Understanding the dynamics of home language instruction in shaping children's foundational learning: Evidence from rural India

Shreedaya Arvind[*], Deepak Kumar[†], Rukmini Banerji,[‡], Wilima Wadhwa[§]

September 2024

Abstract

Evidence suggests that children's learning development during the foundational stage can be hampered when the language spoken at home differs from the instruction language used in school. The recent National Education Policy 2020 in India recommends teaching children in their home language in primary schools. In this context, this study examines the role of home language instruction (advantage) on primary school student's foundational learning in rural India by using the Annual Status of Education Report (ASER) 2022 dataset. The ASER 2022 dataset provides unique additional information on children's language spoken at home and medium of instruction in school which make it possible to match whether a child was instructed in his/her home language or not. Overall, we find that teaching primary school student's in their home language has an advantage in achieving their foundational learning, particularly in different levels of literacy development. There is no evidence of home language instruction advantage in Hindi medium schools, predominant in the Hindi belt. However, we find a strong advantage of home language instruction in the single regional languages states after restricting our sample to these state language (excluding English and Hindi) medium schools. Our study has broader policy implications and highlights a larger need for mother tongue-based education, especially in linguistically diverse regions of the country.

Keywords: Learning Outcomes; Home Language Instruction; Primary Schools; Foundational Literacy and Numeracy Skills; Multilingual Education; India
JEL: I20, I21, O12

[*]Research Intern, ASER Centre, New Delhi
[†]Senior Research Associate, Research Unit, ASER Centre, New Delhi
[‡]Chief Executive Officer, Pratham Education Foundation
[§]Director, ASER Centre

1 Introduction

Attaining the foundational literacy and numeracy (FLN) skills is essential for ensuring effective learning throughout a student's educational journey. This has become a central focus in policy discussions on primary education in developing countries, many of which have made significant progress in expanding access to schooling but continue to grapple with challenges of low learning levels (World Development Report, 2018). For instance, India, the context of this study, has achieved near universal access to primary education, however, children's FLN skills are still low – more than half the students in fifth grade could not read a text at the grade 2 level and more than 55% of them could not solve a two-digit subtraction problem (ASER Report, 2022). The language of instruction in school, while teaching during the student's foundational stage, plays an important role in determining their learning outcomes (Jhingran, 2009; Jain, 2017; Erikkson, 2014). Literature shows that teaching children in their home language during primary education significantly leads to better learning outcomes (Seid, 2019; Africa, Taylor & Von Fintel, 2016). The recent National Education Policy (NEP) 2020 in India also recommends teaching children in their home language in primary schools. Therefore, this study examines the role of home language instruction (advantage), which is the medium of instruction in school, on primary school student's foundational learning in rural India.

Language is a crucial instrument of learning; though learning without language is possible with sole reliance on biological and social indicators of behaviour, language eases learning by presenting a medium for humans to convey ideas (Boroditsky, 2011). In education, this effect is magnified, and facilitates thinking, reading, writing, speaking, and listening. The effect of teaching language in primary education is more pronounced because primary education coincides with a crucial phase of cognitive and socio-emotional development in the child. A home language learning advantage is instrumental in the acquisition of foundational skills (Scheele et al., 2010; Karlsen et al., 2017) and some components of higher-order literacy that affect reading comprehension (Lesaux et al., 2006; Vagh, 2018). The driving argument here is that with home language instruction, there is continuous engagement with reading and communication in one language, and resultant enrichment experiences (or lack thereof) affects the creation of knowledge and by consequence, acquisition of reading skills. Home language instruction also allows familial and cultural forces to better contribute towards the child's literacy development with frequent communication, presence of reading materials, and inclusion of oral culture.

India is highly diverse and home to more than 700 languages across different states. Among these languages, only 22 languages are granted an 'official' status according to the Eighth Schedule of the Indian Constitution (Article 344(1) & 351; Nambissan 1994). Several languages possess dialects, which are regional variations of a language spoken within a broader geographical region. Indian languages also differ in terms of their orthography, which is a crucial deter-

minant of learning. All Indian languages can be broadly categorised into the Indo-Aryan, Dravidian, Austro-Asiatic, or Sino-Tibetan family (Ethnologue, 2019). Classrooms in India mirror this diversity, with several children often possessing bilingual or multilingual capabilities by the time they enter formal classrooms (NCERT, 2006). This diversity stems from the urban areas of India which are hotspots of diversity, attracting people from various corners of the country. Even in rural India, many regions have different home languages and medium of instruction in school, particularly in tribal regions (Nambissan, 1994). Second, diversity is created from language and dialect variation. There is ongoing debate on whether these language systems constitute a dialect or a language of its own (e.g., is Chhattisgarhi a dialect of Hindi or a language of its own). Unfortunately, language-in-education policies have not caught up with the extent of diversity in India's classrooms.

After Independence, the States Reorganisation Act of 1956 reorganised states on the basis of languages spoken. Although newer states have been created since, these separations also developed on the basis of language differences. In light of massive language differences and a strong opposition to elect Hindi as the country's national language, education was made a subject of the concurrent list. This gives each state the autonomy to frame and implement their own education policies, based on guidelines received from the central government. This resulted in varied policies across the country. Recently, NEP 2020 mandates a three-language formula. In primary school, students need to be instructed in their primary language and from upper primary they need to be taught another modern Indian language along with instruction in English. Since education policy is a subject of the concurrent list, many states do not implement the three-language recommendation and opted for language policies that suited language politics within their state (Nambissan, 1994). Political phenomena, especially at regional levels such as the state or district, are largely influenced by linguistic groups and the power they hold within their states. Constitutionally-recognised speakers are more politically powerful and minorities such as tribes who speak tribal languages lack the political lustre necessary to bring out development within their community, such as setting up tribal schools. Some of these tongues remain alienated, which often seeps into the classroom. Further, alienation and the inability to understand the school language in tribal students results in adverse learning outcomes.

In this context, there is a clear need to better understand the match (advantage) and mismatch between a student's home language and medium of instruction in primary schools, and how it affects the children's FLN skills across different regions in rural India. Here, we also examine how the role of home language instruction (advantage) on children learning changes across different official language dominant regions, such as Hindi language region and other state languages, so that effective targeted policies can be made where it matters the most. In our knowledge, this is the first study that examines this issue at rural India level by using the Annual Status of Education Report

(ASER) 2022 survey. The ASER dataset provides the information on children's enrolment and learning outcomes over time which is representative at the district level. However, the ASER 2022 survey has also collected information on children's home language in addition to their test language and medium of instruction in school. Therefore, this study contributes in the area of the role of home language instruction on children's learning outcomes during the foundational stage in developing countries like India. Our findings have broader policy implications based on the socio-cultural and political context of a region.

2 Language Dynamics in India

Literature in the field of language policy and learning and its returns is in no means no scant. A wide array of literature studies in empirical and experimental methods reveals evidence of language influences on educational outcomes, wage, workplace success, and social networks, etc. (Azim et al., 2010; Chakraborty & Kaur, 2016). In India, these relationships are more nuanced owing to India's sheer linguistic diversity, haphazard economic growth, and caste and class dynamics. Hubs of economic growth, such as metropolitan cities, economic zones feature the use of more politically powerful languages such as English or Hindi (Jhingran, 2019; Meghanathan, 2017). But tucked away in remote corners are sparsely spoken languages and dialects (Jhingran, 2009). Caste and class dynamics also affect the language spoken; many hidden dialects are spoken by various tribal groups of India. These languages lack the social capital necessary for the progress of both the language and the group, and existing language hierarchies have continued to oppress speakers of such languages (Jhingran, 2009).

Existing literature is almost entirely concerned with the role of language in educational outcomes and/or its returns. For instance, Chakraborty and Kaur (2016) documented the returns of English-language skills and found that individuals with more access to public schools that taught in English-medium had better labour market outcomes. Azim et al. (2010) echoes these results, with growing English-language premiums especially in urban, cosmopolitan India. Jain (2017) exploits the reorganisation of states along linguistic lines as a natural experiment to study the role of language-in-education policies on educational attainment and its returns. The results showed that having a common language in a district and state reduced costs of communication, and reorganisation along language lines improved growth rates in educational attainment by 46.8% in previously language minority districts. The author also identified that though the transition to mother-tongue based instruction and organisation in minority districts resulted in growth, the transformation was slow. Only by 1991 did minority districts catch up to the majority. Another study by Khatri (2023) explores a similar question without the natural experiment using a district-level panel of ASER 2005-2007 data. Using district-level fixed effects and instrumental variable estimation, the authors identified significant negative effects of linguistic heterogeneity on reading outcomes but no effect on

arithmetic outcomes.

The language hierarchy in India has a tricky history. After Independence, Indian leaders were determined to shed colonial identities by abandoning English and embracing Hindi as the national language and a linguistic principle guiding state re-organisation. But this decision was met with massive opposition, especially from southern Indian states and accelerated into the Dravidian movement. The government abandoned that decision and in an attempt to remain inclusive, drafted the Eighth Schedule of the Indian Constitution that recognised dominantly spoken languages in India in articles 344(1) and 351. This list changed over the years and now consists of the following languages: Hindi, Gujarati, Punjabi, Marathi, Kannada, Malayalam, Tamil, Telugu, Odia, Bengali, Mizo, Assamese, Nepali, Bhojpuri, Bodo, Kashmiri, Konkani, Manipuri, Sanskrit, Santhali, Sindhi, and Maithili (Ramachandran, 2015)[1]. English serves as the official language of administration. However, the population census of 2011 recognises 122 different languages in the country and research by the People's Linguistic Survey of India and Ethnologue (2019) shows that India is home to more than 700 languages. But only a negligible fraction receives 'recognition'. This has created a hierarchy, in a colloquial and political sense. English, Hindi and typically the predominant regional language of a state are associated with better economic returns. Parents encourage their children to learn Hindi or English instead of their tribal mother tongue because it is seen as a step in the ladder of upliftment (World Development Report, 2017; Jhingran, 2019; Meghanathan, 2017). But these tribal students are unable to comprehend these languages, which are so distinct from their home languages, resulting in a marked disadvantage in their learning. Languages like English and Hindi sit atop the Indian language hierarchy, then followed by a state's dominant regional language if any.

3 Data & Methodology

3.1 Data

To investigate this issue, this study employs the latest 2022 survey dataset of the Annual Status of Education Report (ASER) which has been conducted by Pratham, a non-governmental organisation. ASER basic survey is a nationally representative household survey conducted by trained volunteers[2] in every rural district in India, which focuses on collecting information of 3-16-year-olds enrolment status, and assesses the foundational literacy and numeracy skills among 5–16-year-olds (Chavan and Banerji, 2013). It uses a two-stage sampling survey design in which the 30 villages are randomly selected[3] during the

[1] Usually a language from the list of languages

[2] The ASER survey is implemented by a network of partner organisations, volunteers, and local District Institute of Educational Training (DIET) institutes, etc.

[3] The ASER follows a rotating sample, so in every iteration of the basic survey, 10 villages are dropped and 10 new villages are added in their place.

first stage by using probability proportional to size (PPS) sampling method, and then, 20 households are randomly surveyed, which have 3-16-year-olds children, within each of these 30 villages during the second stage. By following this sampling design, it makes the data representative even at the district level, unlike other household survey datasets which are only representative at the state level. The ASER 2022 survey reached 616 districts in rural India. The survey also covered 19,060 villages and 374,554 households, surveying 699,597 children between the ages of 3 and 16. In addition to the household survey, the ASER also surveyed the government school observations and captured information such as teacher attendance, child attendance, and physical infrastructure. For this paper, we use only the household level dataset and do not account for school-level characteristics because ASER only surveys one government school in each surveyed village. Also, our study sample includes only those children who were enrolled in primary schools (i.e., grade 1 to 5) as NEP 2020 recommends to teach children in their native language in primary schools.

Our main outcome variable of interest is the children's foundational literacy and numeracy (FLN) skills. The simple ASER reading and maths assessments tools are designed to assess a child's FLN skills, and it is conducted orally and one-on-one. Both the reading and maths abilities are recorded in an ordinal scale. For instance, the ASER reading assessment assigns each child one of five literacy levels: beginner (cannot identify letters), can identify letters but not words, can read words but not a paragraph, can read a short paragraph but not story, and can read a longer story (which corresponds to a standard 2 level text). Similarly, the ASER maths assessment assigns each child one of five numeracy levels: beginner (can't identify numbers 1–9), can identify numbers 1–9 but not 11–99, can perform two-digit subtraction but not division (3 digit by 1), and can perform division. The ASER learning assessment tool shows "high inter-rater reliability and concurrent validity with more extended measures such as the Fluency Battery and the Read India literacy and maths tests, as well as with early grade reading and mathematics assessments" (Vagh, 2010; Results for Development, 2015; as cited in Alcott and Rose, 2017). In our study, the ordinal abilities of reading and maths assessment is broken down into four dummies, each representing one level of ability in their respective categories, creating eight variables in total. For instance the reading levels are converted into four binary variables such as whether a child can read the alphabets, words, paragraph, and story. It takes the value '1' if a child can read a particular level such as can read the paragraph, otherwise it takes '0'. Similarly, we do this for maths levels.

Our main explanatory variable of interest is the children's language match (Match) between their medium of instruction in school and home language. We created a simple dummy of language match and it takes value 1 if the two languages were identical and 0 otherwise. The ASER 2022 survey is the only learning assessment in India that captures a child's home language, medium of instruction in school, and language of assessment along with their grade en-

rolled, school type, parental education, and household indicators, etc. All of this information makes this data suitable for our study to empirically test the hypothesis that children's learning is positively influenced by language match advantage between home language and medium of instruction in school during the children's primary education. With the inception of the NEP 2020 that emphasises the mother-tongue based instruction in primary school, we restrict our sample to students enrolled from 1 to 5 classes. This restriction is further enhanced by psychological literature that underscores the importance of language in learning in a coinciding phase. Table 1 below provides the sampled children descriptive statistics (and variable description) for the outcome and explanatory variables used in this study. It shows that the majority of children haven't achieved their foundational literacy and numeracy skills. Also, around 42% of children have the same home language and medium of instruction in school (i.e., language matched).

3.2 Estimation Strategy

The key outcome variables of interest are the binary variables of children's reading and maths abilities. For simplification, we use the simple linear probability model and the Ordinary Least Squares regression equation which is used for empirical analysis is as follows:

$$Y_{ivds} = \beta_0 + \beta_1 \cdot Match_{ivds} + \delta \cdot X_{ivds} + \epsilon_{isj} (1)$$

where, Y_{ivds} is the outcome of interest for individual i in village v in district d in state s. $Match_{ivds}$ is our main variable of interest that reflects that the child's home language and medium of instruction in school is the same. The parameter of interest is β_1 which provides the effect of language match on the outcome of interest. X is a set of control variables for the child and household level covariates that affect the children's learning as mentioned in descriptive statistics Table 1. We also include state-fixed effects to account for state variations in education policies and supply-side differences. In this specification, we cluster the standard errors at the village level as the first stage sampling was done at the village level and households are similar at this level.

ASER 2022 was the only iteration that recorded home language and school medium instruction in the series of ASER surveys over the last two decades. It is also the only survey to capture these variables along with an effective learning assessment and covariate measurement. Though this eliminates most empirical strategies and restricts our analysis to a cross-sectional dataset, this paper will still be the first in its kind to accurately quantify the effect of a language match/mismatch on learning outcomes in the Indian context using large-scale survey data. Additionally, we present fixed effects with lower levels of granularity as measures of robustness of our estimates.

The first and largest advantage of using ASER data is the quality of the learning data. Each assessment is developed, piloted and modified before it is introduced in the field for the final field. It also conducts assessments in households instead of schools, minimising selection biases from high-performing school students. Second, although it is a rural survey, ASER data is representative at the district level. The survey also records control variables of different categories – demographic, household, and village. But since the survey implementation in villages rests on implementation by community partners and rural social networks, variables such as caste, religion, and income are not recorded. There are some proxies for income, such as vehicle ownership that we use instead. In this paper, we only employ state-fixed effects because each state specifies its language-in-education policy and distributions of speakers of a language are typically independent of caste and religion.

4 Results

The key findings from this paper have been divided into three sections. The first section describes the results from our main framework, and sections 2 and 3 present results from heterogeneity tests that filter out confounds and sources of endogeneity and give us a better picture of the effects of home language instruction advantage. It is important to reiterate that this paper does not present any causal estimates. Though not causal, this is still among the first few papers to explicitly measure the effects of home language instruction on foundational learning outcomes, using such a thorough assessment at the rural India level.

There are 8 regression equations for reading and maths outcomes, with each outcome denoting the four dummies used to assess learning levels – Reading Levels: the ability to read a letter, a word, a paragraph and a story; and Maths levels: 1-digit number recognition, 2-digit number recognition, performing subtraction, and division. First, we do the overall analysis of all students of grades 1 to 5 (i.e., sample 1) for all 8 outcome variables. We then split the sample for higher-order learning outcomes to match grade-level expectations. Therefore, for grades 1 and 2 students (i.e., Sample 2), we only do the analysis for lower levels of reading (i.e., Letter and Word) and maths (i.e., 1-digit number recognition and 2-digit number recognition) outcomes based on their grade level expectations, while we do the analysis for all four levels of reading and maths outcomes for grade 3 to 5 student's (i.e., Sample 3).

4.1 Home language instruction and foundational learning

Table 2 provides the results of home language instruction effect on children's reading levels for full sample 1 students. These results show a positive and significant effect of home language instruction on all the reading levels in both

samples. For instance, children enrolled (grade 3 to 5) in a school instructing in their home language are 1.6 percentage points (p.p.) more likely to be able to recognise letters than those enrolled in a different instruction medium. For reading words, paragraph, and story is 1.7, 2.1, and 0.2 percentage points, respectively. The mechanism governing this relationship is explained by the rationale of the home language literacy advantage hypothesis itself. That is, if students speak a certain language at home and are instructed in the same language in their school, their comfort level in that language increases and this can fuel improved proficiencies in reading and comprehension (Nag & Vagh, 2018). There is an interesting pattern demonstrated by these coefficients: they increase until the paragraph ability and tapers down for the ability to read a story. This pattern could be because our sample was restricted to primary school students and the ability to read and comprehend a story (in any language) unfortunately, is slightly an advanced task given the reading levels of Indian students.

Table 3 shows results of home language instruction effect on children's maths levels for both sample 1 and 2 students. The coefficient interpretation follows a similar pattern: relative to children facing an instruction language mismatch, among those children with a match are 0.5 percentage points more likely to recognise a 1-digit number, a result significant at the 1% level. As the mathematics tasks level increases, both magnitude and strength of the coefficients reduce. For all other levels, we find that the coefficients are positive but insignificant. Current literature is still debating the role of one's mother tongue in the acquisition of mathematical abilities . One strand of literature, using neuroimaging evidence, has identified language-dependent areas that light up when mathematical calculations are performed (Benn et al., 2012; Pesenti et al., 2000). Other studies by Swanson et al. (2019) find that working memory prowess plays a greater role in mathematical ability than language. The crux of all scholarly literature, however, is that mathematics, at every level, is a language or lens by itself and instruction of mathematics in a certain language requires presence of strong orthography, writing styles, corresponding concepts, and reading material to successfully introduce the basics of arithmetic. As one will observe later, several languages lack the capital to ensure this and this results in poorer understanding of mathematics. Naturally, factors such as quality of pedagogy, student motivation, resources and socioeconomic status also affect performance in mathematics (Kraft and Hill, 2015; Bermejo, 2021).

It is important to acknowledge that the categorisation of abilities studied in this paper are supposed to take place in specific classes or standards, as they are referred to in India. For instance, among the reading outcomes, the ability to read a letter or a simple word are grade 1 level tasks or tasks that students enrolled in or completed grade 1 must be able to do. However, reading a paragraph corresponds to a grade 2 (i.e., completed grade 1) level ability (where a similar logic is applied), and reading a story is the most advanced task, corresponds to a grade 3 level task (i.e., completed grade 2). Among

mathematics outcomes, recognising one digit and two digit numbers is a class 1 task, performing subtraction is a class 2 level task, and performing division is a class 3 level task. The ASER survey also recognises this grade-wise categorisation but does not implement it in its assessment of students. Instead, students are encouraged to try as many tasks as they can. Not acknowledging these grade-wise divisions can bias our estimates, giving us lower effects of the home language instruction because students only in certain grades and above are actually learned enough to perform that task. This becomes crucial especially in testing outcomes like reading a paragraph or performing subtraction. So, we re-run regressions based on the earlier empirical strategy but this time, we split the sample into the corresponding standards of the various tasks.

In doing so, we find stronger and higher coefficients (and highly significant) on all four reading levels, as demonstrated in Table 4. While the coefficients for letter and word remained identical because the sample did not change, coefficients for paragraph and story increased to 2.6 and 2.3 percentage points for children enrolled in home language instruction schools (vis-a-vis those without). For the mathematics outcomes, we find marginal improvements only in these coefficients (Table 5). The operating mechanism is similar from before: in the case of reading abilities, the impact of language is more straightforward than arithmetic abilities (Nag & Vagh, 2018; Benn et al., 2012; Swanson et al., 2019).

4.2 Disentangling diversity in home language instruction and foundational learning

From this bigger picture, what can we infer about the role of linguistic diversity in the Indian context? That is, to what extent is this diversity tainting our outcomes and how do we disentangle those effects? In a country like India, this diversity can be a substantial source of endogeneity and uncovering its effect requires a large, representative dataset that records enough languages (whether at home or in school), which fortunately ASER 2022 does for us.

From this subsection onwards, we are solely interested in the key language or languages of instruction in state, apart from English, because we are interested in tying languages spoken at home and school instruction mediums. English-speaking households are scant in India, accounting for only 0.14% of our sample. We create a category variable, taking values 1, 2, or 3 if a child is enrolled in a Hindi medium, English medium, or other regional language medium (henceforth, regional medium) schools, respectively. We adopt a Hindi versus regional medium approach because from the data, we observed that Hindi happened to be the predominant medium of instruction across most north and central Indian states. Ironically, these states also showcase massive amounts of linguistic diversity in terms of languages spoken. In particular, this supposedly 'Hindi' belt is home to hundreds of dialects of Hindi – although there remains some political controversy that many languages are actually distinct by them-

selves and not dialects of a larger language.

First, results in table 6 & 7 show the effects of home language instruction for children enrolled in Hindi medium schools. It is associated with a 1.3 percentage point increase in the likelihood of recognising a letter, which is significant at the 10% level. Beyond that, all results are null and insignificant. There is stronger evidence of association for mathematics outcomes but coefficients are weaker, as shown in table 7. Contrarily, we find stronger associations of higher magnitudes for children enrolled in regional medium schools. From table 10, we can observe that home language instruction is associated with a 3.9, 6.5, 6.2, 4.6 pp. increase in all four reading outcome variables. These coefficients are not only significantly higher compared to the coefficients for Hindi and English medium schools but are also highly significant. Identical results are observed for mathematics outcomes in Table 11. The pattern of results in mathematics in this section present an interesting comparison. We seem to find statistical significance only among lower-order mathematical tasks and are left with nearly null results with no significance. The patterns between language and mathematical outcomes require further probing to account for the subtleties involved in mathematics instruction and learning, which lie outside the scope of this paper.

These results highlight the importance of language-in-education policies at the state level. As discussed earlier, education sits in the concurrent list of subjects, giving states sole jurisdiction over determining its language-in-education and other relevant policies.

4.3 So What are 'Good' Languages?

The notion of 'good' languages for learning is not entirely arbitrary. Factors such as writing style, orthography, and socio-political capital determine the use and instruction of the language. In this section, we further disentangle the results from the previous subsection and identify languages that provide advantages and disadvantages for foundational learning to its speakers.

For simplicity, we concern ourselves with only two major outcomes from the set of eight outcome variables, for grade 3 to 5 students, based on the expectation that by grade 3 they should be able to do these tasks: reading the paragraph and doing subtraction. We chose these outcomes because the acquisition of these abilities are strongly influenced by the use of language and they are nuanced enough to give us a clear picture of a child's learning level. We re-run the main regression models for grade 3 to 5 students by interacting the state code with home language instruction variables. The statewise predicted probabilities of reading a paragraph and doing subtraction for home language instruction (matched) and different language instruction (Not matched) is provided in tables 12 and 13. This information is provided for all medium schools and without English medium schools, and it also provides the

difference in the probabilities for matched and not-matched languages for each state. In line with the results from the previous subsection, we find a strong home language instruction advantage for all medium schools in the following states: Andhra Pradesh, Assam, Himachal Pradesh, Karnataka, Maharashtra, Manipur, Odisha, Sikkim, Tamil Nadu, Telangana, Tripura, and West Bengal, etc. The results from these states confirm the pattern of a strong regional influence and relevance of home language instruction. We test for the robustness of our results by interacting the home language instruction with a third language variable present in the dataset– namely, the language in which the assessment was given (henceforth, test language). From these test language results, we can infer that languages in states listed above demonstrate a strong home language instruction advantage. It is also interesting to note that some of these states have multiple state languages (for instance, Assam, Orissa, and West Bengal, etc.) spread across regions within a state. This clearly demonstrates the role of language-in-education policies in shaping the foundational learning outcomes within a state.

A potential source of endogeneity from these estimates are students enrolled in English-medium schools. Children enrolled in english-medium schools typically hail from more influential families with higher income and better socioeconomic background. A simple descriptive comparison of key socioeconomic control variables from the ASER dataset (and other national surveys like NSSO) substantiates this fact. This could skew their learning outcomes and bias our estimates upwards. Further, this paper is also aimed at assessing the language-in-education policies of state governments, and English is a medium of instruction mostly in private schools. Though we did not explicitly acknowledge the bias from these students in the previous subsection, we categorised children by school medium which eliminates any bias and should adopt a similar strategy for unbiased coefficients. Upon re-running the regressions on the sample of children enrolled only in non-English medium schools, we stumbled upon a fascinating result – the advantage in most of the southern states of India vanishes almost entirely! In this reduced sample, there is evidence of a home language instruction advantage only in West Bengal, Assam, Kerala, and Maharashtra. Thus, we find evidence of strong associations between home language instruction and learning outcomes, cutting across language and socioeconomic factors, only among these four states. These findings also indicate that further deeper analysis is required within a state to understand it better, and how this dynamic is playing out in urban areas.

5 Discussion

Evidence suggests that the language of instruction in school, while teaching during the student's foundational stage, plays an important role in determining their learning outcomes. Along these lines the recent National Education Policy 2020 in India also recommends teaching children in their home language in primary schools. However, there is not much empirical evidence at the In-

dia level on this issue. In this context, this paper is an empirical exploration to understand the dynamics of home language instruction (advantage) on primary school student's foundational learning in rural India using ASER 2022 dataset. The additional information collected by ASER 2022 survey, on children's language spoken at home and medium of instruction in school, make it possible to match whether a child was instructed in his/her home language or not. Given the linguistic diversity in India, this paper serves as a primer for further research on policy on mother-tongue based instruction and effects on foundational learning. This can further direct larger policy implications on language-in-education policies in India.

Overall, our findings suggest that teaching primary school student's in their home language has an advantage in achieving their foundational learning. We identify a consistently strong home language learning advantage especially among all four reading outcome variables. We stratified our results by grade-wise expectations, school medium categories (Hindi, English, and Regional), and state or test language wise variation. In the first stratification, we observed an intensification of the advantage. In the second, we observed the advantage only among regional medium schools and the concluding heterogeneity test pinpointed accurate sources of this advantage from these regional medium schools. However, we don't find any home language instruction advantage in the Hindi language dominant regions after restricting our sample to only Hindi medium schools.

With these results in hand, we present clear evidence of a home language learning advantage in primary education. From a policy lens, conducive language-in-education policies can foster better learning outcomes, resulting in higher enrolment ratios, lower dropout rates, and healthier integration of diverse social groups within society. Classroom and playground interactions can foster harmonious relationships even in the face of linguistic diversity and shape a population that is tolerant, respectful, and well-educated. It is important to note that these policies must also comply with economic and cultural power of all languages; it is insufficient to simply include mother-tongue based instruction in primary classrooms. More pronounced change requires assisting linguistic minority groups in developing socioeconomic and cultural soft power to ensure gains from the classroom eventually translate to improved quality of life. Finally, our study has broader policy implications and highlights a larger need for mother tongue-based education, especially in linguistically diverse regions of the country.

6 References

1. ASER Centre. "Annual Status of Education Report (Rural) 2022". ASER Centre, 18 January 2023, New Delhi.

2. World Bank. "World Development Report 2018". World Bank, 2018, Washington DC.

3. Jhingran, D. "Hundreds of Home Languages in the Country and many in most Classrooms: Coping with Diversity in Primary Education in India". Chapter 14, Social Justice Through Multilingual Education, Phillipson et al., Multilingual Matters, 2009.

4. Seid, Y. "The impact of learning first in mother tongue: evidence from a natural experiment in Ethiopia". Applied Economics, Vol. 51-6, pp.: 577-593, 2019.

5. Jain, T. "Common Tongue: The Impact of Language on Educational Outcomes".The Journal of Economic History, Vol. 77, No. 2, June 2017. 6. Erikkson, K. "Does the language of instruction in primary school affect later labour market outcomes? Evidence from South Africa". Economic History of Developing Regions 29-2, 311-335, 2014.

7. Taylor, S. "Estimating the impact of language of instruction in South African primary schools: A fixed effects approach". Economics of Education Review, Vol. 50, 75-89, February 2016. 8. Boroditsky, L. "How Language Shapes Thought". Scientific American, Vol. 304-2, pp. 62-65, February 2011.

9. Scheele, A. F., Leseman, P. P. M. Mayo, A. Y. "The home language environment of monolingual and bilingual children and their language proficiency". Applied Psycholinguistics, Vol. 31, 117–140, 2010.

10. Karlsen, J., Lyster, S. Lervåg, A. "Vocabulary development in Norwegian L1 and L2 learners in the kindergarten–school transition". Journal of Child Language, Vol. 44-2, 402–426, 2017.

11. Lesaux, N. K., Koda, K., Siegel, L. S. Shanahan, T. Development of literacy, in: D. August, T. Shanahan (Eds) Developing literacy in second-language learners, 75–122, 2006.

12. Nag, S et al. "Home language, school language and children's literacy attainments: A systematic review of evidence from low- and middle-income countries". Review of Education, Vol. 7-1, pp. 91-150, November 2018.

13. Nambissan, G. "Language and Schooling of Tribal Children: Issues Related to Medium of Instruction". Economic & Political Weekly, Vol. 29-42,

October 15, 1994.

14. Jhingran, D. "Early Literacy and Multilingual Education in South Asia". United Nations, UNICEF Regional Office for South Asia, Nepal, 2019.

15. Prakash, N., Azim, M., Chin,and A. "Mehtabul Azim, Aimee Chin, Nishith Prakash". IZA Discussion Paper Series No. 4802, 2010.

16. Chakraborty, T., and S. Kapur. "English Language Premium: Evidence from a Policy Experiment in India Washington University, St Louis." 2008.

17. Meghanathan, R. "Language policy in education and the role of English in India: From library language to language of empowerment". Dreams and Realities: Developing Countries and the English Language, British Council, 2011.

18. Datta, Ashokankur, and Anushree Khatri. "Effect of State's Language Policy on Learning Outcomes of Linguistic Minorities: Evidence from Indian Districts." 2023.

Table 1: Summary Statistics

Variables	Mean	Std. Dev.	Observations
Reading Level			
Letter or Above	0.83	0.38	2,69,570
Word or Above	0.56	0.50	2,69,570
Para or Above	0.37	0.48	2,69,570
Story	0.22	0.42	2,69,570
Math Level			
1-Digit Number Recognition or Above	0.87	0.34	2,69,593
2-Digit Number Recognition or Above	0.60	0.49	2,69,593
Subtraction or Above	0.27	0.45	2,69,593
Division	0.10	0.31	2,69,593
Home Language Instruction (Match)	0.42	0.49	2,70,002
Male Children	0.51	0.50	2,69,894
Grade Enrolled	3.00	1.42	2,70,002
Age (Years)	8.47	1.91	2,69,969
School Type			
Government	0.72	0.45	2,69,299
Private	0.28	0.45	2,69,299
Madrasa	0.00	0.06	2,69,299
Other	0.00	0.03	2,69,299
Private Tuition	0.25	0.43	2,65,181
Father Gone to School	0.81	0.39	2,54,667
Mother Gone to School	0.70	0.46	2,64,960
Household Type			
Pucca	0.52	0.50	2,67,822
Semi-pucca	0.24	0.43	2,67,822
Katcha	0.23	0.42	2,67,822
HH has 4-Wheeler Motor Vehicle	1.88	0.33	2,64,251
HH has 2-Wheeler Motor Vehicle	0.54	0.50	2,67,268
HH has Electricity Connection	0.95	0.22	2,68,963
HH has Toilet	0.79	0.41	2,68,883
HH has a TV	0.63	0.48	2,66,713
HH has a Mobile	0.95	0.22	2,69,177
HH gets Newspaper	0.05	0.23	2,68,396
HH member knows Computer	0.14	0.35	2,68,837
HH has Reading Material	0.05	0.21	2,63,639
HH has a member who completed 12th			
Yes	0.21	0.41	2,67,084
No	0.77	0.42	2,67,084
Don't Know	0.02	0.14	2,67,084

Table 2: Effect of Home Language Instruction on Reading Outcomes

	(1) Letter	(2) Word	(3) Paragraph	(4) Story
Match	0.0164***	0.0167***	0.0211***	0.0150***
	(6.25)	(5.33)	(7.28)	(5.88)
Female	0.0145***	0.0179***	0.0238***	0.0211***
	(9.89)	(9.90)	(13.37)	(13.10)
Private School	0.0736***	0.121***	0.113***	0.0901***
	(34.98)	(41.83)	(38.52)	(33.42)
Tuition	0.0530***	0.0704***	0.0655***	0.0480***
	(24.61)	(25.24)	(23.09)	(18.49)
Father educ	0.0517***	0.0565***	0.0614***	0.0499***
	(18.12)	(18.00)	(21.94)	(21.57)
Mother educ	0.0592***	0.0783***	0.0788***	0.0600***
	(24.80)	(28.11)	(30.35)	(26.90)
School Class	0.0541***	0.105***	0.107***	0.0753***
	(56.57)	(87.21)	(90.29)	(67.31)
Child Age	0.0296***	0.0435***	0.0389***	0.0295***
	(42.31)	(48.33)	(43.62)	(35.34)
Pucca	0.0149***	0.0294***	0.0312***	0.0299***
	(7.24)	(11.60)	(12.67)	(13.70)
2-wheeler	0.0169***	0.0306***	0.0241***	0.0181***
	(8.49)	(12.78)	(10.32)	(8.67)
Electricity	-0.00490	-0.0136**	-0.0118**	-0.00659
	(-0.93)	(-2.42)	(-2.39)	(-1.61)
Toilet	0.0412***	0.0525***	0.0394***	0.0228***
	(15.17)	(17.49)	(14.33)	(9.87)
TV	-0.0269***	-0.0431***	-0.0376***	-0.0284***
	(-11.94)	(-15.71)	(-14.35)	(-12.33)
Mobile	0.0309***	0.0227***	0.0385***	0.0270***
	(6.08)	(4.35)	(8.78)	(7.69)
Newspaper	0.0125***	0.0229***	0.0276***	0.0139***
	(4.04)	(5.08)	(5.69)	(2.93)
Reading Material	0.00314	0.0278***	0.0298***	0.0217***
	(0.77)	(4.97)	(5.04)	(3.85)
Graduate	0.0287***	0.0438***	0.0424***	0.0332***
	(14.04)	(16.03)	(15.12)	(12.28)
Computer	0.0146***	0.0393***	0.0511***	0.0469***
	(6.53)	(12.40)	(15.14)	(13.94)
Size	-0.00312***	-0.00484***	-0.00349***	-0.00209***
	(-9.24)	(-11.47)	(-8.56)	(-5.52)
Num of Obs	226651	226651	226651	226651

Source: Annual Status of Education Report 2022; Standard errors clustered at village level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Effect of Home Language Instruction on Arithmetic Outcomes

	(1) 1-Digit	(2) 2-Digit	(3) Subtraction	(4) Division
Match	0.00596**	0.00190	0.00283	0.000638
	(2.55)	(0.63)	(1.04)	(0.35)
Female	0.00407***	-0.0131***	0.00173	-0.00579***
	(3.08)	(-7.40)	(1.04)	(-4.70)
Private School	0.0600***	0.153***	0.102***	0.0522***
	(32.36)	(56.44)	(36.07)	(25.75)
Tuition	0.0432***	0.0826***	0.0763***	0.0480***
	(22.45)	(30.70)	(27.45)	(23.64)
Father educ	0.0429***	0.0583***	0.0527***	0.0300***
	(16.24)	(18.96)	(20.23)	(18.45)
Mother educ	0.0440***	0.0781***	0.0681***	0.0350***
	(20.53)	(28.88)	(28.84)	(21.45)
School Class	0.0466***	0.0878***	0.0837***	0.0417***
	(54.40)	(76.41)	(73.55)	(46.05)
Child Age	0.0252***	0.0464***	0.0344***	0.0196***
	(40.58)	(54.51)	(40.43)	(28.80)
Pucca	0.0136***	0.0297***	0.0301***	0.0162***
	(7.32)	(11.99)	(13.09)	(10.43)
2-wheeler	0.0119***	0.0258***	0.0214***	0.00780***
	(6.69)	(10.91)	(9.72)	(5.11)
Electricity	-0.00484	-0.0152***	-0.00336	-0.0000430
	(-1.00)	(-2.75)	(-0.78)	(-0.01)
Toilet	0.0345***	0.0562***	0.0302***	0.0111***
	(13.87)	(18.88)	(12.34)	(6.92)
TV	-0.0220***	-0.0437***	-0.0316***	-0.0164***
	(-10.88)	(-16.32)	(-12.93)	(-9.80)
Mobile	0.0359***	0.0388***	0.0174***	0.00364
	(7.41)	(7.50)	(4.51)	(1.56)
Newspaper	0.00129	0.0176***	0.0325***	0.0172***
	(0.47)	(4.10)	(6.85)	(4.66)
Reading Material	0.000849	0.0224***	0.0210***	0.0177***
	(0.23)	(4.33)	(3.84)	(4.21)
Graduate	0.0186***	0.0416***	0.0415***	0.0253***
	(10.23)	(15.68)	(15.00)	(11.92)
Computer Use	0.0164***	0.0455***	0.0485***	0.0309***
	(8.48)	(15.13)	(14.64)	(11.89)
Size	-0.00219***	-0.00447***	-0.00405***	-0.00179***
	(-7.08)	(-11.12)	(-10.37)	(-6.12)
Num of Obs	226688	226688	226688	226688

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Class-wise Effect: Reading Outcomes

	(1) Letter	(2) Word	(3) Paragraph	(4) Story
Match	0.0164***	0.0167***	0.0265***	0.0231***
	(6.25)	(5.33)	(7.72)	(6.27)
Female	0.0145***	0.0179***	0.0286***	0.0331***
	(9.89)	(9.90)	(13.44)	(13.76)
Private School	0.0736***	0.121***	0.124***	0.121***
	(34.98)	(41.83)	(36.16)	(30.78)
Tuition	0.0530***	0.0704***	0.0730***	0.0652***
	(24.61)	(25.24)	(22.30)	(17.87)
Father educ	0.0517***	0.0565***	0.0712***	0.0680***
	(18.12)	(18.00)	(21.37)	(20.01)
Mother educ	0.0592***	0.0783***	0.0920***	0.0836***
	(24.80)	(28.11)	(29.80)	(25.45)
School Class	0.0541***	0.105***	0.111***	0.0943***
	(56.57)	(87.21)	(75.99)	(48.49)
Child Age	0.0296***	0.0435***	0.0369***	0.0273***
	(42.31)	(48.33)	(36.47)	(24.41)
Pucca	0.0149***	0.0294***	0.0355***	0.0397***
	(7.24)	(11.60)	(12.23)	(12.60)
2-wheeler	0.0169***	0.0306***	0.0284***	0.0250***
	(8.49)	(12.78)	(10.31)	(8.21)
Electricity	-0.00490	-0.0136**	-0.0184***	-0.0143**
	(-0.93)	(-2.42)	(-3.03)	(-2.30)
Toilet	0.0412***	0.0525***	0.0481***	0.0343***
	(15.17)	(17.49)	(14.49)	(9.91)
TV	-0.0269***	-0.0431***	-0.0433***	-0.0403***
	(-11.94)	(-15.71)	(-13.80)	(-11.80)
Newspaper	0.0125***	0.0229***	0.0273***	0.00884
	(4.04)	(5.08)	(4.96)	(1.34)
Reading Material	0.00314	0.0278***	0.0285***	0.0226***
	(0.77)	(4.97)	(4.30)	(3.00)
Graduate	0.0287***	0.0438***	0.0484***	0.0465***
	(14.04)	(16.03)	(14.71)	(11.90)
Computer Use	0.0146***	0.0393***	0.0563***	0.0582***
	(6.53)	(12.40)	(14.37)	(12.14)
Size	-0.00312***	-0.00484***	-0.00398***	-0.00289***
	(-9.24)	(-11.47)	(-8.22)	(-5.21)
Mobile	0.0309***	0.0227***	0.0550***	0.0505***
		(4.35)	(10.31)	(9.39)
Cons	0.255***	-0.304***	-0.522***	-0.539***
	(23.88)	(-26.31)	(-41.85)	(-38.53)
Num of Obs	226651	226651	180792	135340

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Class-wise Effect: Mathematics Outcomes

	(1) 1-Digit	(2) 2-Digit	(3) Subtraction	(4) Division
Match	0.00596**	0.00190	0.00448	0.00169
	(2.55)	(0.63)	(1.38)	(0.62)
Female	0.00407***	-0.0131***	0.00274	-0.00663***
	(3.08)	(-7.40)	(1.36)	(-3.45)
Private School	0.0600***	0.153***	0.120***	0.0823***
	(32.36)	(56.44)	(35.56)	(26.19)
Tuition	0.0432***	0.0826***	0.0874***	0.0710***
	(22.45)	(30.70)	(27.14)	(23.38)
Father educ	0.0429***	0.0583***	0.0614***	0.0423***
	(16.24)	(18.96)	(19.87)	(17.19)
Mother educ	0.0440***	0.0781***	0.0816***	0.0527***
	(20.53)	(28.88)	(28.93)	(21.14)
School Class	0.0466***	0.0878***	0.0917***	0.0727***
	(54.40)	(76.41)	(65.56)	(46.40)
Child Age	0.0252***	0.0464***	0.0345***	0.0197***
	(40.58)	(54.51)	(35.40)	(21.70)
Pucca	0.0136***	0.0297***	0.0345***	0.0233***
	(7.32)	(11.99)	(12.65)	(9.87)
2-wheeler	0.0119***	0.0258***	0.0262***	0.0132***
	(6.69)	(10.91)	(9.97)	(5.67)
Electricity	-0.00484	-0.0152***	-0.00736	-0.00522
	(-1.00)	(-2.75)	(-1.39)	(-1.11)
Toilet	0.0345***	0.0562***	0.0371***	0.0192***
	(13.87)	(18.88)	(12.52)	(7.64)
TV	-0.0220***	-0.0437***	-0.0380***	-0.0251***
	(-10.88)	(-16.32)	(-13.00)	(-9.77)
Mobile	0.0359***	0.0388***	0.0286***	0.0134***
	(7.41)	(7.50)	(6.16)	(3.64)
Newspaper	0.00129	0.0176***	0.0365***	0.0231***
	(0.47)	(4.10)	(6.63)	(4.15)
Reading Material	0.000849	0.0224***	0.0200***	0.0261***
	(0.23)	(4.33)	(3.15)	(4.26)
Graduate	0.0186***	0.0416***	0.0479***	0.0384***
	(10.23)	(15.68)	(14.54)	(11.72)
Computer Use	0.0164***	0.0455***	0.0580***	0.0475***
	(8.48)	(15.13)	(14.85)	(11.95)
Size	-0.00219***	-0.00447***	-0.00469***	-0.00249***
	(-7.08)	(-11.12)	(-9.98)	(-5.50)
Cons	0.385***	-0.236***	-0.489***	-0.436***
	(38.60)	(-20.74)	(-41.86)	(-39.44)
Num of Obs	226688	226688	180839	135377

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Reading Outcomes: Hindi Medium

	(1) Letter	(2) Word	(3) Paragraph	(4) Story
Match	0.0129***	0.00185	0.00203	-0.000695
	(3.03)	(0.40)	(0.50)	(-0.20)
Female	0.0105***	0.00782***	0.0117***	0.00835***
	(4.26)	(2.85)	(4.51)	(3.58)
Private School	0.0961***	0.137***	0.127***	0.101***
	(26.92)	(29.64)	(27.49)	(23.01)
Tuition	0.0810***	0.0838***	0.0816***	0.0651***
	(20.96)	(18.36)	(18.90)	(17.12)
Father educ	0.0613***	0.0553***	0.0564***	0.0436***
	(15.15)	(13.39)	(15.65)	(14.34)
Mother educ	0.0627***	0.0710***	0.0707***	0.0526***
	(18.92)	(19.62)	(21.24)	(18.21)
School Class	0.0726***	0.109***	0.0978***	0.0688***
	(48.39)	(62.94)	(59.95)	(45.38)
Child Age	0.0326***	0.0451***	0.0394***	0.0302***
	(31.10)	(36.00)	(33.04)	(27.32)
Pucca	0.0124***	0.0250***	0.0320***	0.0225***
	(3.76)	(6.65)	(9.31)	(7.48)
2-wheeler	0.0192***	0.0319***	0.0245***	0.0187***
	(5.93)	(9.10)	(7.47)	(6.51)
Electricity	0.00808	0.000794	-0.00435	0.00575
	(1.19)	(0.12)	(-0.75)	(1.16)
Toilet	0.0411***	0.0450***	0.0322***	0.0190***
	(10.89)	(11.42)	(9.20)	(6.48)
TV	-0.0302***	-0.0470***	-0.0368***	-0.0280***
	(-8.92)	(-12.22)	(-10.45)	(-8.98)
Mobile	0.0253***	0.0121*	0.0303***	0.0220***
	(3.66)	(1.79)	(5.51)	(4.95)
Newspaper	0.0135**	0.0340***	0.0309***	0.0264***
	(1.97)	(3.86)	(3.47)	(3.04)
Reading Material	0.0112	0.0331***	0.0328***	0.0322***
	(1.38)	(3.04)	(3.00)	(3.27)
Graduate	0.0408***	0.0535***	0.0483***	0.0417***
	(11.54)	(12.17)	(11.24)	(10.22)
Computer Use	0.0182***	0.0312***	0.0431***	0.0452***
	(4.09)	(5.34)	(7.30)	(7.87)
Size	-0.00366***	-0.00482***	-0.00300***	-0.00152***
	(-7.20)	(-8.05)	(-5.30)	(-2.95)
Cons	0.128***	-0.377***	-0.458***	-0.384***
	(8.65)	(-24.64)	(-33.84)	(-32.37)
Num of Obs	99704	99704	99704	99704

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Mathematics Outcomes: Hindi Medium

	(1) 1-Digit	(2) 1-Digit	(3) Subtraction	(4) Division
Match	0.00694*	0.00888*	0.00693*	0.00466*
	(1.85)	(1.96)	(1.80)	(1.72)
Female	-0.000974	-0.0336***	-0.0128***	-0.0151***
	(-0.44)	(-12.15)	(-5.44)	(-8.29)
Private School	0.0719***	0.150***	0.0899***	0.0485***
	(22.08)	(32.54)	(20.18)	(13.99)
Tuition	0.0640***	0.105***	0.0942***	0.0705***
	(18.17)	(23.11)	(22.44)	(21.90)
Father educ	0.0508***	0.0594***	0.0503***	0.0285***
	(13.68)	(14.23)	(15.36)	(12.92)
Mother educ	0.0430***	0.0664***	0.0594***	0.0303***
	(14.57)	(18.52)	(19.90)	(13.91)
School Class	0.0657***	0.0996***	0.0700***	0.0365***
	(48.50)	(58.80)	(46.04)	(30.06)
Child Age	0.0277***	0.0487***	0.0339***	0.0201***
	(29.66)	(39.70)	(30.18)	(22.04)
Pucca	0.0114***	0.0282***	0.0236***	0.0121***
	(3.82)	(7.54)	(7.54)	(5.36)
2-wheeler	0.0150***	0.0255***	0.0194***	0.00602***
	(5.27)	(7.18)	(6.42)	(2.80)
Electricity	0.00323	-0.00476	0.00264	0.00443
	(0.52)	(-0.70)	(0.52)	(1.19)
Toilet	0.0338***	0.0437***	0.0207***	0.00976***
	(9.89)	(11.11)	(6.84)	(4.57)
TV	-0.0236***	-0.0447***	-0.0317***	-0.0164***
	(-7.82)	(-11.69)	(-9.98)	(-7.19)
Mobile	0.0334***	0.0284***	0.0119**	0.000385
	(4.99)	(4.30)	(2.56)	(0.12)
Newspaper	0.00168	0.0221**	0.0352***	0.0188***
	(0.28)	(2.45)	(4.07)	(2.65)
Reading Material	0.00626	0.0361***	0.0215**	0.0205***
	(0.85)	(3.74)	(2.32)	(2.81)
Graduate	0.0264***	0.0582***	0.0508***	0.0335***
	(8.36)	(13.02)	(12.14)	(10.17)
Computer Use	0.0236***	0.0451***	0.0366***	0.0316***
	(6.14)	(7.64)	(6.36)	(6.64)
Size	-0.00246***	-0.00517***	-0.00319***	-0.00160***
	(-5.24)	(-8.84)	(-6.10)	(-3.89)
Cons	0.273***	-0.340***	-0.380***	-0.235***
	(19.67)	(-22.42)	(-30.43)	(-25.06)
Num of Obs	99790	99790	99790	99790

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Reading Outcomes: English

	(1) Letter	(2) Word	(3) Paragraph	(4) Story
Match	-0.0261	-0.0534**	-0.0330	0.00421
	(-1.35)	(-1.98)	(-1.05)	(0.13)
Female	0.0122***	0.0205***	0.0315***	0.0327***
	(5.56)	(5.91)	(8.73)	(9.69)
Private School	0.0288***	0.0860***	0.0887***	0.0697***
	(7.54)	(14.07)	(14.08)	(12.68)
Tuition	0.0171***	0.0312***	0.0197***	0.00110
	(6.39)	(6.96)	(4.04)	(0.24)
Father educ	0.0138**	0.0293***	0.0563***	0.0474***
	(2.31)	(3.63)	(7.76)	(7.91)
Mother educ	0.0285***	0.0563***	0.0592***	0.0532***
	(6.48)	(8.78)	(9.19)	(9.42)
School Class	0.0237***	0.0986***	0.130***	0.104***
	(18.13)	(44.84)	(57.29)	(47.32)
Child Age	0.0185***	0.0335***	0.0296***	0.0217***
	(17.35)	(19.88)	(16.84)	(13.21)
Pucca	0.0126***	0.0377***	0.0402***	0.0337***
	(3.48)	(7.12)	(7.47)	(6.93)
2-wheeler	0.00113	0.0187***	0.0215***	0.0156***
	(0.36)	(3.89)	(4.21)	(3.31)
Electricity	-0.00282	-0.0187	-0.00669	-0.0133
	(-0.26)	(-1.28)	(-0.44)	(-1.06)
Toilet	0.0208***	0.0612***	0.0481***	0.0262***
	(3.51)	(7.27)	(6.14)	(4.00)
TV	-0.0142***	-0.0265***	-0.0307***	-0.0221***
	(-3.80)	(-4.51)	(-5.11)	(-4.17)
Mobile	0.0268**	0.0511***	0.0501***	0.0332***
	(2.28)	(3.11)	(3.62)	(2.96)
Newspaper	0.0107***	0.0117*	0.0215***	0.0242***
	(2.99)	(1.83)	(2.99)	(3.34)
Reading Material	0.00267	0.0245***	0.0301***	0.0277***
	(0.59)	(3.48)	(3.49)	(3.14)
Graduate	0.0137***	0.0289***	0.0285***	0.0209***
	(4.97)	(6.44)	(5.89)	(4.38)
Computer Use	0.0163***	0.0413***	0.0580***	0.0617***
	(5.71)	(8.93)	(11.19)	(12.08)
Size	-0.00189***	-0.00446***	-0.00440***	-0.00383***
	(-3.90)	(-6.01)	(-5.61)	(-4.94)
Cons	0.594***	-0.116***	-0.446***	-0.408***
	(26.25)	(-3.75)	(-15.75)	(-17.18)
Num of Obs	55845	55845	55845	55845

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Mathematics Outcomes: English Medium

	(1) 1-Digit	(2) 2-Digit	(3) Subtraction	(4) Division
Match	-0.0261	-0.0534**	-0.0330	0.00421
	(-1.35)	(-1.98)	(-1.05)	(0.13)
Female	0.0122***	0.0205***	0.0315***	0.0327***
	(5.56)	(5.91)	(8.73)	(9.69)
Private School	0.0288***	0.0860***	0.0887***	0.0697***
	(7.54)	(14.07)	(14.08)	(12.68)
Tuition	0.0171***	0.0312***	0.0197***	0.00110
	(6.39)	(6.96)	(4.04)	(0.24)
Father educ	0.0138**	0.0293***	0.0563***	0.0474***
	(2.31)	(3.63)	(7.76)	(7.91)
Mother educ	0.0285***	0.0563***	0.0592***	0.0532***
	(6.48)	(8.78)	(9.19)	(9.42)
School Class	0.0237***	0.0986***	0.130***	0.104***
	(18.13)	(44.84)	(57.29)	(47.32)
Child Age	0.0185***	0.0335***	0.0296***	0.0217***
	(17.35)	(19.88)	(16.84)	(13.21)
Pucca	0.0126***	0.0377***	0.0402***	0.0337***
	(3.48)	(7.12)	(7.47)	(6.93)
2-wheeler	0.00113	0.0187***	0.0215***	0.0156***
	(0.36)	(3.89)	(4.21)	(3.31)
Electricity	-0.00282	-0.0187	-0.00669	-0.0133
	(-0.26)	(-1.28)	(-0.44)	(-1.06)
Toilet	0.0208***	0.0612***	0.0481***	0.0262***
	(3.51)	(7.27)	(6.14)	(4.00)
TV	-0.0142***	-0.0265***	-0.0307***	-0.0221***
	(-3.80)	(-4.51)	(-5.11)	(-4.17)
Mbile	0.0268**	0.0511***	0.0501***	0.0332***
	(2.28)	(3.11)	(3.62)	(2.96)
Newspaper	0.0107***	0.0117*	0.0215***	0.0242***
	(2.99)	(1.83)	(2.99)	(3.34)
Reading Material	0.00267	0.0245***	0.0301***	0.0277***
	(0.59)	(3.48)	(3.49)	(3.14)
Graduate	0.0137***	0.0289***	0.0285***	0.0209***
	(4.97)	(6.44)	(5.89)	(4.38)
Computer Use	0.0162***	0.0413***	0.0580***	0.0617***
	(5.71)	(8.93)	(11.19)	(12.08)
Size	-0.00189***	-0.00446***	-0.00440***	-0.00383***
	(-3.90)	(-6.01)	(-5.61)	(-4.94)
Cons	0.594***	-0.116***	-0.446***	-0.408***
	(26.25)	(-3.75)	(-15.75)	(-17.18)
Num of Obs	55845	55845	55845	55845

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Reading Outcomes: Regional Medium

	(1) Letter	(2) Word	(3) Paragraph	(4) Story
Match	0.0391***	0.0649***	0.0621***	0.0463***
	(7.28)	(9.28)	(9.27)	(8.06)
Female	0.0244***	0.0341***	0.0388***	0.0340***
	(9.19)	(10.25)	(11.57)	(11.45)
Private School	0.0392***	0.0645***	0.0679***	0.0643***
	(8.73)	(9.67)	(9.63)	(9.75)
Tuition	0.0531***	0.0907***	0.0905***	0.0716***
	(13.11)	(17.27)	(16.25)	(13.81)
Father educ	0.0439***	0.0643***	0.0690***	0.0582***
	(8.55)	(10.72)	(12.68)	(13.68)
Mother educ	0.0607***	0.0907***	0.0893***	0.0650***
	(12.52)	(15.38)	(16.52)	(14.73)
School Class	0.0605***	0.115***	0.109***	0.0680***
	(31.57)	(47.15)	(43.45)	(28.68)
Child Age	0.0252***	0.0394***	0.0380***	0.0289***
	(16.99)	(20.00)	(18.94)	(15.17)
Pucca	0.0173***	0.0246***	0.0200***	0.0322***
	(5.18)	(5.68)	(4.51)	(8.15)
2-wheeler	0.0207***	0.0310***	0.0207***	0.0155***
	(6.03)	(7.07)	(4.80)	(4.05)
Electricity	-0.0275**	-0.0342***	-0.0164	-0.0254***
	(-2.49)	(-2.72)	(-1.41)	(-2.95)
Toilet	0.0343***	0.0550***	0.0482***	0.0276***
	(7.31)	(9.91)	(9.19)	(6.24)
TV	-0.0166***	-0.0292***	-0.0278***	-0.0199***
	(-3.99)	(-5.62)	(-5.47)	(-4.50)
Mobile	0.0256***	0.0318***	0.0591***	0.0439***
	(2.89)	(3.26)	(7.05)	(6.68)
Newspaper	0.0125**	0.0209**	0.0238***	-0.0193**
	(2.05)	(2.51)	(2.61)	(-2.27)
Reading Material	0.00103	0.0282***	0.0286***	0.00376
	(0.15)	(2.92)	(2.84)	(0.39)
Graduate	0.0195***	0.0389***	0.0439***	0.0304***
	(4.80)	(7.32)	(7.82)	(5.65)
Computer Use	0.0169***	0.0462***	0.0479***	0.0278***
	(3.77)	(7.50)	(7.19)	(4.21)
Size	-0.00341***	-0.00552***	-0.00398***	-0.00189**
	(-4.74)	(-5.87)	(-4.50)	(-2.35)
Cons	0.289***	-0.313***	-0.554***	-0.446***
	(13.93)	(-13.22)	(-25.53)	(-24.11)
Num of Obs	66790	66790	66790	66790

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Mathematics Outcomes:Regional Medium

	(1) 1-Digit	(2) 2-Digit	(3) Subtraction	(4) Division
Match	0.0196***	0.0629***	0.0513***	0.0513***
	(4.05)	(9.33)	(8.90)	(8.90)
Female	0.0111***	0.00208	0.0167***	0.0167***
	(4.53)	(0.63)	(5.42)	(5.42)
Private School	0.0366***	0.0809***	0.0499***	0.0499***
	(9.24)	(12.43)	(7.35)	(7.35)
Tuition	0.0459***	0.0927***	0.0909***	0.0909***
	(12.71)	(17.22)	(17.26)	(17.26)
Father educ	0.0367***	0.0643***	0.0538***	0.0538***
	(7.74)	(10.95)	(11.20)	(11.20)
Mother educ	0.0525***	0.0958***	0.0711***	0.0711***
	(11.73)	(16.80)	(15.65)	(15.65)
School Class	0.0516***	0.110***	0.0803***	0.0803***
	(29.59)	(45.50)	(33.21)	(33.21)
Child Age	0.0237***	0.0414***	0.0308***	0.0308***
	(17.31)	(21.46)	(16.06)	(16.06)
Pucca	0.0158***	0.0219***	0.0243***	0.0243***
	(5.04)	(5.02)	(5.88)	(5.88)
2-wheeler	0.0122***	0.0296***	0.0239***	0.0239***
	(3.90)	(6.83)	(6.03)	(6.03)
Electricity	-0.0178*	-0.0284**	-0.0133	-0.0133
	(-1.79)	(-2.32)	(-1.43)	(-1.43)
Toilet	0.0310***	0.0603***	0.0415***	0.0415***
	(7.02)	(10.79)	(9.11)	(9.11)
TV	-0.0181***	-0.0351***	-0.0257***	-0.0257***
	(-4.71)	(-6.69)	(-5.59)	(-5.59)
Mobile	0.0269***	0.0350***	0.0279***	0.0279***
	(3.31)	(3.46)	(3.85)	(3.85)
Newspaper	0.00225	0.0216**	0.0132	0.0132
	(0.38)	(2.53)	(1.48)	(1.48)
Reading Material	-0.00394	0.0141	0.00699	0.00699
	(-0.59)	(1.38)	(0.72)	(0.72)
Graduate	0.0152***	0.0370***	0.0456***	0.0456***
	(4.07)	(6.92)	(8.32)	(8.32)
Computer Use	0.0180***	0.0444***	0.0320***	0.0320***
	(4.44)	(7.30)	(5.02)	(5.02)
Size	-0.00260***	-0.00520***	-0.00460***	-0.00460***
	(-3.86)	(-5.79)	(-5.67)	(-5.67)
Cons	0.390***	-0.309***	-0.461***	-0.461***
	(20.39)	(-13.07)	(-24.15)	(-24.15)
Num of Obs	66739	66739	66739	66739

Source: Annual Status of Education Report 2022; Standard errors clustered at village level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: State-wise Probabilities: Reading Outcomes

State	Paragraph (All)			Paragraph (w/o English MI)		
	Match=0	Match=1	Diff.	Match=0	Match=1	Diff.
ANDHRA PRADESH	0.42	0.50	0.08	0.42	0.47	0.05
ARUNACHAL PRADESH	0.48	0.39	-0.09	0.41	0.30	-0.11
ASSAM	0.46	0.57	0.11	0.41	0.53	0.12
BIHAR	0.50	0.49	-0.01	0.42	0.44	0.02
CHHATTISGARH	0.59	0.62	0.03	0.56	0.58	0.02
GUJARAT	0.60	0.58	-0.02	0.62	0.55	-0.07
HARYANA	0.63	0.61	-0.02	0.56	0.57	0.01
HIMACHAL PRADESH	0.68	0.71	0.03	0.68	0.68	0.00
JAMMU KASHMIR	0.56	0.56	0.00	0.54	0.55	0.00
JHARKHAND	0.50	0.51	0.01	0.44	0.46	0.02
KARNATAKA	0.35	0.43	0.08	0.40	0.40	0.00
KERALA	0.61	0.55	-0.06	0.36	0.52	0.17
MADHYA PRADESH	0.42	0.42	0.00	0.38	0.39	0.00
MAHARASHTRA	0.46	0.64	0.17	0.43	0.60	0.18
MANIPUR	0.62	0.63	0.01	0.63	0.51	-0.12
MEGHALAYA	0.55	0.42	-0.13	0.68	0.37	-0.30
MIZORAM	0.57	0.62	0.04	0.41	0.58	0.18
NAGALAND	0.60	0.59	-0.01	0.56	0.71	0.14
ODISHA	0.56	0.64	0.08	0.51	0.59	0.08
PUNJAB	0.58	0.63	0.04	0.62	0.59	-0.03
RAJASTHAN	0.47	0.43	-0.04	0.44	0.40	-0.04
SIKKIM	0.48	0.54	0.06	0.60	0.50	-0.10
TAMIL NADU	0.31	0.46	0.14	0.45	0.43	-0.02
TELANGANA	0.34	0.44	0.10	0.40	0.41	0.01
TRIPURA	0.47	0.55	0.09	0.35	0.50	0.15
UTTAR PRADESH	0.54	0.51	-0.02	0.47	0.48	0.00
UTTARAKHAND	0.60	0.54	-0.06	0.57	0.51	-0.06
WEST BENGAL	0.41	0.58	0.18	0.33	0.53	0.20

Table 13: State-wise Probabilities: Mathematics

State	Subtraction (All)			Subtraction (w/o English MI)		
	Match=0	Match=1	Diff.	Match=0	Match=1	Diff.
ANDHRA PRADESH	0.50	0.49	0.00	0.48	0.46	-0.02
ARUNACHAL PRADESH	0.54	0.74	0.20	0.41	0.63	0.22
ASSAM	0.37	0.42	0.05	0.30	0.39	0.09
BIHAR	0.46	0.44	-0.02	0.38	0.39	0.01
CHHATTISGARH	0.37	0.40	0.03	0.34	0.37	0.04
GUJARAT	0.23	0.28	0.05	0.19	0.25	0.06
HARYANA	0.52	0.45	-0.08	0.42	0.41	-0.01
HIMACHAL PRADESH	0.51	0.54	0.03	0.46	0.52	0.06
JAMMU & KASHMIR	0.50	0.54	0.04	0.48	0.69	0.21
JHARKHAND	0.44	0.41	-0.03	0.38	0.36	-0.01
KARNATAKA	0.35	0.37	0.02	0.33	0.34	0.01
KERALA	0.39	0.30	-0.09	0.27	0.28	0.01
MADHYA PRADESH	0.29	0.31	0.02	0.24	0.27	0.03
MAHARASHTRA	0.25	0.34	0.09	0.20	0.31	0.11
MANIPUR	0.58	0.37	-0.21	0.62	0.34	-0.28
MEGHALAYA	0.35	0.28	-0.07	0.45	0.24	-0.20
MIZORAM	0.49	0.51	0.02	0.38	0.47	0.09
NAGALAND	0.45	0.61	0.16	0.44	0.50	0.06
ODISHA	0.35	0.46	0.11	0.30	0.42	0.11
PUNJAB	0.53	0.50	-0.03	0.58	0.46	-0.11
RAJASTHAN	0.27	0.29	0.02	0.23	0.27	0.04
SIKKIM	0.49	0.56	0.07	0.40	0.53	0.13
TAMIL NADU	0.28	0.31	0.03	0.35	0.28	-0.07
TELANGANA	0.47	0.48	0.01	0.44	0.45	0.01
TRIPURA	0.36	0.38	0.02	0.26	0.33	0.08
UTTAR PRADESH	0.45	0.41	-0.04	0.37	0.37	0.00
UTTARAKHAND	0.33	0.30	-0.03	0.30	0.27	-0.03
WEST BENGAL	0.29	0.37	0.08	0.20	0.32	0.12