Village dominance and learning gaps in rural India *

Neha Bailwal

University of Göttingen

Email: neha.bailwal95@gmail.com

and

Sourabh Bikas Paul

Indian Institute of Technology Delhi, New Delhi

Email: sbpaul@hss.iitd.ac.in

*The authors would like to thank Prof. Kunal Sen, Prof. Katharina Michaelowa, Prof. Isabel Günther, Dr. Amit Thorat, Dr. Reetika Khera, Dr. Sisir Debnath, Dr. Debasis Mondal, Dr. Rahul Lahoti and other seminar participants at NADEL, ETH Zurich, University of Zurich, University of Göttingen, KDI School, Development Impact Evaluation World Bank, Cornell University, UNU-WIDER and IIT Delhi for comments and feedback. The authors are responsible for the remaining errors.

Abstract

In this paper, we show that the village social structure shapes the learning outcome of marginalised children in rural India. Exploiting the variation in the dominant caste in the village, we find a significant impact of village dominance. The reading and arithmetic skills of Scheduled Castes are better in their own caste-dominated villages compared to higher caste-dominated villages. Our findings show that village dominance affects learning outcomes primarily through differential teachers' behaviour. Children from marginalised castes are more likely to be scolded, beaten up, and treated unfairly by teachers in villages dominated by higher castes compared to their own castes. We do not find evidence of other mechanisms like the difference in parent and children's aspirations, caste diversity, discrimination in the village, or positive group externalities, which could explain why children from Scheduled Castes fare better in their own caste-dominated villages.

Keywords: education; caste; rural India; learning gaps; teacher quality

JEL Classification: I20; I24; I25

1 Introduction

In this paper, we look into the extent of learning gaps across social groups in rural India and how the social identity of the dominant caste group in a village affects this gap. India has achieved almost universal primary school enrolment, with more than 96 percent of children aged between 6-14 years now enrolled in schools (Pratham et al., 2016). However, the learning outcomes of primary school children have not improved to the desired level. Numerous national and international learning assessments show that a large number of students in India are not equipped with basic reading and writing skills (Pritchett, 2013; World Bank, 2017).

Annual Status of Education Report (ASER) finds that the learning outcome of primary school children is abysmally low (Pratham et al., 2018). For instance, only half of the children enrolled in class V can read standard II-level text, and less than one-third possess arithmetic skills like division. The international comparison provided by the Programme for International Student Assessment (PISA) gives a more grim picture of India, ranking 72 out of 73 countries in 2009 (Walker, 2011). In India, learning levels are not only low but so are learning trajectories over time. Muralidharan and Zieleniak (2013) find that an additional year of schooling does not add much to the learning levels of the kids.

In addition to low learning levels, it is well established in the literature that learning outcomes vary across social groups[1] (Borooah, 2012; Desai and Thorat, 2012). Centuries of caste-based discrimination and exclusion of socially and

[1]The caste system divides Indian society into a hierarchical structure where some groups are placed at the top (Brahmins and Forward Castes), and some are at the bottom (Dalits). The caste system is hereditary, as membership in particular group is decided by birth.

economically backward castes placed them at the bottom of the social and economic ladder. In efforts to uplift these two marginalised groups, the Government of India aggregated these castes into separate constitutional schedule known as "Scheduled Castes" (SC) and "Scheduled Tribes" (ST) and provided them several affirmative actions in education and employment. However, despite these efforts, the two marginalised groups, which constitute 16 and 8 percent of India's total population, remain overrepresented among the illiterate, low levels of occupation distribution, consumption, and wages (Hnatkovska et al., 2012, 2013). Children from non-SCST castes perform significantly better than SC and ST in reading, writing, and arithmetic skills. Despite the best attempts and policy initiatives taken by the governments at the centre as well as in states, the learning gaps are still persistent across social groups (Borooah, 2017; Bailwal and Paul, 2021). The Right to Education Act 2009 (RTE), though, improved access to schools and infrastructure facilities, but it failed to have any significant impact on learning skills (Shah and Steinberg, 2019). This inadequate learning outcome with widening social gaps is one of the biggest challenges, which requires attention not only because it is associated with labour market outcome and economic mobility but also because early learning deficit magnifies over time (World Bank, 2017).

A number of researchers have looked into low learning levels and their association with economic status, parental education, aspiration, and other household characteristics (Santhakumar et al., 2016; Parcel and Dufur, 2009; Filmer and Pritchett, 2001). However, a rich body of literature also suggests that school-level factors such as type of school (public vs. private), student-teacher ratio,

and teacher's qualifications are also important factors that determine learning achievement (Muralidharan and Sundararaman, 2015; Dreze and Kingdon, 2001; Ramachandran and Naorem, 2013; Karopady, 2014). Though individual, household, and school-related factors offer an explanation for low learning achievements, in this paper, we argue that village-level caste structure, particularly caste dominance has an important role in determining gaps in learning outcomes across social groups. India, which is known for its rigid caste-based stratification system, the social identity of an individual plays an important role in all spheres of life (Munshi, 2019). The caste of a child is associated with the teacher's attitudes/behaviour, interactions with peers, and sometimes even test scores (Hanna and Linden, 2012; Rawal and Kingdon, 2010). However, emerging literature also shows that the social identity of a larger geographic unit, like a village, plays an important role in influencing the provision and quality of public schools in rural India (Bailwal and Paul, 2021).

Evidence from vast literature suggests the neighbourhood and community of a child not only influence the educational outcome but are the most important factors of social and economic mobility (Chetty et al., 2016; Wodtke and Parbst, 2017; Lei, 2018). In this paper, we intend to understand the influence of the neighbourhood (captured in terms of the dominant social group in a village) on learning outcomes, particularly for children from marginalised caste groups. The performance of the socially and economically backward castes in the spaces which are dominated by the higher caste groups is ambiguous. Residing in villages dominated by higher caste groups could lead to an increase in learning outcomes of the marginalised castes by having better access to public schools

(Bailwal and Paul, 2021). Conversely, there could be a negative impact on the learning outcomes as SC and ST kids may experience more oppression and humiliation as their low caste status is more salient in such settings (Hoff and Pandey, 2006; Bajoria, 2014). Therefore, which of these effects dominates is an interesting empirical question that we seek to investigate in this paper.

This paper contributes to the literature in the following ways. First, the problem of low learning levels with high inequality is not specific to India but is also common in many middle and low-income countries (World Bank, 2017; Singh, 2020). The existing literature has often focused on household income, schooling quality and inputs, teacher quantity and quality, community participation, and governance as possible explanations for low learning levels (Glewwe and Muralidharan, 2016; McEwan, 2015; Kremer et al., 2013). However, with this paper, we aim to contribute to the existing literature by focusing on the role of village dominance[2] as one of the possible reasons for low levels of learning.

Second, we contribute to the literature focusing on how the neighbourhood influences the economic outcomes. Much of the existing literature focusing on residential segregation, poor neighbourhood, and its impact on economic outcomes is based in the Unite States of America and Europe (Chetty et al., 2016; Chetty and Hendren, 2018; Andersson et al., 2014). We aim to contribute to this strand of literature by not only providing evidence from developing country but also by understanding how the social structure of a community/ neighbourhood affects the learning outcome of children, particularly belonging to marginalised groups.

[2]Village dominance is measured by social group with maximum economic power (land ownership) in a village.

Third, there is emerging literature on the identity-based disadvantage of a larger geographical unit (village) and its impact on economic outcomes (Anderson, 2011; Iversen et al., 2014; Coffey et al., 2019). Our study is very close to the two important studies in this area (Anderson, 2011; Iversen et al., 2014). We are similar to both studies as they also look at the impact of caste dominance on the economic performance of the marginalised castes. However, we depart from existing studies in two important ways, one, while Anderson (2011) and Iversen et al. (2014) focus on agricultural yields, incomes of the household, and its impact on growth and poverty, our focus is education, particularly learning outcomes of children. We also look at the possible channel through which village dominance manifests in low learning levels. Second, we use a more recent data set (IHDS 2011) compared to the other two studies, based on 1997-98 and 2005-06 datasets, respectively.

We also contribute to the literature on teachers' characteristics and their influence on the educational attainment of marginalised groups (Karachiwalla, 2019; Muralidharan and Sheth, 2016). Existing literature suggests that teachers sharing the same identity with students plays an important role in determining educational outcomes (Rawal and Kingdon, 2010). We contribute to this strand of literature by showing that it is not only teachers' characteristics but also teachers' behaviour and attitude toward marginalised children that shapes the learning outcomes. The behaviour of teachers is not uniform towards the backward castes; rather, it varies across villages.

Lastly, our study indirectly adds to the behavioural economics and psychological literature, which aims to understand the performance of an individual in

the same gender or ethnic group compared to diverse groups (Booth et al., 2018; Huguet and Regner, 2007). Literature suggest that females in single-sex classrooms are more likely to perform better than their counterparts in co-educational classes because of the low stereotype threat or positive role model effect. We also find that minority groups perform better in their own caste-dominated villages compared to villages dominated by others. However, because of data constraints, we cannot test the stereotype or positive role model hypothesis but instead focus on teachers' behaviour as a possible mechanism to understand our main findings.

We find that SC kids are more likely to obtain a higher score on reading and arithmetic skills when they reside in villages dominated by their own caste groups. However, for ST, we do not find any significant impact of village dominance on learning outcomes. This may be because ST resides in isolation, with more than 70 percent of the ST population residing in villages dominated by their own caste groups. We also try to understand the mechanism through which village dominance influences the learning outcomes. We pay particular attention to teacher's behaviour, and treatment in schools as teachers are believed to be a fundamental agents of the learning process. A key finding is that teacher behaviour is an important channel through which village caste dominance manifests into lower learning outcomes for the marginalised group. SC children are more likely to be beaten, scolded, and treated unfairly when they reside in villages dominated by non-SCST compared to residing in their own caste-dominated villages.

The rest of the paper is organised as follows. Section 2 describes the background and context of the village dominance. In Section 3, we describe the data

and provide descriptive analysis. Section 4 illustrates the empirical strategy. In Section 5, we present our main empirical results, followed by a possible explanation in Section 6. Section 7 provides alternative mechanisms which may explain our main findings. Finally, we present robustness checks in Section 8 and conclude the paper in Section 9.

2 Village dominance: background and context

According to Census 2011, there are more than 600 thousand villages in India, with varied economic and social structure depending on their size, primary occupation, demography, and geographical location. One of the unique features of Indian villages is their population composition and diversity, where people from different castes and religions reside in the same village. However, households are often segregated by hamlets within a village, with SC/ST settlements mostly at the periphery of the villages (Shah et al., 2006).

According to the social hierarchy of the caste system, based on the notion of purity and pollution, SC are placed at the bottom and often referred to as *untouchables* and *outcasted*. While SC faces discrimination because of their low ranking in the social hierarchy, ST, who are behind SC both economically and educationally, face discrimination of a different kind. They are mostly geographically isolated, with a lack of basic infrastructure and public goods and services. The practice of untouchability and discrimination based on caste was legally prohibited by the Indian constitution more than 65 years ago; still, caste is an important determinant of social exclusion and discrimination in Indian villages (Shah et al., 2006; Armstrong et al., 2010; Deshpande, 2011; Thorat and New-

man, 2012).

With considerable economic resources and political influences, upper castes have been the most powerful group in the villages for a very long time. However, with the abolition of the Zamindari system[3] and introduction of the several affirmative programmes (education, occupation, and political representation) in the 1950s, the SCST communities emerged as a dominant caste in some villages by their relative size, improved economic situation, and political representation. The concept of dominant caste was first given by M. N. Srinivas in 1955, where he defined the dominant caste in a village as one with a majority population and the greatest economic and political power. He emphasised that the concept of dominant caste is important to understand the intercaste relation at the local level and in rural societies. The size or strength of caste groups in a village influence the kind of relations it has with other castes in the village. The dominant caste groups have command over resources in the villages and control the local panchayat and village councils (Srinivas, 1959). They have a maximum say in all village matters and activities like legal matters (settling disputes), public resources, organising festivals, etc. However, later in 1970, Dumont (1980) suggested that the dominant caste in a village is only determined by the economic power which arises from land ownership. This later definition of dominant caste based on land ownership given by Dumont (1980) has been widely used in the recent empirical literature (Anderson, 2011; Iversen et al., 2014).

[3]The Britishers introduced the Zamindari system in 1793. Under the *Zamindari system*, *Zamindars* (mostly from higher castes) were made the permanent owners of the land and were asked to collect the rent from the cultivators or the tenants (mostly from backward castes working on those lands). The abolition of *Zamindari system* was an important land reform that helped redistribute the land from large *Zamindars* to tenants working on those land.

In this paper, we borrow the definition of dominant caste in a village from Anderson (2011), which defines caste dominance as based on economic power from land ownership. For instance, if the non-SCST owns the largest share of the total village land, then the dominant social group in the village is non-SCST, and the village is said to be non-SCST dominated. Similarly, we define SC and ST dominated villages using the largest land share criterion. For a robustness check of our results, we use two alternative definitions of dominance based on population share and a combination of land ownership and population shares.

3 The data and descriptive statistics

We use Indian Human Development Survey 2011-12 (IHDS) (Desai and Vanneman, 2011)[4] for our analysis. IHDS data is a nationally representative survey that covers 42,152 households in 1503 villages. This survey covers numerous topics, including education, health, employment, and earnings, among others. This data also collects information on village-level characteristics like population composition, village infrastructure, availability of schools, and individual and household information. A separate schedule on education and health is canvassed for all the households in the sample. This schedule includes the learning assessment test, administered to all the children in the sample household aged 8-

[4]IHDS is a national panel survey. Data was collected in 2004-05 for the first round, and the same households were reinterviewed in 2011- 12. For our main analysis, we use only the recent round because learning outcomes are tested for children aged 8-11 years; therefore, it is not possible to track the learning outcomes of the same child in two rounds. Using both rounds will not help us to exploit the panel structure of the data; it will only give learning outcomes for children over two time periods. However, we test our main hypothesis using IHDS 2005 dataset as well (Appendix Table 20 and Table 21) and find that the influence of the village dominance on learning outcomes for SC has persisted over time.

11 years. We merge individual and household-level data with village information for our analysis. We divide the total population into the three castes groups: SC, ST, and non-SCST[5].

IHDS tests the reading, writing, and arithmetic skills of all available children aged 8-11 years in the sample households. Learning achievements of 12000 kids are available in the final dataset. Around 22 percent of kids belong to SC, and about 8 percent belong to ST, representing their share in the total population of India. In this paper, we focus on two outcomes: reading and arithmetic skills[6]. The score on reading skills ranges from 0 to 4, based on whether a child cannot read at all (0), read letters (1), read words (2), read paragraphs (3), or read a story (4). Similarly, for arithmetic ability, a score of 0 is given to children who cannot identify numbers, a score of 1 is given to those who can recognise numbers, those who can do subtraction are given a score of 2, and the highest score of 3 is given to those who can divide.

Figure 1 shows the reading and math skills of children across social groups in rural India. We see that among the non-SCST, a lower proportion of them cannot read at all or recognise numbers, whereas, among SC and ST, a higher proportion of them cannot read at all or recognise numbers. For instance, while 14 percent of non-SCST students cannot read at all, the same figures for SC

[5]non-SCST includes Brahmin, Forward castes, and Other Backward Castes (OBC). Although OBCs are also classified as economically and socially backward castes in the Indian constitution, we club them with non-SCST groups in our main analysis. This is for two reasons: first, our focus in this paper is on the well-being of the two most disadvantaged groups in India, and second, the economic and social status of OBCs varies across different states in India. Still, as an additional measure of robustness check, we do include OBC as a separate category and do not find any association between village dominance and learning outcomes for OBC kids (results in Appendix Table 17 and Table 18).

[6]As a measure of robustness check, we also use writing skills as our dependent variable and find that our results hold true for writing skills also.

and ST are 20 and 19 percent, respectively. Similarly, 26 percent of SC and 24 percent of ST cannot recognise numbers, and the proportion is much lower for students from the non-SCST group (19 percent).

Before discussing how SC and ST learning outcomes differ across villages, we show some important village and household characteristics across village types in Table 1 and Table 2, respectively. We divide villages into three types (non-SCST, SC, and ST) based on which caste is dominant in a village. Table 1 depicts the distribution of population composition, the pattern of residence in villages, and the status of infrastructure facilities in three types of villages. It is evident from the table that in villages dominated by non-SCST, 20 per cent of the population belongs to SC, and 5 per cent belongs to ST. This suggests that SC are not confined to their own dominated villages but also reside in non-SCST dominated villages. Similarly, in villages dominated by ST, 70 per cent of the population belongs to ST, suggesting lower integration of ST in other types of villages. This pattern of residence, particularly for SC, allows us to exploit the variation in village-level dominance to understand its effects on the gaps in learning outcomes across caste groups.

While 20 percent of the population in non-SCST dominated villages is SC, almost 63 percent of non-SCST dominated villages are segregated by caste as SC live in separate hamlets, highlighting the residential segregation within villages. The last two columns of Table 1 show whether village characteristics differ significantly across different types of villages. We see that except for population composition, the pattern of residence and few infrastructure facilities like banks, post offices, and provision of middle schools, SC and non-SCST villages

are not significantly different from each other. However, ST and non-SCST dominated villages are different in terms of infrastructure facilities pointing towards the lower economic development of the ST-dominated villages compared to non-SCST dominated villages.

Next, in Table 2, we show how the household characteristics vary across different village types[7]. The primary source of income in all three villages is cultivation. We also see that the economic status of the household, type of floor, and roof are significantly different across non-SCST and SC and ST villages. The wide prevalence of untouchability practice in rural areas is shown in the last rows of Table 2. Around 34 percent of the households in non-SCST dominated villages reported practicing untouchability, the corresponding figures are 22 and 27 percent in SC and ST dominated villages, which highlights that SC and ST are more vulnerable in villages dominated by general castes.

As we have seen in Figure 1, the higher proportion of SC and ST are over-represented at the lower level of learning; next, we depict how these marginalised groups fare in terms of reading and arithmetic skills in a different types of villages. Figure 2 shows the learning outcome of social groups across different types of villages. We consider the lowest and highest scores on reading (cannot read and can read a story) and arithmetic skills (cannot recognise numbers and can do division). Figure 2 suggests that, on average, children from marginalised castes perform differently when residing in villages dominated by different castes. For instance, a higher proportion of SC kids can read a story or do division, and a lower proportion cannot read or recognise numbers when they reside in the

[7]We also compare the similar household outcomes of SC and ST residing in both types of villages (Appendix Table 16).

villages dominated by their own caste groups compared to non-SCST. However, for ST, the pattern is not very clear. They score marginally higher in reading skills when they reside in villages dominated by non-SCST. However, the proportion of ST kids who cannot read or recognise numbers is lower in ST-dominated villages compared to non-SCST dominated villages.

The main findings from the descriptive statistics are as follows. First, there are differences in learning outcomes across social groups. Children from socially backward castes are in a disadvantaged position compared to non-SCST across both learning skills. Second, children from the same caste perform differently when staying in different types of villages. These summary statistics are preliminary evidence of the association between village dominance and learning outcomes, which will be further investigated in the remainder of this paper. In the next section of the paper, we discuss our empirical strategy.

4 Empirical strategy

We first examine how much learning gaps across social groups can be explained by individual, household, and school-related factors. In the next specification, we extend this model by including our main variable of interest, village dominance. Since our dependent variable is ordinal, we use an ordered probit model, which is derived from a latent variable model, formally defined as follows:

$$Y^*_{i,h} = \beta_1 SC_i + \beta_2 ST_i + \boldsymbol{X}_{i,h}\boldsymbol{\delta} + \boldsymbol{Z}_s\boldsymbol{\Gamma} + state\_dummies + \epsilon_{i,h}$$

where $Y^*_{i,h}$ is a latent variable capturing the learning outcome of child i in household h. Observed learning outcome (reading and arithmetic) $Y_{i,h}$ is determined by the cutoff parameters (μ_k) and is linked to the latent variable $Y^*_{i,h}$ such that :

$$Y_{i,h} = k \quad if \quad \mu_{k-1} < Y^*_{i,h} \leq \mu_k \quad \text{for } k = 0, \ldots, 4 \ (reading\ skill)$$
$$k = 0, \ldots, 3 \ (arithmetic\ skill)$$

SC and ST are dummy variables for caste with non-SCST as the reference category. $\boldsymbol{X}_{i,h}$ includes individual and household controls that influence the learning outcomes of a child. We control for individual characteristics like age and gender of a child, grade in which child is studying, and household characteristics like household size, economic status of the household captured by assets ownership[8], parent's education, and occupation. $\boldsymbol{Z}_s$ includes school characteristics like type of school (private or public), whether a child gets any incentive (free uniforms and books) from the school, and the quality of the school[9]. We also include state dummies to control for all state-level aggregate effects.

The estimated β_1 and β_2 coefficients give the relative gaps in learning outcomes across castes after controlling for above mentioned individual, household, and school factors. If the coefficients β_1 and β_2 turn out to be negative, there is statistical evidence of low learning levels of SC and ST compared to non-SCST. However, these gaps may be partially explained by how caste dominance in a village influences the learning experience of a child. Therefore, in the next

[8]Assets ownership is a composite index which is created using 9 variables: Bicycle, motorcycle/scooter, colour television, mobile phone, fridge/refrigerator, electric fan, pressure cooker, cable/dish TV, and computer. A household having all 9 assets gets value of 9 and one with the none of these assets gets a score of zero.

[9]Measured by class teacher's absenteeism in the class.

specification, we extend our model by including village caste dominance.

The main estimating equation is :

$$\begin{aligned} Y^{*}_{i,h,v} &= \beta_1 SC_i + \beta_2 ST_i + \beta_3 VDSC_i + \beta_4 VDST_i + \beta_5 SC_i * VDSC_i \\ &. + \beta_6 ST_i * VDSTi + \boldsymbol{X}_{i,h}\boldsymbol{\delta} + \boldsymbol{\Delta}_v + state\_dummies + \epsilon_{i,h,v} \end{aligned}$$

where $VDSC_i$ takes value 1 if the village of residence of child i is dominated by SC and 0 otherwise. Similarly, $VDST_i$ is a dummy variable depicting village dominance by ST. We define village dominance by the largest land share criteria. Our main identification assumption is that the dominance of a village, defined using land ownership, is assumed to be exogenous because the share of land owned is historically determined and remains unchanged for many years (Banerjee and Somanathan, 2007; Anderson, 2011). Using two rounds of IHDS data, we also find that the change in household land ownership between 2005 and 2011 is close to zero (Appendix Figure 3). Further, the proportion of villages dominated by different social groups remains fairly constant between two rounds of IHDS, which strengthens our claim that village dominance is exogenous (Appendix Table 14). The other concern that people may migrate to villages with better economic outcomes is less likely in our sample because 94 percent of households in our sample are staying in their village of residence for more than 50 years (Appendix Table 15).

While we show that the village dominance is exogenous, however; there are several other variables that may introduce the problem of endogeneity in the model, which we need to control. As we have seen in Table 2, the household endowments vary across villages systematically, which may confound the main

findings of the paper. Therefore, we include a set of controls like parent's education, occupation, economic status of household (captured by assets ownership), grade in which child is studying, type of school a child is going to, the health of child and mother[10], and whether a child receives any incentive like free books and uniform from the school, captured by $\boldsymbol{X}_{i,h}$ in the above equation. We also include the village controls like village development[11], and the caste of the village head (($\boldsymbol{\Delta}_v$)) in the estimations because the identity of the village head may vary across villages, and it may also influence the education outcomes directly (Lahoti and Sahoo, 2020; Bhalotra et al., 2014).

The coefficient β_1 can be interpreted as the difference in learning outcomes between SC and non-SCST in villages dominated by non-SCST. Similarly, the coefficient on β_2 can be interpreted as learning gaps between ST and non-SCST in villages dominated by non-SCST. The coefficient β_3 captures the gap in learning outcomes between the kids in SC-dominated villages and non-SCST dominated villages, whereas β_4 captures the gap between kids living in ST-dominated villages and non-SCST dominated villages. The β_5 coefficient associated with $SC * VDSC$ is interpreted as the difference in learning outcomes between SC and non-SCST children living in SC dominated villages. Similarly, β_6 captures the gap between ST and non-SCST living in ST dominated villages. Our main interest is the summation of β_3 and β_5, which reflects the difference in learning outcomes for SC children living in villages dominated by their own caste group and SC children living in non-SCST dominated villages. Similarly, for

[10]Health of a mother and her child is captured by Body Mass Index (BMI).

[11]Village development is captured by a Village Development Index (VDI). VDI is a composite index of 3 variables: pucca road, bank and post offices.

ST children, summation of β_4 and β_6 gives how they fare in ST-dominated villages compared to non-SCST dominated villages. If these effects turn out to be positive, it provides statistical evidence that SC and ST perform better in their own caste-dominated villages compared to higher caste dominated villages. We present the results of these estimations in the next section of the paper.

5 Results

We first present results of how much gap in learning outcomes across castes is explained by socio-economic status and school characteristics. The first three columns of Table 3 depict results of cannot read at all (lowest score), and the following three columns show the result of can read a story (highest score). Similarly, column 1-3 of Table 4 present results of cannot recognise the number (lowest score), and column 4-6 depicts the result of can do division (highest score). Model 1 does not control for anything, whereas we control for individual and household factors in model 2 and add school-level controls and state fixed effects along with individual and household controls in model 3.

We find significant inter-caste differences in both reading and arithmetic skills. Children belonging to socially backward castes are performing worse than their non-SCST counterparts in both learning outcomes; specifically, the disadvantage is higher for ST. After taking into account a child's individual, household, and school characteristics, the learning gap of SC and ST reduces; however, it remains significant. For instance, an SC child is 8.7 percent less likely to read a story compared to a non-SCST child. However, this probability reduces to 4 percent after controlling for socioeconomic status and school quality. Similarly,

for ST, the disadvantage in reading stories reduces from 11 percent to 4.7 percent once we include individual, household, and school controls.

Next, we present the results of our main hypothesis - whether children from marginalised castes perform better when residing in villages dominated by their own castes. Table 5 and Table 6 depicts the result of reading and arithmetic skills, respectively. Model 1 presents the relative gaps in learning across castes. In model 2 and model 3, we include village dominance, caste, and their interaction without and with controls, respectively.

We find two important results. First, SC children score higher on reading and arithmetic skills when they reside in villages dominated by their own caste group than in higher caste-dominated villages. Second, learning outcomes gaps between SC and non-SCST is insignificant in SC-dominated villages. The probability that an SC child residing in a village dominated by SC cannot read at all is 10 percentage points less compared to an SC child residing in a village dominated by non-SCST. Even after controlling for important household, school, and village level characteristics, the own caste dominance advantage remains significant, as depicted in model 3 of Table 5 and Table 6. We do not find any significant association between village dominance and learning outcome for ST. One possible reason for this could be the pattern of residence of ST. Around 70 percent of the population in ST-dominated villages belongs to ST, which suggests ST's degree of integration in non-SCST dominated villages is low. Therefore, it could be because of the lack of variation in village dominance for ST that we fail to see any association between village dominance and learning outcomes for ST.

6 Possible explanation

The evidence in the previous section raises an important question: why do SC children perform better in their own caste-dominated villages? In this section, we investigate a possible channel through which non-SCST village dominance manifests into lower learning outcomes for the marginalised castes.

Teachers are believed to be fundamental agents who influence the learning process in schools. Hanushek and Rivkin (2006) highlight that teachers are the most important resources in schools that contribute to the learning achievements of students. In India, under the RTE Act, physical and mental harassment by teachers is banned and punishable by law. However, 30 percent of parents reported their kids being beaten up, and more than 40 percent reported being scolded in school (Desai and Vanneman, 2011). This suggests that despite legal prohibition, the practice of verbal abuse and corporal punishment continue to be the most common way to ensure discipline in schools.

We hypothesise that teachers' treatment towards socially and economically backward caste children varies in different types of villages. First of all, to check whether teachers' behaviour plays any role in explaining the association between village dominance and learning outcomes, we include the teachers' behaviour[12] variable in our main specification. If the variable of interest (VDSC + SC*VDSC) in the main specification becomes weak or insignificant by including the above dummy variable, it suggests teacher behaviour is an important

[12]Teacher behaviour is captured by a dummy variable which takes a value of 1 if a child is beaten, scolded, or treated unfairly. It is important to note that these variables are constructed based on parents' responses. Variable scolded and beaten is a dummy variable that gets a value of one if parents respond positively that the child has been beaten up and scolded in last 30 days. Similarly, unfair treatment is also a dummy variable which takes a value of one if parents respond that the class teacher treats the child unfairly and 0 otherwise.

channel through which village dominance influences the learning outcome of SC children[13].

Next, we examine how teachers' behaviour towards the marginalised groups varies across villages; we test the hypothesis: whether SC children are less likely to be scolded, beaten up, or treated unfairly in villages dominated by their own castes. We specify the following probit model[14]:

$$\begin{aligned} P(y_i = 1) &= \Phi(\beta_0 + \beta_1 SC_i + \beta_2 VDSC_i + \beta_3 SC_i * VDSC_i \\ &+ \boldsymbol{X}_{i,h}\boldsymbol{\delta} + \boldsymbol{Z}_s + state\_dummies + \epsilon_i) \end{aligned}$$

where Y_i is a binary variable that takes value 1 if a child is beaten up, scolded or treated unfairly in the last 30 days and 0 otherwise. The set of controls $\boldsymbol{X}_{i,h}$,and $\boldsymbol{Z}_s$ includes age, gender and grade of a child, parent's education and occupation, economic status of households, type of school, gender and location of the class teacher. Once again, our main hypothesis is captured by the summation of β_2 and β_3, which is interpreted as a difference in the probability of scolded, beaten up, and unfair treatment for SC children in villages dominated by SC compared to SC children in the non-SCST dominated village. If $\beta_2 + \beta_3$ turns

[13]We note that the teacher's behaviour may be an endogenous variable. A teacher may react to the performance of students belonging to different caste groups. Therefore, it is important that results from these estimations should be interpreted as a mere association, and they should not be interpreted with causal inference. We also believe that physical and verbal punishment and abuse are illegal and punishable by law under the RTE act; therefore, nothing should justify (even low performance) teachers' bad behaviour in schools. We also show in our Appendix Table 19 that even with the same learning outcomes, teachers' behaviour varies across village types, suggesting that the trigger of the bad behaviour of teachers is less likely to be associated with only low learning levels.

[14]We remove ST from our sample for this section as village dominance does not explain low learning levels for ST children.

out to be negative, it suggests that SC kids are less likely to be scolded, beaten up, or treated unfairly in villages dominated by their same caste groups. The coefficient β_3 associated with the interaction term can be interpreted as difference in teachers' behaviour towards SC compared to non-SCST in SC-dominated villages.

Table 7 depicts that when we control for teacher's behaviour in our main specification, the village dominance variable (VDSC + SC*VDSC) becomes weak for arithmetic and no longer a significant variable for reading skills. This suggests that the positive effect of own village dominance for SC is linked to teacher behaviours. The association of teacher's behaviour (which captures the probability of scolding, beating, and treating unfairly) with learning outcomes is negative, suggesting that the verbal and physical abuse by a teacher increases the likelihood of scoring low on both reading and arithmetic skills for a child.

Next, we show how teachers' behaviour varies across village types. Table 8 highlights the teachers' differential treatment across all three indicators of teacher's behaviour: beating, scolding, and unfair treatment. The probability of being beaten up, scolded, and unfairly treated is lower for an SC kid residing in a SC-dominated village than in higher caste-dominated villages. For beaten up, we do not find any significant difference between SC and non-SCST. However, we find that SC children are less likely to be scolded or treated unfairly compared to non-SCST in their own caste dominated villages.

Why do teachers misbehave with SC children in non-SCST dominated villages compared to own caste-dominated villages? One of the reasons could be that teacher behaviour are consistent with their preferences and bias that

makes them discriminates against children from backward castes (Becker, 1971). Teachers preferences and behaviour may be influenced by historical and social reasons. Given the hierarchical nature of the caste system, it is not uncommon for teachers to discriminate against children from marginalised caste (Team, 1999; Shastry and Linden, 2007; Hanna and Linden, 2012). Calling students (verbal abuse) from their castes specific names, not sharing the same source of drinking water with lower castes children, assigning menial tasks to the children from marginalised castes to caste specific segregations in classrooms and while eating MDM are some anecdotal evidence of teachers behaviour against SC children in schools (Ramachandran and Naorem, 2013; Balagopalan and Subrahmanian, 2003). The previous literature also highlights that teachers sharing the same caste with students help them learn better as teachers understand children's background better (Rawal and Kingdon, 2010). Therefore, it could be the higher social distance[15] between the children and teachers in non-SCST dominated villages that biases the teachers' behaviour towards SC children.

The other possible reason for teachers to behave differently in two types of villages could be that in a non-SCST dominated village, teachers exert more influence and power in such villages. Teachers know that SC is a minority caste in the village, and even if they behave badly or unequally, there are lower chances that it will rebound on them because SC parents lack voice and agency in such villages. On the other hand, in SC-dominated villages, the SC is the

[15]We find that, the probability of having SC teacher in non-SCST dominated villages is significantly lower compared to SC dominated villages. Unfortunately, IHDS does not allow us to match child specific information with teachers caste data. Otherwise, we could have tested empirically whether social distance between teacher and child can explain our main findings.

most powerful caste with better social and personal networks at the village and administrative levels. The better social networks in SC-dominated villages may help local people to mobilise themselves better, raise their voices and take strong action against the teachers in case of any unequal and unfair treatment of SC children[16].

7 Alternative channels

The main finding of the paper is that SC children are likely to score better in reading and arithmetic in own caste-dominated villages compared to non-SCST dominated villages. We present that one of the possible explanations for that is teachers are less likely to scold, beat and mistreat SC children in villages dominated by SC. However, there are other alternative explanation that may explain why SC children perform better in own caste-dominated village that may not be linked to teachers' behaviour. In this section of the paper, we explore other possible mechanisms that may explain our main findings.

7.1 School quality and access to schools

One alternative explanation for better performance of SC in their own dominated villages could be that SC children have better access to good quality schools in their own dominated villages compared to non-SCST dominated villages. Since the share of SC in their own caste-dominated village is higher, they may be able to mobilise themselves better and therefore demand a better quality of

[16]Using social networks section of IHDS 2011 data, we find that SC households in SC dominated villages are more likely to have personal acquaintance with local politicians, government officers, and police compared to SC in non-SCST dominated villages.

schools in their own dominated villages. Consequently, it could be the due to better access of schools and not due to teachers' behaviour that SC children perform better in their own dominated villages. However, Table 1 shows no significant difference in access to primary and quality of schools in two types of villages (SC and non-SCST). Access to primary schools is almost universal in both types of villages which in confirmation with existing literature that there is no systematic bias in provision of primary schools in rural India (Bailwal and Paul, 2021). The quality of schools measured by school infrastructure, teachers' training, and teachers' presence in a class are also not significantly different in SC and non-SCST dominated villages, as depicted in Table 1. Therefore, the conjecture that access and quality of public schools may be deriving our main results may not be true in our case.

7.2 Discrimination in village

Another possible reason for better performance for SC children in their own caste-dominated villages could be that the overall level of discrimination in villages dominated by the same caste is lower compared to villages dominated by non-SCST. SC children may face more humiliation, oppression and may feel like an outsider in non-SCST dominated village, which may lower their self confidence (because of negative stereotypes) and hence result in lower learning levels in such villages (Hoff and Pandey, 2006). Table 2 also shows that the overall the practice of untouchability is significantly higher in non-SCST dominated villages compared to SC-dominated villages. This suggests that it could be that it is not the teachers' behaviour but the overall high level of discrimination against

SC in non-SCST dominated villages which explains why SC children perform worse in non-SCST dominated villages. To test whether this is true, we include an indicator of discrimination in village[17] in our main specification. We find that the difference in practice of untouchability across villages does not explain why SC performs better in own caste-dominated villages. Table 9 shows that although the overall high level of discrimination in villages is associated with lower learning outcomes for all children, it does not explain our main findings. Our main variable of interest remains robust even when we include the measure of untouchability in our main specification.

7.3 Parents/children aspiration and involvement

Parents and children's aspiration play out an important role in shaping the education attainment of children (Bernard et al., 2019; Galab et al., 2013). It could be possible that in own caste-dominated villages, friends and families may be able to pass on information about the returns to education and better jobs and thereby inspire both parents and children to have high aspirations. On the other hand, in non-SCST dominated villages, because of higher social distance between groups, access to information may be limited, hence the aspirations of both parents and children. These differences in aspiration across villages may explain why SC children fair better in own caste-dominated villages compared to non-SCST dominated villages.

In our data, we do not have any direct measure of parent's or children as-

[17]Discrimination in a village is dummy variable which takes value one if household in the village practice untouchability or if household report that they have a problem if SC enter their house and kitchen and zero otherwise.

pirations. However, we have some indicators on parent's involvement in child education and time spent by a child in schools and at home studying. We believe that parent's involvement in child education through attending school meetings and discussing child progress with the teacher regularly is positively linked to parent's aspirations. Similarly, the time spent by the child in school and at home studying is also associated with child's aspiration level. Therefore, we include these indirect measures of parent's and children's aspirations in our main specification and find that our main findings remain robust, as depicted in Table 9. Hence, while child and parent aspirations are positively associated with higher learning levels, it does not appear that these differences in child or parent aspirations derive our main results.

7.4 Caste fractionalisation index

It could be possible that village dominance in our model could be picking up the effect of caste fractionalisation in the village. The dominant caste in a village also implies less heterogeneity and low caste fractionalisation index. The wide literature in developed, as well as developing countries shows that higher caste fractionalisation is negatively associated with economic outcomes (Alesina et al., 1999; Banerjee and Somanathan, 2007). Table 1 also shows non-SCST dominated villages are ethnically diverse (proportion of population from different caste groups reside in non-SCST dominated villages) compared to SC-dominated villages[18]. The higher caste diversity in non-SCST may be linked to lower access to public goods and services like public schools and hence lower learning

[18]Overall caste diversity index in non-SCST dominated villages is 0.485 and in SC dominated is 0.460

outcomes for everyone, including children from marginalised castes. However, we find the caste diversity does not explain the lower learning outcomes of SC children in non-SCST dominated villages. Table 9 shows that our main findings remain robust even when we include the measure of caste diversity[19] in our main specification.

7.5 Positive group size effect

Another possibility of better performance of SC in own caste-dominated villages could be because the educated SC parents are more in numbers in SC dominated villages and these educated parents would pull up the average learning outcome among SC children, both directly and through positive within-group learning externalities. To test whether the positive group size effect that explains the better performance of SC in own caste-dominated villages or not, we calculate the proportion of educated SC households[20] in the village and include that in our main specification. Once again, our conjecture that positive group size effect explains why SC performs better in own caste-dominated villages does not hold true in our case.

8 Robustness check

The results of three additional robustness checks are presented in Table 11 - Table 13. We use two alternative measures of village dominance as our robustness check. We define village dominance based on numerical strength instead of land

[19]Caste diversity is measured by 1 minus Herfindahl concentration index, i.e., 1 - s^2 where s^2 is square of population share of caste group c.

[20]SC household with any member having secondary and above education.

ownership in model 1. In model 2, we define village dominance based on both land share and population share as formalised by (Dasgupta and Pal, 2021). $VDSC$ in model 1 of Table 11 is a dummy variable that takes value 1 if the share of SC population is greater than share of non-SCST and ST population and 0 otherwise. Similarly, $VDST$ is a dummy variable that takes value 1 if the share of the ST population is highest in the total village population. $VDSC$ in model 2 is a dummy variable which takes value 1 if the multiplication of population share and land ownership of SC is greater than the multiplication of population share and land ownership of ST and non-SCST and 0 otherwise. We find that our results are robust to alternative definition of village dominance. SC performs significantly better in both reading and arithmetic skills in villages dominated by their own caste groups compared to higher caste-dominated villages, as depicted in Table 11. Similar to our previous results, we do not find any significant impact of the share of land ownership on learning outcomes for ST.

For the second robustness check, we replace the village dominance dummy variable with the degree of dominance. Dominance is captured by the share of land owned and population share of each social group in model 1 and model 2 of Table 12, respectively. In model 3, the degree of dominance is captured by a combination of land and population share. Table 12 depicts the result of our second robustness check. Once again, we find that SC children perform better in reading and arithmetic skills when the share of land owned by SC increases.

For our last robustness check, we use writing skills as our main dependent variable instead of reading and arithmetic shown in 13. Once again, we find that probability of scoring high on writing skills (child can write without mistakes) is

significantly higher in SC-dominated villages compared to non-SCST dominated villages. Similarly, the probability of cannot write at all is lower for SC children when they reside in villages dominated by the same caste group.

9 Conclusion

Despite achieving universal enrolment rates at the primary level, the learning levels among school-going children remain poor in rural India. Not only the learning outcomes are low, but the learning deficit is largest for marginalised caste groups. This paper shows that the low learning level among the marginalised castes, particularly SC, is driven not only by socio-economic factors but is related to much deeper and structural problems in rural India. Village caste dominance, defined as the social identity of the group with the largest share of village land, is an important factor that explains the lower learning outcomes of SC children. We find that SC performs better across both reading and arithmetic skills when they reside in a village dominated by their own castes group compared to villages dominated by higher castes. We also identify an important channel through which village caste dominance manifests itself in low learning outcomes for SC. We find that teacher treatment towards SC in different villages is an important channel. The probability of being beaten up, scolded, and treated unfairly against the SC kids is high when they reside in villages dominated by non-SCST caste groups compared to their own castes.

The findings of this paper highlight a few concerns and policy suggestions. The first concern that emerges from this paper is low learning with widening social gaps. Given that lower caste dominant villages are already disadvan-

taged in terms of both provision and quality of public schools (Bailwal and Paul, 2021, 2019), this rigid nature of village dominance adds more to their disadvantage. One way to compensate for the rigidity of village dominance is to improve the provision and quality of public schools located in villages dominated by marginalised community.

The second issue that is of concern and requires immediate policy attention is teachers' behaviour and treatment towards children in schools. The paper suggests that teachers' treatment towards SC children is more abusive in non-SCST dominated villages compared to SC and ST dominated villages, which makes the schooling experience for SC children far from positive and thereby influences their learning outcomes. One solution for this could be to train teachers and thereby make them more sensitive towards the children from marginalised castes in schools. Apart from this, there should be an effective monitoring system to check ill-treatment and discriminatory practices adopted by teachers in schools against students from the vulnerable section. Efforts should be made to create an environment where learning is enjoyable and not a burden for students, particularly those coming from marginalised backgrounds.

References

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Anderson, S. (2011). Caste as an impediment to trade. *American Economic Journal: Applied Economics*, 3(1):239–63.

Andersson, R., Musterd, S., and Galster, G. (2014). Neighbourhood ethnic composition and employment effects on immigrant incomes. *Journal of Ethnic and Migration Studies*, 40(5):710–736.

Armstrong, D., Davenport, C., Klasing, A. M., Macwan, M., Pradeep, M., Vania, S., and Varma, M. (2010). Understanding untouchability: A comprehensive study of practices and conditions in 1589 villages.

Bailwal, N. and Paul, S. (2021). Caste discrimination in provision of public schools in rural India. *The Journal of Development Studies*, pages 1–22.

Bailwal, N. and Paul, S. B. (2019). Does population composition of a village influence public school quality? Evidence from rural India. *mimeo, Indian Institute of Technology Delhi*.

Bajoria, J. (2014). *'They say we're dirty: Denying an education to India's marginalized*. Human Rights Watch.

Balagopalan, S. and Subrahmanian, R. (2003). Dalit and Adivasi children in schools. *IDS Bulletin*, 34(1):43–54.

Banerjee, A. and Somanathan, R. (2007). The political economy of public goods: Some evidence from India. *Journal of Development Economics*, 82(2):287–314.

Becker, G. (1971). *The economics of discrimination*. University of Chicago Press, Chicago.

Bernard, T., Dercon, S., Orkin, K., and Taffesse, A. S. (2019). Parental aspirations for children's education: Is there a "girl effect"? Experimental evidence from rural Ethiopia. In *AEA Papers and Proceedings*, volume 109, pages 127–32.

Bhalotra, S., Clots-Figueras, I., Cassan, G., and Iyer, L. (2014). Religion, politician identity and development outcomes: Evidence from India. *Journal of Economic Behavior & Organization*, 104:4–17.

Booth, A. L., Cardona-Sosa, L., and Nolen, P. (2018). Do single-sex classes affect academic achievement? An experiment in a coeducational university. *Journal of Public Economics*, 168:109–126.

Borooah, V. K. (2012). Social identity and educational attainment: The role of caste and religion in explaining differences between children in India. *Journal of Development Studies*, 48(7):887–903.

Borooah, V. K. (2017). *The Progress of Education in India: A quantitative analysis of challenges and opportunities*. London: Palgrave Macmillan.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Coffey, D., Deshpande, A., Hammer, J., and Spears, D. (2019). Local social inequality, economic inequality, and disparities in child height in India. *Demography*, 56(4):1427–1452.

Dasgupta, I. and Pal, S. (2021). Touch thee not: Group conflict, caste power and untouchability in rural india. *Journal of Comparative Economics*, 49(2):442–466.

Desai, S. and Thorat, A. (2012). Social inequalities in education. *India Infrastructure Report*, pages 44–52.

Desai, S. and Vanneman, R. (2011). National Council of Applied Economic Research, New Delhi. India Human Development Survey-II (IHDS-II, 2011-12. ICPSR36151-v2). *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].*

Deshpande, A. (2011). *The grammar of caste: economic discrimination in contemporary India*. Oxford University Press, Oxford, New York.

Dreze, J. and Kingdon, G. G. (2001). School participation in rural India. *Review of Development Economics*, 5(1):1–24.

Dumont, L. (1980). *Homo hierarchicus: The caste system and its implications*. University of Chicago Press.

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India. *Demography*, 38(1):115–132.

Galab, S., Vennam, U., Komanduri, A., Benny, L., and Georgiadis, A. (2013). *The impact of parental aspirations on private school enrolment: Evidence from Andhra Pradesh, India*.

Glewwe, P. and Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education*, volume 5, pages 653–743. Elsevier.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–68.

Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078.

Hnatkovska, V., Lahiri, A., and Paul, S. (2012). Castes and labor mobility. *American Economic Journal: Applied Economics*, 4(2):274–307.

Hnatkovska, V., Lahiri, A., and Paul, S. B. (2013). Breaking the caste barrier intergenerational mobility in india. *Journal of Human Resources*, 48(2):435–473.

Hoff, K. and Pandey, P. (2006). Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211.

Huguet, P. and Regner, I. (2007). Stereotype threat among school girls in

quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, 99(3):545.

Iversen, V., Kalwij, A., Verschoor, A., and Dubey, A. (2014). Caste dominance and economic performance in rural India. *Economic Development and Cultural Change*, 62(3):423–457.

Karachiwalla, N. (2019). A teacher unlike me: Social distance, learning, and intergenerational mobility in developing countries. *Economic Development and Cultural Change*, 67(2):225–271.

Karopady, D. (2014). Does school choice help rural children from disadvantaged sections? Evidence from longitudinal research in Andhra Pradesh. *Economic and Political weekly*, pages 46–53.

Kremer, M., Brannen, C., and Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130):297–300.

Lahoti, R. and Sahoo, S. (2020). Are educated leaders good for education? Evidence from India. *Journal of Economic Behavior & Organization*, 176:42–62.

Lei, L. (2018). The effect of neighborhood context on children's academic achievement in china: Exploring mediating mechanisms. *Social science research*, 72:240–257.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, 85(3):353–394.

Munshi, K. (2019). Caste and the Indian economy. *Journal of Economic Literature*, 57(4):781–834.

Muralidharan, K. and Sheth, K. (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources*, 51(2):269–297.

Muralidharan, K. and Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics*, 130(3):1011–1066.

Muralidharan, K. and Zieleniak, Y. (2013). Measuring learning trajectories in developing countries with longitudinal data and item response theory. *UC San Diego*.

Parcel, T. L. and Dufur, M. (2009). Family and school capital explaining regional variation in math and reading achievement. *Research in Social Stratification and Mobility*, 27(3):157–176.

Pratham, P. et al. (2016). Annual status of education report (rural). Technical report.

Pratham, P. et al. (2018). Annual status of education report (rural). Technical report.

Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Brookings Institution Press for Center for Global Development, Washington, D.C.

Ramachandran, V. and Naorem, T. (2013). What it means to be a Dalit or Tribal

child in our schools: A synthesis of a six-state qualitative study. *Economic and Political weekly*, pages 43–52.

Rawal, S. and Kingdon, G. (2010). Akin to my teacher: Does caste, religious or gender distance between student and teacher matter? Some evidence from India. *UCL Institute of Education. Department of Quantitative Social Science*.

Santhakumar, V., Gupta, N., and Sripada, R. M. (2016). *Schooling for all: Can we neglect the demand?* Oxford University Press.

Shah, G., Mander, H., Thorat, S., Deshpande, S., and Baviskar, A. (2006). *Untouchability in rural India*. Sage.

Shah, M. and Steinberg, B. (2019). The right to education act: Trends in enrollment, test Scores, and school Quality. In *AEA Papers and Proceedings*, volume 109, pages 232–38.

Shastry, G. K. and Linden, L. (2007). Identifying agent discretion: Exaggerating student attendance in response to a conditional school nutrition program. *Work. Pap., Columbia Univ.*

Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 18(4):1770–1813.

Srinivas, M. N. (1959). The dominant caste in Rampura. *American anthropologist*, 61(1):1–16.

Team, P. (1999). *Public report on basic education in India*. Oxford University, New Delhi, IN.

Thorat, S. and Newman, K. S. (2012). *Blocked by caste: Economic discrimination in modern India*. Oxford University Press, New Delhi.

Walker, M. (2011). Pisa 2009 plus results: Performance of 15-year-olds in reading, mathematics and science for 10 additional participants.

Wodtke, G. T. and Parbst, M. (2017). Neighborhoods, schools, and academic achievement: A formal mediation analysis of contextual effects on reading and mathematics abilities. *Demography*, 54(5):1653–1676.

World Bank (2017). *World development report 2018: Learning to realize education's promise*. The World Bank.

Tables and Figures

Table 1: Village characteristics by village dominance

	Village dominated by			Difference	Difference
	Non-SCST	SC	ST	(NonSCST-SC)	(NonSCST-ST)
Population composition					
Brahmin	5.595	3.086	1.473	2.509*	4.122***
Forward	22.706	12.386	4.176	10.319***	18.530***
OBC	45.685	27.716	14.254	17.968***	31.431***
SC	19.979	49.272	9.229	-29.293***	10.751***
ST	5.021	6.001	70.168	-0.981	-65.147***
Residence within village					
Different caste reside in separate hamlets	0.662	0.415	0.400	0.247***	0.262***
Separate mohalla for SC	0.632	0.378	0.353	0.254***	-0.127***
Source of drinking water					
Piped	0.479	0.463	0.333	0.016	0.146***
Tubewell	0.117	0.085	0.180	0.031	-0.063**
Handpump	0.314	0.305	0.340	0.009	-0.026
Infrastructure					
Bus facility	0.634	0.634	0.500	-0.001	0.134***
Bank	0.283	0.195	0.113	0.088*	0.169***
Post office	0.570	0.451	0.353	0.119**	0.216***
Pucca road	0.853	0.829	0.733	0.023	0.119***
PDS shops	0.878	0.829	0.820	0.049	0.058**
Medical facilities					
Primary health center	0.855	0.866	0.880	-0.011	-0.025
Community health center	0.964	0.988	0.940	-0.024	0.024
Public and private schools					
Anganwadi	0.987	1.000	0.960	-0.013	0.027**
Primary school	0.989	0.976	0.980	0.013	0.009
Middle school	0.805	0.671	0.667	0.134***	0.138***
Secondary school	0.371	0.293	0.267	0.078	0.104**
College	0.025	0.012	0.040	0.012	-0.015
Primary school quality					
Infrastructure	2.886	2.779	2.644	0.107	0.242***
Teachers position vacant	0.396	0.403	0.362	-0.007	0.034
Teachers in-service training	0.836	0.844	0.826	-0.008	0.010
Nearest town (in km)	13.003	13.725	20.887	-0.722	-7.884***
Distance to district headquarter	44.495	43.210	57.872	1.285	-13.38***
Nearest railway station	26.546	23.988	58.140	2.559	-31.59***
Number of villages	1139	82	150		

Notes: Table reports the village characteristics across village types in our sample. Observation are villages which are divided in three types based on which social group is dominant based on land ownership criterion: non-SCST, SC and ST. Last two columns reports difference in characteristics between non-SCST and SC and non-SCST and ST villages respectively.

Table 2: Household characteristics by village dominance

	Village dominated by			Difference	Difference
	Non-SCST	SC	ST	(NonSCST-SC)	(NonSCST-ST)
Prinicpal source of income for household					
Cultivation	0.383	0.388	0.492	0.006	-0.109***
Agriculture wage labour	0.133	0.162	0.128	-0.029***	0.004
Non agriculture wage labour	0.231	0.236	0.199	-0.005	0 .033***
Salaried	0.097	0.076	0.104	0.021***	-0.007*
Business	0.079	0.069	0.050	0.010***	0.029***
Others	0.078	0.070	0.034	0.008**	0.045***
Economic status					
Household without electricity	0.174	0.171	0.228	0.003	-0.054***
Household without toilets	0.594	0.620	0.677	-0.026***	-0.083
Assets*	3.946	3.744	2.785	0.202***	1.161***
Roof type					
Grass, Thatch, Mud, Wood	0.184	0.144	0.289	0.040***	-0.105***
Tile	0.115	0.184	0.222	-0.069***	-0.108***
Cement	0.197	0.157	0.057	0.040***	0.139***
Brick and stone	0.174	0.118	0.049	0.056***	0.125***
GI Metal, Asbestos	0.120	0.217	0.238	-0.097***	-0.119***
Floor type					
Mud	0.070	0.018	0.018	0.052***	0.051***
Brick and stone	0.072	0.094	0.026	-0.022***	0.047***
Cement	0.372	0.363	0.225	0.009	0.147***
Education					
Literate	0.626	0.621	0.587	0.005	0.040***
Means years of schooling	4.495	4.417	3.879	0.079	0.616***
Untouchability					
Household practice untouchability	0.345	0.221	0.272	0.124***	0.073***
SC household experience untouchability	0.232	0.186		0.046***	
Number of households	21532	1,647	2,616		

Notes: Table reports households characteristics across village types in our sample. Observation are households in three types of villages where villages are categorised based on which social group is dominant based on land ownership criterion : non-SCST, SC and ST. Last two columns reports difference in household characteristics between non-SCST and SC and non-SCST and ST villages respectively.

*Assets is composite index which is created using 9 variables: Bicycle, motorcycle/scooter, colour television, mobile phone, fridge/refrigerator, electric fan, pressure cooker, cable/dish TV, and computer. A household having all 9 assets gets value of 9 and one with the none of these assets gets a score of zero.

Table 3: Marginal effects: Reading gaps explained by individual, household and school factors

	cannot read at all			can read story		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.063***	0.039***	0.028**	-0.087***	-0.057***	-0.040***
	(0.018)	(0.013)	(0.011)	(0.022)	(0.017)	(0.016)
ST	0.083***	0.044***	0.033***	-0.109***	-0.064***	-0.047***
	(0.014)	(0.011)	(0.010)	(0.017)	(0.015)	(0.013)
Individual and household control	No	Yes	Yes	No	Yes	Yes
School control: school type and quality	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	8100	7681	7567	8100	7681	7567
R-sq	0.005	0.086	0.110	0.005	0.086	0.110

Notes: Table reports the marginal effects after Oprobit regression of reading skills on social identity, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Individual and household controls include age and gender of child, grade in which child is studying, household size, parents education, parents occupation, and economic status of households. School controls include type of school (private or public), whether a child get any incentive (free uniforms and books) from school and quality of school, measured by teachers absence in classroom. SC=Scheduled castes, ST=Scheduled Tribes. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Marginal effects: Arithmetic gaps explained by individual, household and school factors

	cannot recognise numbers			can do division		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.063***	0.032**	0.026***	-0.047***	-0.025**	-0.020***
	(0.019)	(0.014)	(0.010)	(0.013)	(0.010)	(0.007)
ST	0.111***	0.059***	0.040***	-0.073***	-0.043***	-0.030***
	(0.014)	(0.014)	(0.011)	(0.009)	(0.010)	(0.008)
Individual and household control	No	Yes	Yes	No	Yes	Yes
School control: school type and quality	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	8063	7651	7534	8063	7651	7534
R-sq	0.005	0.103	0.131	0.005	0.103	0.131

Notes: Table reports the marginal effects after Oprobit regression of arithmetic skills on social identity, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Individual and household controls include age and gender of child, grade in which child is studying, household size, parents education, parents occupation, economic status of household. School controls include type of school (private or public), whether a child get any incentive (free uniforms and books) from school and quality of school, measured by teachers absence in schools. SC=Scheduled castes, ST=Scheduled Tribes. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Marginal effects of village dominance and caste on reading skills

	cannot read at all			can read a story		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.063***	0.073***	0.027**	-0.087***	-0.092***	-0.037**
	(0.018)	(0.017)	(0.013)	(0.022)	(0.019)	(0.017)
ST	0.083***	0.092***	0.030**	-0.109***	-0.108***	-0.041**
	(0.014)	(0.027)	(0.017)	(0.017)	(0.026)	(0.021)
VDSC		-0.008	-0.030*		0.011	0.045*
		(0.029)	(0.016)		(0.043)	(0.027)
VDST		0.002	-0.023		-0.003	0.034
		(0.023)	(0.016)		(0.033)	(0.028)
SC*VDSC		-0.041	-0.001		0.067	0.003
		(0.029)	(0.024)		(0.047)	(0.042)
ST*VDST		0.085***	0.015		-0.100***	-0.024
		(0.029)	(0.018)		(0.038)	(0.028)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	7947	7947	7334	7947	7947	7334
R-sq	0.005	0.006	0.110	0.005	0.006	0.110
$\beta_3 + \beta_5$		-0.100***	-0.052**		0.137***	0.078**
		(0.025)	(0.021)		(0.048)	(0.034)
$\beta_4 + \beta_6$		-0.006	-0.038		0.005	0.051
		(0.028)	(0.023)		(0.025)	(0.033)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on reading skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index, and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Marginal effects of village dominance and caste on arithmetic skills

	Cannot recognise numbers			Division		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.063***	0.073***	0.028***	-0.047***	-0.048***	-0.021***
	(0.019)	(0.017)	(0.009)	(0.013)	(0.011)	(0.006)
ST	0.111***	0.113***	0.048**	-0.073***	-0.068***	-0.033***
	(0.014)	(0.024)	(0.021)	(0.009)	(0.010)	(0.013)
VDSC		-0.019	-0.029**		0.016	0.024*
		(0.025)	(0.014)		(0.020)	(0.014)
VDST		-0.007	-0.025		0.006	0.025
		(0.027)	(0.017)		(0.023)	(0.019)
SC*VDSC		-0.054**	-0.007		0.050*	0.007
		(0.027)	(0.021)		(0.027)	(0.020)
ST*VDST		0.123***	0.046		-0.075***	-0.037
		(0.036)	(0.031)		(0.026)	(0.026)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	7914	7914	7,308	7914	7914	7,308
R-sq	0.005	0.007	0.130	0.005	0.007	0.130
$\beta_3 + \beta_5$		-0.121***	-0.051***		0.093***	0.040**
		(0.021)	(0.019)		(0.024)	(0.017)
$\beta_4 + \beta_6$		0.002	-0.027		-0.001	0.017
		(0.040)	(0.024)		(0.017)	(0.016)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on arithemtic skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index, and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Marginal effects of village dominance, caste and teacher's behaviour

	cannot read at all	can read story	cannot recognise numbers	can do division
SC	0.029***	-0.042***	0.033***	-0.025***
	(0.011)	(0.015)	(0.009)	(0.006)
ST	0.042**	-0.056***	0.058***	-0.040***
	(0.017)	(0.020)	(0.021)	(0.013)
VDSC	-0.032**	0.053*	-0.025*	0.021*
	(0.015)	(0.028)	(0.013)	(0.012)
VDST	-0.019	0.028	-0.022	0.018
	(0.020)	(0.032)	(0.015)	(0.015)
Teacher behaviour	0.046***	-0.071***	0.048***	-0.039***
	(0.008)	(0.013)	(0.008)	(0.006)
SC*VDSC	0.016	-0.028	0.008	-0.007
	(0.023)	(0.039)	(0.022)	(0.020)
ST*VDST	0.010	-0.016	0.041	-0.035
	(0.019)	(0.032)	(0.027)	(0.023)
Controls	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
N	6378	6378	6357	6357
R-sq	0.114	0.114	0.139	0.139
$\beta_3 + \beta_5$	-0.039	0.058	-0.037*	0.028
	(0.024)	(0.038)	(0.021)	(0.018)
$\beta_4 + \beta_6$	-0.052**	0.070**	-0.040	0.026
	(0.025)	(0.033)	(0.026)	(0.017)

Notes: This table reports the marginal effects from Oprobit model of caste, village dominance and teachers behaviour on reading and arithmetic skills, weighted using sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index, and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Marginal effects: village dominance and teacher's behaviour

	Beaten	Scolded	Unfair treatment
SC	0.013	-0.001	-0.002
	(0.023)	(0.019)	(0.021)
VDSC	-0.055	-0.052	-0.068**
	(0.074)	(0.062)	(0.033)
SCxVDSC	-0.055	-0.127**	-0.031*
	(0.101)	(0.059)	(0.018)
Controls	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
N	6393	6391	7162
R-sq	0.202	0.213	0.177
$\beta_2 + \beta_3$	-0.108**	-0.150***	-0.090**
	(0.050)	(0.049)	(0.038)

Notes: This table reports the marginal effects from probit model of village dominance on probability of being beaten up, scolded and unfair treatment, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include age of a child, gender of a child, grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), gender of a class teacher, and location of a class teacher. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Alternative mechanisms: Marginal effects of village dominance on reading skills

	cannot read at all					can read a story				
	model 1	model 2	model 3	model 4	model 5	model 1	model 2	model 3	model 4	model 5
SC	0.040***	0.031**	0.031***	0.029**	0.019	-0.056***	-0.044**	-0.045***	-0.042**	-0.026
	(0.010)	(0.015)	(0.012)	(0.013)	(0.013)	(0.014)	(0.021)	(0.017)	(0.018)	(0.016)
VDSC	-0.030	-0.034*	-0.034*	-0.035*	-0.039*	0.045	0.054	0.055	0.054	0.057
	(0.020)	(0.019)	(0.018)	(0.019)	(0.021)	(0.036)	(0.037)	(0.034)	(0.036)	(0.036)
SCxVDSC	0.005	0.011	0.004	0.005	-0.001	-0.010	-0.019	-0.007	-0.009	0.002
	(0.025)	(0.030)	(0.028)	(0.031)	(0.034)	(0.046)	(0.055)	(0.052)	(0.057)	(0.057)
Discrimination in village	0.036***					-0.053***				
	(0.013)					(0.017)				
Parents aspiration and involvement : participation in school committee		0.010					-0.014			
		(0.010)					(0.015)			
Attended PTA meeting		-0.041***					0.063***			
		(0.012)					(0.021)			
Child aspiration: time spend per week in school			0.000					-0.001		
			(0.000)					(0.000)		
time spend per week in doing homework			-0.004***					0.006***		
			(0.001)					(0.002)		
Caste fractionalization index				0.032					-0.048	
				(0.023)					(0.034)	
Prop. of educated SC households					-0.047					0.065
					(0.030)					(0.040)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6411	5889	6244	6340	5435	6411	5889	6244	6340	5435
R-sq	0.112	0.115	0.116	0.114	0.113	0.112	0.115	0.116	0.114	0.113
$\beta_3 + \beta_5$	-0.055***	-0.050**	-0.055***	-0.052**	-0.053**	0.082**	0.074**	0.085***	0.080**	0.078**
	(0.021)	(0.023)	(0.019)	(0.022)	(0.025)	(0.035)	(0.036)	(0.032)	(0.036)	(0.039)

Notes: This table reports the marginal effects from oprobit model of caste and village dominance on reading skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Alternative mechanisms: Marginal effects of village dominance on arithmetic skills

	cannot recognise numbers					can do division				
	model 1	model 2	model 3	model 4	model 5	model 1	model 2	model 3	model 4	model 5
SC	0.036***	0.027***	0.031***	0.028***	0.020**	-0.028***	-0.020***	-0.024***	-0.022***	-0.015**
	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.006)	(0.006)	(0.007)
VDSC	-0.029	-0.029*	-0.037***	-0.033**	-0.038*	0.024	0.024	0.032**	0.028	0.032
	(0.018)	(0.018)	(0.014)	(0.016)	(0.021)	(0.019)	(0.018)	(0.016)	(0.017)	(0.022)
SCxVDSC	0.006	-0.015	0.000	-0.009	-0.005	-0.006	0.014	-0.001	0.009	0.026
	(0.023)	(0.024)	(0.023)	(0.022)	(0.025)	(0.023)	(0.023)	(0.024)	(0.022)	(0.027)
Discrimination in village	0.026*					-0.022*				
	(0.015)					(0.011)				
Parents aspiration and involvement : participation in school committee		-0.031***					0.027***			
		(0.006)					(0.006)			
Attended PTA meeting		-0.013					0.011			
		(0.013)					(0.012)			
Child aspiration: time spend per week in school			0.001					-0.001		
			(0.001)					(0.000)		
time spend per week in doing homework			-0.005***					0.004***		
			(0.001)					(0.001)		
Caste fractionalization index				-0.001					0.001	
				(0.039)					(0.033)	
Prop. of educated SC households					-0.054					0.043
					(0.058)					(0.046)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6411	5889	6226	6321	5418	6411	5889	6226	6321	5418
R-sq	0.131	0.134	0.138	0.133	0.133	0.131	0.134	0.138	0.133	0.133
$\beta_3 + \beta_5$	-0.061***	-0.060***	-0.060***	-0.061***	-0.056***	0.051***	0.052***	0.052***	0.052***	0.048**
	(0.017)	(0.016)	(0.015)	(0.016)	(0.021)	(0.017)	(0.016)	(0.015)	(0.016)	(0.020)

Notes: This table reports the marginal effects from oprobit model of caste and village dominance on arithmetic skills, weighted using sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Robustness check 1: Marginal effects of population and combination of land and population on learning outcomes

	Reading				Arithemetic			
	cannot read at all		can read a story		Cannot recognise numbers		Can do division	
	model 1	model 2	model 1	model 2	model 1	model 2	model 1	model 2
SC	0.033***	0.031***	-0.043***	-0.042***	0.032***	0.031***	-0.023***	-0.023***
	(0.010)	(0.010)	(0.014)	(0.014)	(0.012)	(0.012)	(0.008)	(0.008)
ST	0.018	0.024	-0.025	-0.033	0.052***	0.055***	-0.036***	-0.038***
	(0.015)	(0.016)	(0.021)	(0.022)	(0.018)	(0.019)	(0.011)	(0.012)
VDSC	-0.028	-0.020	0.042	0.029	-0.057**	-0.042**	0.052*	0.037**
	(0.020)	(0.016)	(0.034)	(0.027)	(0.025)	(0.018)	(0.028)	(0.018)
VDST	-0.036*	-0.026	0.062*	0.043	-0.017	-0.027	0.013	0.023
	(0.018)	(0.017)	(0.036)	(0.031)	(0.025)	(0.020)	(0.023)	(0.019)
SC*VDSC	-0.042	0.003	0.078	-0.005	0.001	0.022	-0.001	-0.022
	(0.029)	(0.025)	(0.052)	(0.042)	(0.036)	(0.030)	(0.041)	(0.029)
ST*VDST	0.049**	0.036	-0.081*	-0.057	0.049**	0.045*	-0.008**	-0.037*
	(0.024)	(0.022)	(0.042)	(0.036)	(0.024)	(0.025)	(0.004)	(0.022)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7476	7476	7947	7476	7449	7449	7449	7449
R-sq	0.110	0.111	0.110	0.111	0.132	0.132	0.132	0.132
$\beta_3 + \beta_5$	-0.084***	-0.043**	0.137***	0.060**	-0.080***	-0.050**	0.069***	0.038*
	(0.018)	(0.019)	(0.037)	(0.030)	(0.025)	(0.025)	(0.027)	(0.021)
$\beta_4 + \beta_6$	-0.006	-0.014	0.008	0.019	-0.037	-0.037	0.024	0.024
	(0.020)	(0.021)	(0.027)	(0.028)	(0.024)	(0.025)	(0.016)	(0.016)

Notes: This table reports the marginal effects from oprobit model of caste and village dominance on arithmetic skills, weighted using sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Robustness check 2: Marginal effects of degree of dominance on learning outcomes

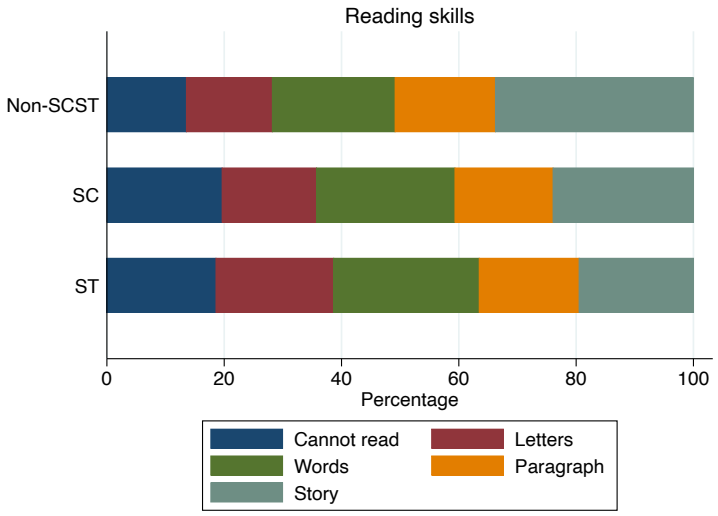
	Reading						Arithmetic					
	Cannot read			Story			Cannot recognise numbers			Can do division		
	model 1	model 2	model 3	model 1	model 2	model 3	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.036***	0.041***	0.035***	-0.047***	-0.053***	-0.046***	0.037***	0.039***	0.034***	-0.026***	-0.027***	-0.026***
	(0.013)	(0.015)	(0.012)	(0.017)	(0.018)	(0.016)	(0.009)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)
ST	0.025*	0.031**	0.026**	-0.035*	-0.043**	-0.036**	0.052**	0.070***	0.057**	-0.036**	-0.047***	-0.039***
	(0.014)	(0.016)	(0.013)	(0.019)	(0.020)	(0.017)	(0.024)	(0.023)	(0.023)	(0.015)	(0.012)	(0.013)
share VDSC	-0.030	-0.057***	-0.042	0.031	0.074***	0.046	-0.096***	-0.089***	-0.154***	0.073**	0.068***	0.121***
	(0.026)	(0.021)	(0.045)	(0.038)	(0.027)	(0.066)	(0.034)	(0.025)	(0.047)	(0.028)	(0.020)	(0.039)
share VDST	-0.039**	-0.061**	-0.099***	0.059**	0.091**	0.150***	-0.040	-0.065	-0.055	0.031	0.050	0.043
	(0.020)	(0.026)	(0.027)	(0.029)	(0.038)	(0.040)	(0.034)	(0.042)	(0.049)	(0.027)	(0.032)	(0.041)
Share VDSC*SC	-0.141***	-0.142**	-0.175***	0.175***	0.171**	0.219***	-0.156***	-0.129*	-0.163**	0.103***	0.084*	0.109**
	(0.049)	(0.066)	(0.051)	(0.056)	(0.074)	(0.059)	(0.060)	(0.070)	(0.083)	(0.039)	(0.044)	(0.054)
Share VDST*ST	-0.022	-0.040	-0.024	0.028	0.049	0.032	-0.045	-0.091**	-0.064	0.027	0.049***	0.037
	(0.022)	(0.027)	(0.023)	(0.027)	(0.030)	(0.029)	(0.043)	(0.044)	(0.049)	(0.022)	(0.019)	(0.024)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7476	7395	7452	7476	7395	7452	7449	7369	7426	7449	7369	7426
R-sq	0.114	0.114	0.114	0.114	0.114	0.114	0.137	0.137	0.137	0.137	0.137	0.137

Notes: This table reports the marginal effects from oprobit model of caste and village dominance on reading and arithmetic skills, weighted using sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

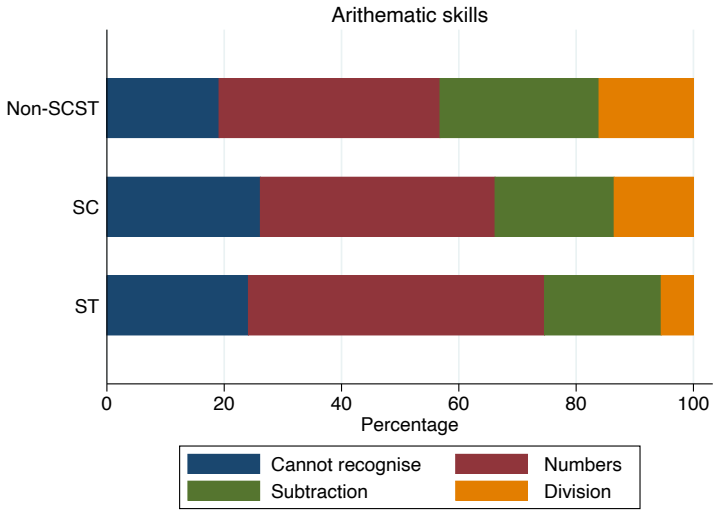
Table 13: Robustness check 3: Marginal effects of village dominance and caste on writing skills

	Cannot write			No mistake		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.087***	0.099***	0.056***	-0.086***	-0.093***	-0.056***
	(0.019)	(0.018)	(0.016)	(0.018)	(0.016)	(0.016)
ST	0.100***	0.104***	0.050	-0.097***	-0.096***	-0.049
	(0.021)	(0.037)	(0.034)	(0.021)	(0.031)	(0.032)
VDSC		-0.050**	-0.061***		0.052*	0.068***
		(0.024)	(0.021)		(0.027)	(0.026)
VDST		-0.015	-0.013		0.016	0.013
		(0.033)	(0.034)		(0.036)	(0.035)
SC*VDSC		0.003	0.067**		-0.004	-0.076*
		(0.031)	(0.034)		(0.036)	(0.040)
ST*VDST		0.120**	0.045		-0.113**	-0.046
		(0.049)	(0.031)		(0.050)	(0.033)
School control: school type and quality	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	7876	7876	7289	7876	7876	7289
R-sq	0.005	0.007	0.117	0.005	0.007	0.117
$\beta_3 + \beta_5$		-0.126***	-0.063***		0.122***	0.062***
		(0.037)	(0.022)		(0.040)	(0.022)
$\beta_4 + \beta_6$		0.001	-0.018		-0.001	0.017
		(0.052)	(0.046)		(0.047)	(0.044)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on writing skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.



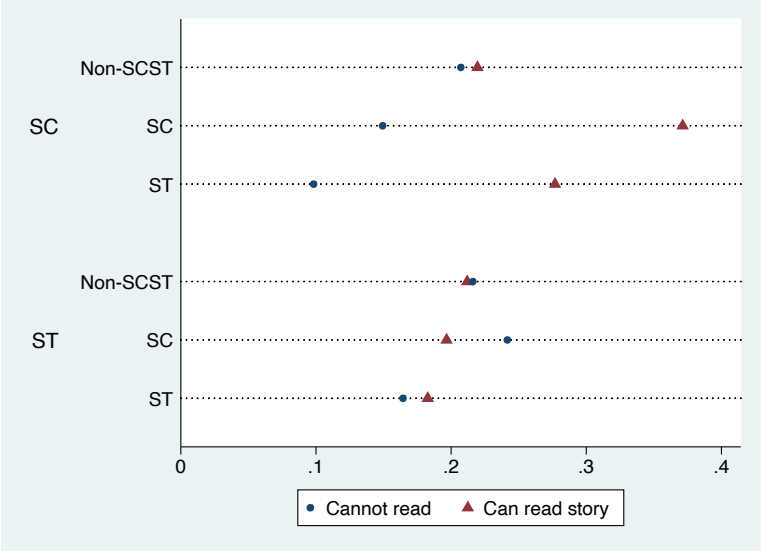
(a) Reading



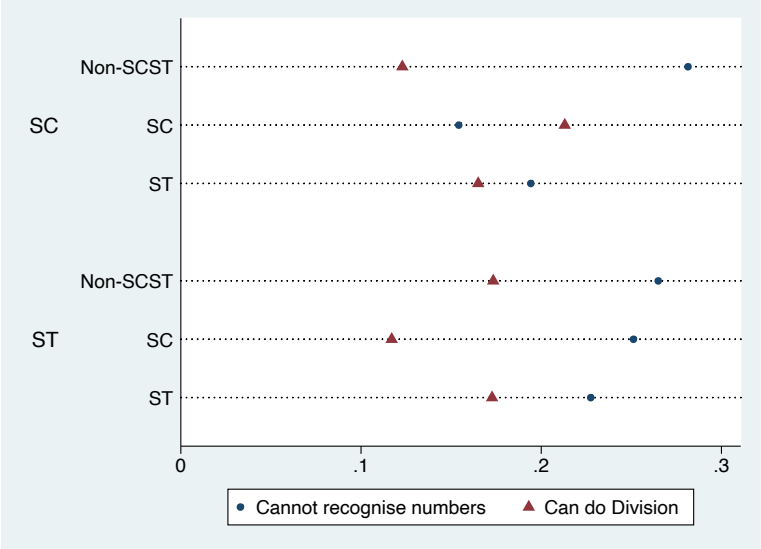
(b) Arithmetic

Figure 1: Learning outcome across social groups

Notes: Figure displays the learning outcomes of children aged 8-11 years across three social groups: non-SCST, SC and ST. Panel (a) display results for reading skills, where children are divided into 5 categories: cannot read at all (lowest score), can read letters, can read words, can read paragraph and can read story (highest score). Panel (b) display results for arithmetic skills, where children are divided into 4 categories: cannot recognise numbers (lowest score), can read numbers, can do subtraction and can do division (highest score).



(a) Reading



(b) Arithmetic

Figure 2: Learning outcome of SC/ST and village dominance

Notes: Figures depicts the learning outcomes of SC and ST children in different type of villages, where villages vary by which caste is dominated. Villages are divided in three types: non-SCST, SC and ST based on which caste is dominated in a village. Panel (a) display minimum (cannot read at all) and maximum (can read a story) score on reading skills, while Panel (b) display minimum (cannot recognise numbers) and maximum (can do division) score on arithmetic skills.

Appendix

Table 14: Village dominance over time

Village dominance	2005	2011
SC	6.792	5.896
ST	10.753	10.842
non-SCST	82.455	83.262
Number of villages	1,501	1,410

Notes: This table shows the proportion of villages dominated by SC,ST and non-SCST in 2005-06 and 2011-12. Village dominance is measured by land ownership criteria.

Table 15: Household years of residence

Years ago household came to village of residence	Frequency
Forever	90.682
50-85 years	3.542
49-15 years	4.372
14-10 years	0.653
Less than 9 years	0.751
Number of households	27,576

Notes: This table shows the years ago the the households in our sample came to the village of residence. Observations are number of households in rural India.

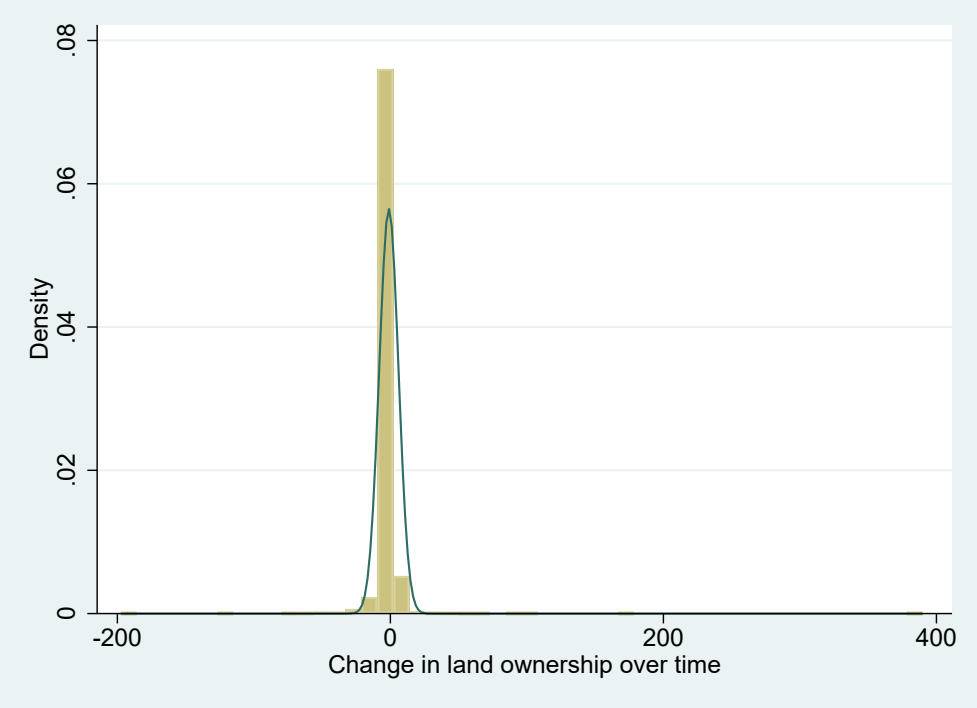


Figure 3: Change in land ownership (in acres) between 2005-06 and 2011-12

Notes: This plot shows the change in ownership of household land between 2005-06 and 2011-12 using IHDS panel data.

Table 16: Household characteristics of SC/ST by village dominance

Characersitics of households	Village dominated by			Village dominated by		
	Non-SCST	SC	Difference (NonSCST-SC)	Non-SCST	ST	Difference (NonSCST-ST)
Prinicpal source of income for household						
Cultivation	0.191	0.299	-0.109***	0.344	0.535	-0.192***
Agriculture wage labour	0.228	0.203	0.025***	0.225	0.116	0.109***
Non agriculture wage labour	0.376	0.278	0.098***	0.288	0.191	0.097***
Salaried	0.100	0.086	0.013**	0.072	0.105	-0.033***
Business	0.047	0.069	-0.022***	0.025	0.033	-0.008***
Others	0.058	0.064	-0.005	0.046	0.020	0.026***
Economic status						
Assets ownership	3.370	3.625	-0.255***	2.836	2.500	0.336***
Household without electricity	0.198	0.194	0.004	0.212	0.251	-0.039***
Household without toilets	0.703	0.613	0.090***	0.786	0.692	0.094***
Education						
Literate	0.564	0.635	-0.071***	0.525	0.563	-0.038***
Mean years of schooling	3.823	4.377	-0.553***	3.546	3.587	-0.041
Untouchability						
Household practice untouchability	0.162	0.123	0.038***	0.316	0.259	0.057***
SC household experience untouchability	0.232	0.186	0.046***			
Number of households	5087	778		1755	1146	

Notes: Table reports the summary statistics of our sample: house characteristics across village types. Observation are households in three types of villages where villages are categorised based on which social group is dominant: non-SCST, SC and ST. Last two columns reports difference in household characteristics between non-SCST and SC and non-SCST and ST villages respectively.

Table 17: Marginal effects of village dominance and caste on reading skills

	Cannot read at all			Can read a story		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.108***	0.108***	0.034**	-0.134***	-0.130***	-0.047**
	(0.022)	(0.021)	(0.015)	(0.023)	(0.023)	(0.020)
ST	0.137***	0.132***	0.037**	-0.148***	-0.145***	-0.050**
	(0.020)	(0.033)	(0.019)	(0.017)	(0.028)	(0.023)
OBC	0.053***	0.054***	0.009	-0.075***	-0.072***	-0.013
	(0.009)	(0.011)	(0.006)	(0.014)	(0.014)	(0.009)
VDSC		-0.006	-0.030*		0.006	0.046
		(0.033)	(0.018)		(0.048)	(0.032)
VDST		-0.006	-0.028		0.010	0.044
		(0.026)	(0.019)		(0.040)	(0.031)
VDOBC		0.003	-0.003		-0.002	0.005
		(0.012)	(0.010)		(0.016)	(0.014)
SC*VDSC		0.002	0.005		-0.002	-0.008
		(0.034)	(0.026)		(0.051)	(0.047)
ST*VDST		0.154***	0.035**		-0.167***	-0.056**
		(0.031)	(0.016)		(0.043)	(0.026)
OBC*VDOBC		0.018**	-0.008		-0.025**	0.011
		(0.009)	(0.009)		(0.012)	(0.014)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	7947	7947	7478	7947	7947	7478
R-sq	0.007	0.009	0.112	0.007	0.009	0.112
VDSC + SC*VDSC		-0.094***	-0.053***		0.106**	0.077**
		(0.033)	(0.020)		(0.047)	(0.030)
VDST + ST*VDST		0.016	-0.029		-0.012	0.038
		(0.036)	(0.025)		(0.027)	(0.031)
VDOBC + OBC*VDOBC		-0.037*	-0.020		0.044*	0.029
		(0.020)	(0.014)		(0.023)	(0.019)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on reading skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, OBC= Other Backward Castes, VDSC= Village dominated by SC, VDST=village dominated by ST, VDOBC=village dominated by OBC. *** p<0.01, ** p<0.05, * p<0.1.

Table 18: Marginal effects of village dominance and caste on arithmetic skills

	Cannot recognise numbers			Can do division		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.115***	0.116***	0.035***	-0.075***	-0.073***	-0.026***
	(0.020)	(0.021)	(0.012)	(0.012)	(0.013)	(0.008)
ST	0.171***	0.160***	0.055**	-0.091***	-0.089***	-0.038***
	(0.018)	(0.026)	(0.022)	(0.010)	(0.011)	(0.013)
OBC	0.067***	0.063***	0.006	-0.051***	-0.047***	-0.005
	(0.009)	(0.011)	(0.007)	(0.009)	(0.009)	(0.006)
VDSC		-0.009	-0.032**		0.007	0.026*
		(0.022)	(0.015)		(0.017)	(0.014)
VDST		-0.010	-0.032*		0.009	0.028
		(0.027)	(0.018)		(0.024)	(0.020)
VDOBC		0.014	-0.004		-0.009	0.003
		(0.016)	(0.012)		(0.012)	(0.009)
SC*VDSC		-0.008	-0.002		0.007	0.002
		(0.023)	(0.017)		(0.019)	(0.016)
ST*VDST		0.199***	0.059		-0.108***	-0.049
		(0.045)	(0.036)		(0.031)	(0.031)
OBC*VDOBC		0.032***	0.003		-0.023***	-0.002
		(0.011)	(0.008)		(0.009)	(0.007)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	7914	7914	7451	7914	7914	7451
R-sq	0.008	0.010	0.132	0.008	0.010	0.132
VDSC + SC*VDSC		-0.111***	-0.062***		0.067***	0.049***
		(0.029)	(0.017)		(0.023)	(0.015)
VDST + ST*VDST		0.029	-0.028		-0.009	0.018
		(0.045)	(0.029)		(0.015)	(0.018)
VDOBC + OBC*VDOBC		-0.021	-0.008		0.013	0.006
		(0.016)	(0.013)		(0.010)	(0.010)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on arthimetic skills, weighted using IHDS sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, OBC= Other Backward Castes, VDSC= Village dominated by SC, VDST=village dominated by ST, VDOBC=village dominated by OBC. *** p<0.01, ** p<0.05, * p<0.1.

Table 19: Teachers behaviour with same learning level

	Can read story			Can do division		
	Beaten	Scolded	Unfair treatment	Beaten	Scolded	Unfair treatment
SC	-0.015	-0.051**	-0.041**	-0.022	-0.047	-0.051
	(0.031)	(0.020)	(0.019)	(0.036)	(0.039)	(0.031)
VDSC	-0.141***	0.060	-0.027	-0.018	-0.080	-0.092***
	(0.047)	(0.093)	(0.055)	(0.103)	(0.101)	(0.032)
SCxVDSC	-0.073	-0.332***	-0.081	-0.201	-0.354***	-0.036
	(0.094)	(0.073)	(0.058)	(0.136)	(0.122)	(0.056)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1945	1968	2075	936	930	947
R-sq	0.256	0.229	0.191	0.261	0.338	0.19
VDSC + SC*VDSC	-0.190**	-0.180***	-0.061**	-0.164**	-0.329***	-0.079**
	(0.076)	(0.055)	(0.026)	(0.077)	(0.057)	(0.031)

Notes: This table reports the marginal effects from probit model of village dominance on probability of being beaten up, scolded and unfair treatment, weighted using IHDS sample weights. Observations are rural children aged 8-11 years who can read stories and do division. Robust standard errors, clustered at village level are reported in parentheses. Controls include age of a child, gender of a child, grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), gender of a teacher, and location of a teacher. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 20: Marginal effects of village dominance and caste on reading skills, using 2005 IHDS

	cannot read at all			can read a story		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	0.067***	0.070***	0.032***	-0.098***	-0.101***	-0.053***
	(0.010)	(0.010)	(0.007)	(0.010)	(0.010)	(0.010)
ST	0.048***	0.053***	0.011	-0.069***	-0.078***	-0.020
	(0.006)	(0.014)	(0.013)	(0.009)	(0.014)	(0.021)
VDSC		-0.010	-0.009		0.011	0.012
		(0.021)	(0.012)		(0.035)	(0.023)
VDST		-0.050***	-0.006		0.107***	0.010
		(0.013)	(0.023)		(0.035)	(0.044)
SC*VDSC		0.005	-0.005		-0.009	0.009
		(0.026)	(0.014)		(0.046)	(0.028)
ST*VDST		0.095***	0.021		-0.176***	-0.036
		(0.011)	(0.028)		(0.035)	(0.052)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	8471	8471	7712	8471	8471	7712
R-sq	0.005	0.0064	0.132	0.005	0.006	0.132
$\beta_3 + \beta_5$		-0.062**	-0.036**		0.085**	0.061**
		(0.025)	(0.018)		(0.036)	(0.031)
$\beta_4 + \beta_6$		-0.008	0.004		0.009	-0.006
		(0.023)	(0.024)		(0.027)	(0.039)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on reading skills, weighted using IHDS 2005 sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index, and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.

Table 21: Marginal effects of village dominance and caste on arithmetic skills, using 2005 IHDS

	Cannot recognise numbers			Can do division		
	model 1	model 2	model 3	model 1	model 2	model 3
SC	-0.098***	0.105***	0.051***	-0.081***	-0.085***	-0.046***
	(0.010)	(0.016)	(0.008)	(0.012)	(0.013)	(0.008)
ST	-0.069***	0.126***	0.021	-0.079***	-0.090***	-0.020
	(0.009)	(0.023)	(0.016)	(0.017)	(0.014)	(0.015)
VDSC		-0.041***	-0.023		0.039**	0.023
		(0.015)	(0.016)		(0.016)	(0.020)
VDST		-0.007	0.036*		0.006	-0.034*
		(0.021)	(0.021)		(0.020)	(0.018)
SC*VDSC		0.041	0.034		-0.045	-0.036
		(0.035)	(0.030)		(0.037)	(0.033)
ST*VDST		0.090***	-0.029		-0.072***	0.025
		(0.031)	(0.031)		(0.022)	(0.027)
Controls	No	No	Yes	No	No	Yes
State fixed effects	No	No	Yes	No	No	Yes
N	8442	8442	7,681	8442	8442	7,681
R-sq	0.007	0.008	0.147	0.007	0.008	0.147
$\beta_3 + \beta_5$		-0.093***	-0.036**		0.071**	0.031**
		(0.030)	(0.015)		(0.028)	(0.013)
$\beta_4 + \beta_6$		-0.043	-0.012		0.024	0.011
		(0.042)	(0.024)		(0.025)	(0.022)

Notes: This table reports the marginal effects from Oprobit model of caste and village dominance on arithmetic skills, weighted using IHDS 2005 sample weights. Observations are rural children aged 8-11 years. Robust standard errors, clustered at village level are reported in parentheses. Controls include grade in which child is studying, parents education, parents occupation, economic status of household, type of school (private or public), health of child and mother, whether a child get any incentive (free uniforms and books) from school, village development index, and caste of village head. SC=Scheduled Castes, ST=Scheduled Tribes, VDSC= Village dominated by SC, VDST=village dominated by ST. *** p<0.01, ** p<0.05, * p<0.1.