | 0.166** | 0.202*** | 0.177* | 0.202** | 0.010 | 0.022* |
| | (0.071) | (0.063) | (0.100) | (0.093) | (0.012) | (0.011) |
| Number of observations | 23,458 | 23,458 | 23,458 | 23,458 | 23,458 | 23,458 |
| No controls | X | | X | | | |
| ASER controls | | X | | X | X | X |

Notes: The independent variable is the coefficient on the interaction term between *Post* and *Treatment* districts. Columns 1 and 4 do not include any controls, except district fixed effects. Columns 2 and 5 include ASER controls including district fixed effects, indicator for male gender, household size, and indicator variables for a child's mother having attended school, village having a road, village having a school, and village having a ration shop. Finally, columns 3 and 6 include district level controls including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools, average proportion of male teachers, average proportion of female teachers, average proportion of regular teachers, average proportion of teachers with a graduate degree and average proportion of teachers with a bachelor in education degree. Standard errors are in parentheses, are clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

## Appendix Tables and Figures

**Appendix Table 1. Descriptive Statistics for 2021 NAS - SECC Sample**

| | N | Mean | SD |
|---|---|---|---|
| | [1] | [2] | [3] |
| ***Academic Performance - District Level*** | | | |
| Language score (%) | 968 | 58.99 | 8.645 |
| Math score (%) | 968 | 50.68 | 10.1 |
| ***District Variables (NAS)*** | | | |
| Student - Likes to go to school | 968 | 97.72 | 1.922 |
| Student - Home language same as medium of instruction | 968 | 83.04 | 8.803 |
| Student - Understand teachers | 968 | 96.53 | 2.195 |
| Student - Go out and play | 968 | 75.05 | 7.553 |
| Student - Have access to digital device | 968 | 68.93 | 14.62 |
| Student - Has internet | 968 | 52.21 | 11.83 |
| Student - Has parental support | 968 | 82.19 | 8.175 |
| Teacher - Have adequate teaching learning materials | 968 | 38.07 | 15.86 |
| Teacher - Have adequate workspace | 968 | 55.55 | 17.16 |
| Teacher - Overloaded with work | 968 | 35.73 | 13.48 |
| Teacher - Building needs repair | 968 | 25.24 | 12.62 |
| Teacher - Lack of drinking water | 968 | 15.48 | 12.08 |
| Teacher - Inadequate toilet facilities | 968 | 15.82 | 11.44 |
| Teacher - Participated in professional development | 968 | 55.74 | 14.61 |
| Teacher - Parents take interest | 968 | 89.09 | 7.506 |
| Teacher - Know COVID reporting protocol | 968 | 97.72 | 2.942 |
| Teacher - Know of wellbeing of children | 968 | 98.04 | 2.475 |
| Teacher - Aware of school reopen guidelines | 968 | 97.7 | 3.443 |
| Principal - Have adequate qualified staff | 968 | 78.41 | 12.93 |
| Principal - Have adequate support staff | 968 | 55.45 | 13.16 |
| Principal - Have adequate audio-visual equipment | 968 | 42.94 | 20.2 |
| Principal - Have adequate library | 968 | 50.53 | 19.28 |
| Principal - Participate in sports | 968 | 96.84 | 2.896 |
| Principal - School has library | 968 | 87.47 | 13.57 |
| ***District Variables (SECC)*** | | | |
| Share of households - highest earning member monthly income greater than 10,000 INR | 968 | 0.0919 | 0.116 |
| Share of individuals who have completed primary school or above | 968 | 0.493 | 0.202 |
| Scheduled caste share | 968 | 0.0547 | 0.0918 |
| Share of scheduled tribe population | 968 | 0.0672 | 0.117 |
| Percentage of households that own land | 968 | 0.534 | 0.286 |

Notes:

[1] Sample is restricted to children in Grades 3 and 5

[2] Source: National Achievement Survey 2021 and SECC 2012

## Appendix Table 2. NAS District Results for Reading and Math Outcomes

| | Language Score | | | Math Score | | |
|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] |
| *Panel A: Total* | | | | | | |
| **Proportion of students at district level with medium of instruction same as mother tongue** | 0.139*** | 0.137*** | 0.111** | 0.299*** | 0.257*** | 0.210*** |
| | (0.038) | (0.034) | (0.035) | (0.042) | (0.039) | (0.042) |
| Number of observations | 968 | 968 | 968 | 968 | 968 | 968 |
| *Panel B: Grade 3* | | | | | | |
| **Proportion of students at district level with medium of instruction same as mother tongue** | 0.049 | 0.082 | 0.068 | 0.099* | 0.119** | 0.092* |
| | (0.049) | (0.043) | (0.045) | (0.046) | (0.039) | (0.041) |
| Number of observations | 484 | 484 | 484 | 484 | 484 | 484 |
| *Panel C: Grade 5* | | | | | | |
| **Proportion of students at district level with medium of instruction same as mother tongue** | 0.0326 | 0.0404 | 0.0219 | 0.105** | 0.0912* | 0.0447 |
| | (0.039) | (0.037) | (0.041) | (0.039) | (0.036) | (0.042) |
| Number of observations | 484 | 484 | 484 | 484 | 484 | 484 |
| No controls | X | | | X | | |
| NAS district controls - Student, Teacher, and Headteacher | | X | X | | X | X |
| SECC district controls | | | X | | | X |
| State fixed effects | X | X | X | X | X | X |

Notes: The independent variable is an indicator showing the district-level percentage of children whose medium of instruction is the same as the mother tongue. Columns 1 and 4 do not include any controls, except state fixed effects. Columns 2 and 5 include student, teacher, and headteacher survey responses aggregated at the district level as controls, including district fixed effects. Columns 3 and 6 include controls from the Socio-Economic and Caste Census. These controls are all district-level shares - SC households, ST households, households with monthly income greater than 10,000 INR, share of individuals with at least a primary-level education, and households which own land. Standard errors are in parentheses, are clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

**Appendix Table 3. IV Estimates on Attendance**

| | School Attendance Rate (Grades I-V) | School Attendance Rate (Grades VI-VIII) | Teacher Attendance Rate |
|---|---|---|---|
| | [1] | [2] | [3] |
| **Proportion of students in village with mother tongue instruction** | 0.497*** | 0.466*** | -0.048 |
| | (0.088) | (0.112) | (0.091) |
| N | 17,731 | 17,731 | 17,731 |

Notes: The unit of observation in all columns is at the school level. All columns include village level controls including indicator variables if a village has a pucca road, village has access to electricity, village has a bank, village has an internet café, village has a government primary school, village has a government middle school, village has a government secondary school, and village has a private school and district level controls including percent of schools approachable by an all-weather road, average working hours for teachers, proportion of minority managed schools, average proportion of male teachers, average proportion of female teachers, average proportion of regular teachers, average proportion of teachers with a graduate degree and average proportion of teachers with a bachelor in education degree. Standard errors are in parentheses, clustered by district. *Significant at 10%. **Significant at 5%. ***Significant at 1%.