Tribal Identity in Education: Does Teacher Ethnicity Affect Student Performance?

Girija Borker*

Brown University

March 2017

[Latest Version]

Abstract

Reducing achievement gaps between minorities and non-minorities by raising the achievement of minority students is considered a critical component of promoting inclusive growth, both in terms of human capital and economic well-being. Using large scale administrative panel data for all schools in Jharkhand, I study the effect of having a co-ethnic teacher on the extensive margin of student enrollment. In particular, I estimate the impact of having a tribal teacher on enrollment of tribal students. I find that co-ethnic teachers are 0.44 standard deviations per year more effective at retaining students relative to a teacher belonging to a different ethnic group. Specifically, increasing the proportion of tribal teachers by a standard deviation increases tribal enrollment by 0.57 standard deviations. This effect is persistent and greater than the caste and gender teacher effects. I also develop a theoretical model which identifies whether the observed effects are caused by a decrease in discrimination by the teacher or from an increase in the marginal return to teacher effort, from switching to a co-ethnic teacher.

Keywords: Education, Effect of teachers, India, Tribal identity, Ethnicity.

JEL classification: I21, J15, O15

*Preliminary version. I am grateful to Andrew Foster, Kaivan Munshi and Sriniketh Nagavarapu for their help and advice. I also thank Rai Mahimapat Ray for helping with the data collection and organizing visits to schools in rural Jharkhand. Geospatial data on the location of schools was kindly provided by the Jharkhand Space Application Centre. I also appreciate the helpful comments from participants at the NEUDC Conference, the ICODECON Conference, the seminar at Watson Institute for International Studies, the seminar at Population Studies and Training Center, and the development economics breakfast at Brown University. Research grants from the PSTC and Graduate Program in Development are gratefully acknowledged. All errors are my own.

1 Introduction

The large and persistent achievement gaps separating minorities and non-minorities are one of the most important problems in education today. These achievement gaps have long term effects that reinforce the economic disparity between minorities and non-minorities. Hence reducing, and ultimately eliminating, these gaps at an early stage by raising the achievement of minority students is critical in promoting inclusive growth both in terms of human capital and economic well being. Previous research suggests that differences in observable characteristics do not have sufficient explanatory power in explaining these gaps, and that explanations are more likely to come from more nuanced hypotheses about the dynamics within families, schools and classrooms (Jencks and Phillips 1998). One such explanation is based on the most fundamental relationship in school, that between a student and a teacher, which affects achievement of students. The aim of this paper is to estimate how the ethnic interactions of a student and a teacher affect cognition in school.

While the effect of demographic interactions between students and teachers has been widely studied in developed countries the evidence for developing countries is very limited. Moreover, all of these studies focus on the aggregate impact of these demographic interactions and are unable to identify the exact mechanism behind the observed effects. In this paper I study the effect of having a co-ethnic teacher on the extensive margin of student enrollment in India. In particular, I estimate the impact of having a tribal teacher on enrollment of tribal students using a large scale administrative dataset from the state of Jharkhand, India. I also develop a theoretical model that identifies whether the observed effects are caused by a decrease in discrimination (*active teacher effects*) or from an increase in the marginal return to teacher effort (*passive teacher effects*) from switching to a co-ethnic teacher.

For this study I use longitudinal data on enrollment at the school-grade level over five years for *all* of the 47,000 schools in Jharkhand. The dataset also contains detailed information on teacher characteristics. The combination of panel data and variation in the ethnicity of the pool of teachers for each grade allows me to estimate the casual impact of an ethnicity match between students and teachers. Identification concerns are addressed by showing that the causal estimates do not change substantially under increasingly restrictive set of specifications including district fixed effects and village controls, school and school-grade fixed effects and additional teacher characteristics. Based on qualitative evidence I also show that teachers are not assigned to schools or to grades within schools on the basis of student ethnicity.

I find that teachers are 0.44 standard deviations (SD) per year more effective at retaining co-ethnic students relative to a teacher belonging to a different ethnic group. This effect is cumulative and persists over time. Hence, tribal teachers are 0.44 SD more effective at reducing the ethnic gap in enrollment relative to non-tribal teachers. This result is robust to allowing for heterogeneity by teacher characteristics which are correlated with teacher ethnicity and teacher quality. In addition to co-ethnic students, the overall effectiveness of a teacher also depends on her ability to teach students belonging to a different ethnic group. The second result indicates that tribal teachers are unable to retain non-tribal students. Specifically, increasing the likelihood of having a tribal teacher by 100 per cent decreases non-tribal enrollment by 0.24 SD per year while it increases the enrollment of tribals by 0.57 SD. A third of the decrease in enrollment is due to students dropping out and the remaning from students' switching to other schools in the neighborhood. These results do not differ across gender of students or grade type.[1] Finally, I see that ethnicity match effects are stronger than the gender match and caste match effects in this sample.

My work relates to several strands of literature. First, this study relates to the literature on how demographic matches can improve economic outcomes more generally. Studies have found that cultural proximity between two agents improves efficiency of market outcomes in several contexts. In the Indian credit market, borrowers who had a loan officer from the same social group as them enjoyed increased access to credit, reduced collateral requirements and showed better repayment records (Fisman et al. 2012). Public service delivery in India has been found to be more efficient when the seller and the buyer belong to the same caste group (Nagavarapu and Sekhri 2014). These studies focus on improved informational dynamics and enforcement as drivers of the observed effects. In contrast, in my theoretical model I explore the absence of discrimination and increased complementarity in the form of increased marginal return to effort as the main mechanisms underlying the positive aggregate effects from having a co-ethnic teacher.

Second, my results are consistent with the various "teacher like me" studies which assess the effect of matching student and teacher characteristics like gender, race and ethnicity. In the case

[1] Grade type include Primary Grades (1-5) or Upper Primary Grades (6-8).

of gender match, studies in the US (Dee 2007, Dee 2005) and in India (Muralidharan and Sheth 2015,[2] Kingdon and Rawal 2010) have found an improvement in test scores, teacher perception and engagement in class for girls when taught by a female teacher in school. Most of the studies examining the effect of shared race and ethnicity are based in the US and there is little rigorous evidence available for developing countries. These studies have found large positive effects on academic performance in school (Dee 2004, Kingdon and Rawal 2010), drop out rates, subjective evaluation of students (Ehrenberg et al. 1995, Dee 2005) and long term outcomes such as course selection and degree completion in community college (Farlie et al. 2011). Most of the papers in this literature employ a student fixed effects approach to use the variation of teacher ethnicity within a student across subjects in order to identify the ethnicity match effect. In such an estimation strategy it can be difficult to interpret the magnitudes of the estimated effects without knowing the ethnic composition of the teachers in a subject in previous grades.[3] In this paper, the use of five years of panel data for all primary and upper primary schools in the state of Jharkhand allows me to estimate the impact of an ethnicity match on enrollment in a year. I am also able to look at the long term effects of an ethnicity match and ascertain whether the effects are cumulative or not. Additionally, unlike most of the other studies, I observe students in elementary school (Grade 1 to Grade 8), when they are young. Not only is this age range important because early childhood education is critical for cognitive development (NSCDC 2007) but also at such a young age the role of shared ethnicity is likely to be especially important for their decision to continue in school. This is also the age range for which there is free and compulsory schooling provided by the Government in India and is most relevant for policy design to reduce education gaps between minority and non-minority kids in developing countries since the majority of students do not complete more than eight years of school education.[4]

Third, this study relates to the burgeoning literature on the effects of teacher quality on student

[2]Muralidharan and Sheth extend the approach of this paper in the gender context by actually looking at student learning and attendance, hence being able to look at both the intensive and the extensive margin in one unified framework (though they eventually find no effect of a gender match on the extensive margin).

[3]As pointed out by Muralidharan and Sheth (2015), based on this approach if I find that a tribal student in 5th grade who has a tribal language teacher and a non-tribal math teacher, does better in language then the interpretation of the point estimate is confounded by the possibility that the tribal student is also more likely to have had tribal language teachers in earlier grades.

[4]For instance, in India 41% children drop out before reaching Grade 8 and the drop rate is even higher at 55% for tribal children (Statistics of School Education 2010-11, Ministry of Human Resource Development, GoI).

outcomes. There is wide consensus among policy makers, researchers and parents that teachers are one of the most important institutional component of an academic environment. However, there is considerable debate on the specific characteristics of a teacher that affect student learning. Assigning teachers a "value-added" measure of quality based on student test scores, studies in the US have shown that having a teacher at the 85th percentile of the "quality distribution" vs. one at the 15th percentile leads to 8-20 percentile points higher test-scores in math and reading (Rivkin et al. 2005, Aaronson et al. 2007, Kane and Staiger 2008, Chetty et al. 2014b). Similar effects, based on the value-added quality measure, have been found for students' non-cognitive ability (Jackson 2012) and their long-term outcomes like probability of attending college and wages (Chetty et al. 2014b). Beyond the value-added measure, researchers have found mixed results for the impact of standard resume teacher characteristics such as experience, tenure and educational credentials on student outcomes (Muralidharan and Sundararaman 2013, Harris and Sass 2011, Clotfelter et al. 2005, Clotfelter et al. 2007, Aaronson et al. 2007).[5] Aslam and Kingdon (2007) find pedagogical practices such as lesson planning and questioning students during class to be more important for student learning as compared to the traditional human capital measures. In my study I specifically look at the effects of a salient demographic characteristic like the ethnicity of the teacher on enrollment, conditional on other observable teacher characteristics.

My work expands the existing literature in several ways. First, to my knowledge this is the first study that develops a theoretical model which disentangles two well-documented potential mechanisms behind an aggregate match effect between students and teachers. The model is based on two assumptions: that discrimination by a teacher is triggered solely by the student's tribal ethnicity, independent of teacher and student ability. And that the complementarity effect or increased marginal return to teacher effort is likely to interact with both student and teacher ability. The theoretical model generates predictions that, based on the simulation results, are likely to be testable for a wide range of parameter values. Identification of the exact mechanisms behind the aggregate match effect is important for policy. If it is biases in teacher behavior that manifest themselves in overt discrimination against *out-group* students then the key policy recommendation would be very different than if it is the positive role model effects that are at play.

[5] See Glewwe et al. (2013) for a review of studies from 1990-2010 assessing the effect of various teacher characteristics on student achievement in developing countries.

Second, the results of this study are based on a larger sample of schools than any of the previous studies in a developing country. My sample consists of the universe of primary and upper primary schools in the Indian state of Jharkhand. Hence, the results can truly speak to the effects that matter for policy making for tribal students in India. Third, for this study I have constructed a rich data set in which one can track cohorts of students over time, in all the schools in Jharkhand. My data also has very detailed information on the school teachers, something that is hard to find in a low-income setting and has the potential to empirically answer some key policy questions.

The remainder of this paper proceeds as follows: Section 2 provides a brief background on education in India, the tribal population and the state of Jharkhand. Section 3 lays out the conceptual framework including the simple theoretical model and simulation results. Section 4 describes the data and Section 5 discusses the empirical strategy. Section 6 presents the results and finally Section 7 concludes.

2 Background

India has one of the largest elementary schooling systems in the world, with around 200 million children enrolled across a million schools.[6] However, it is still far behind in achieving the United Nations' Millennium Development Goal of "Universal Elementary Education".[7] As per the United Nation's Initiative on Education, an estimated 8.1 million children between 6-13 years are out of school in India and 41 per cent of school-going children drop out before reaching Grade 8.[8] In terms of its structure, the Indian education system can be divided into three main categories: Elementary Education, Secondary Education and Higher Education based on the age of students enrolled. Children in the age group of 6-14 are a part of the Elementary level of education ranging from Grades 1-8, and are entitled to free and compulsory education as mandated by the Right of Children to Free and Compulsory Education Act, 2009 (RTE). The Elementary level is further divided into two sub-categories, Primary School (Grades 1-5) and Upper Primary School (Grades 6-8). Post Elementary level, a child undergoes Secondary Education for 4 years (Grades 9-12).[9]

[6] All India U-DISE data (2013-14)

[7] Universal Elementary Education was mandated by the 86th amendment to the Constitution of India after 2000-01.

[8] School Education in India: A Handbook (2015).

[9] Secondary education is further split into Secondary (Grades 9-10) and Higher Secondary (Grades 11-12).

India is also the country with the longest history of affirmative action for historically disadvantaged groups including tribals. Tribal groups are considered to be the earliest inhabitants of the country. The state and discourse in India use the term Scheduled Tribes (ST) to refer to these tribal groups. In this study, the term tribals refers to this political-administrative category of ST as identified by the Constitution.[10] The Government of India today identifies 533 tribes of which 32 are located in the state of Jharkhand.[11] While there is a lot of diversity within this social group in terms of language, extent of acculturation, habitation and level of development, the "tribal" identity continues to be salient.[12]

In India, the "*Adivasis*" or tribals are a marginalized minority group. Although numerically only about 8.6 per cent of the country's population (2011), they disproportionately represent the people living below the poverty line, suffer from poor physical health and have some of the worst educational outcomes. Over 90 per cent of tribals live in rural areas and 80 per cent of all tribals workers are engaged in agriculture related activities.[13] The poverty rate among tribals in 2004-05 was 45 per cent compared to a rate of 37 per cent amongst the other historically disadvantaged group of Scheduled Castes (SC) and a population-wide rate of 28 per cent.[14] In terms of health outcomes, the child mortality rate for tribals is 35.8 deaths per thousand children which is almost double of the country-wide average of 18.4 deaths per thousand (2005-06).[15]Additionally, only 17.7 per cent of pregnant tribal women have an institutional delivery compared to an average of 38.7 per cent of women in the country (2005-06).[16] Tribals also have the lowest literacy rate across various social groups in India with only 59 per cent of them being able to read and write compared to 66 per cent of SCs and an overall literacy rate of 73 per cent in India (2010-11).[17] Moreover, over 33 per cent of tribal children drop out of school before they complete Grade 5, compared to 17 per cent of SC children and a country-wide drop out rate of 20 per cent (2010-11).[18] In fact, the

[10]Article 366 (25) of the Constitution of India refers to STs as those communities, who are scheduled in accordance with Article 342 of the Constitution. These groups of people are identified based on five criteria, including their primitive traits, distinctive culture, limited contact with the outside community, geographical isolation and backwardness (Lokur Committee, Government of India, 1965).

[11]http://www.tribal.nic.in/

[12]Ministry of Tribal Affairs, Government of India, 2014.

[13]Census of India, 2011.

[14]Ahluwalia, 2011.

[15]National Family Health Survey, 2005-06.

[16]*Ibid.*

[17]Statistics of School Education, 2010-2011.

[18]*Ibid.*

state of education for tribals has remained in such a dismal condition for so long that the 12th Five Year Plan (2012) cites tribal exclusion as “the single most important challenge in universalizing elementary education [in India]”.

This study is based in Jharkhand, one of the youngest states in the country. Jharkhand was carved out of the state of Bihar in 2000 fulfilling the demand of tribals for separate statehood. Located in eastern India, over 26 per cent of the population of Jharkhand is tribal.[19] From Figure 1, we can see that apart from some areas in the north,[20] tribals are spread out fairly evenly across the state. Despite being endowed with 40 per cent of India’s mineral resources Jharkhand is one of the poorest and most underdeveloped states in the country. Over 13 million (41 per cent) people of a population of 32 million people in the state are classified as poor.[21] More than a third of the population is illiterate[22] and almost 60 per cent of the children (0-3 years) are underweight.[23] Overall, Jharkhand is ranked 19 out of 23 in India in terms of the Human Development Index.[24] Jharkhand is also one of the most education-poor states in India, in 2012-13 it was ranked 34 out of 35 states and union territories in India in its elementary education outcomes.[25] As of 2014, there are still over 300 thousand 6-14 year old children who are out of school.[26] Of those *in* school, 55 per cent of students in Grade 3-5 cannot read a Grade 1 level text and their arithmetic learning is even worse, with around 60 per cent of students in Grade 5 not being able to do basic subtraction.[27]

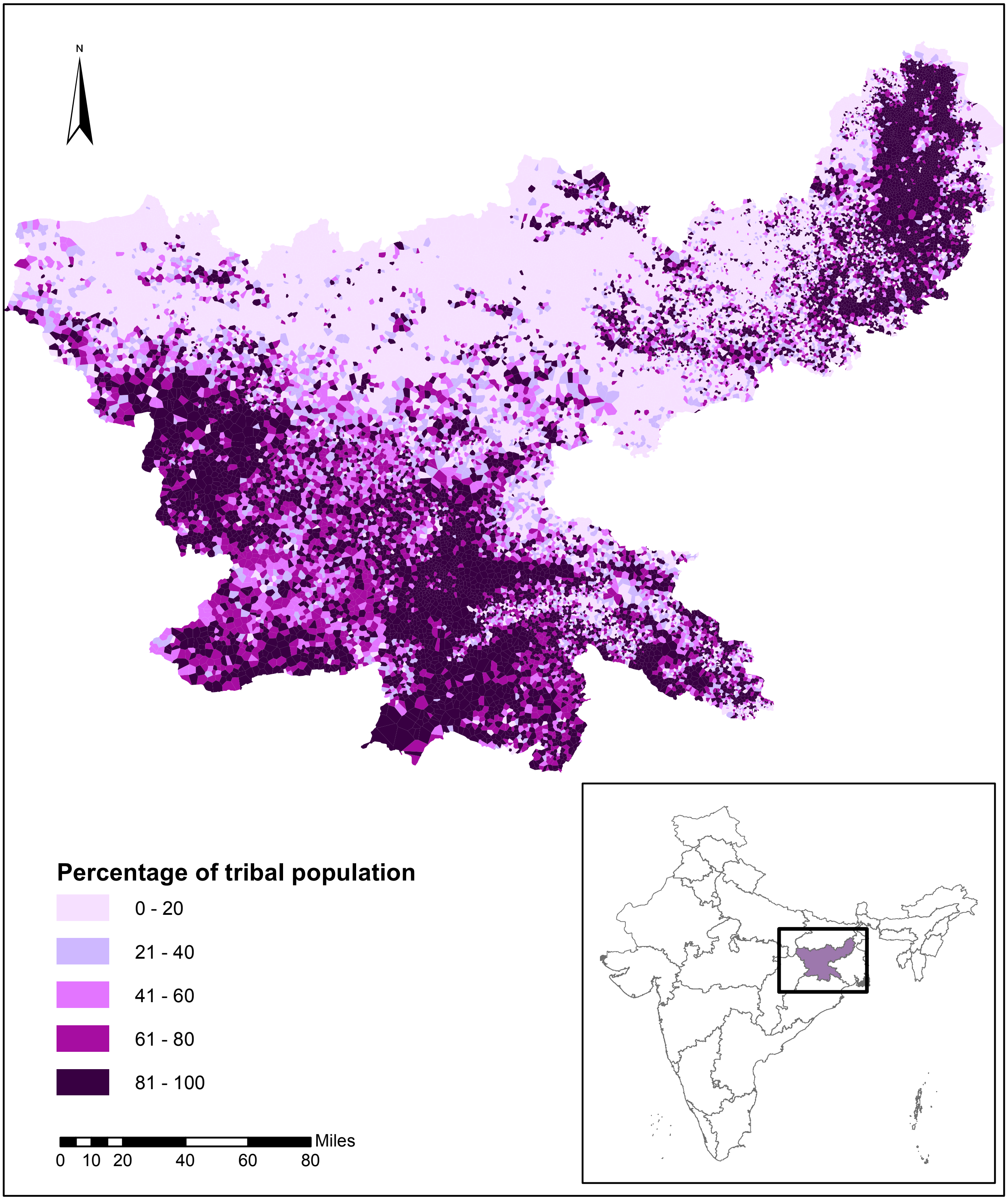
Hence this study is especially important from a development point of view, as I analyze how to improve the educational outcomes of a marginalized minority group in one of the poorest parts of the world.

3 Conceptual framework

The extant literature suggests two main ways in which demographic matches between teachers and students can influence educational outcomes (Dee 2005). One set of explanations involve *passive*

[19] Census of India (2011).
[20] North Jharkhand is densely populated with SCs.
[21] UNDP 2011.
[22] Census of India 2011.
[23] NFHS 2005-06.
[24] UNDP 2011.
[25] National University Education Planning and Administration, Government of India.
[26] http://www.jepc.nic.in.
[27] ASER 2014.

Figure 1: Percentage of tribal population by village in Jharkhand (2011)



Note: This is based on population data from the Census of India, 2011. The village boundaries are created based on the centroid of villages from the 2001 Census.

teacher effects. These effects are simply triggered by a teacher's racial, ethnic, or gender identity and not by any explicit teacher behaviors. These include "role model" effects, according to which students are more engaged, behave more appropriately and perform better when taught by a demographically similar teacher. These role models may provide children with examples of educated, economically well-off, successful individuals of their gender, ethnicity or social background and thereby improve their attitudes towards education and effort in school (Kingdon and Rawal 2010). Passive effects can also include effects which raise the marginal product of teacher effort such as the unintended effect of speaking the same language as a student, which helps the student better understand the teacher. A negative type of passive teacher effect is the threat of being perceived as a negative stereotype where the fear of poor performance causes apprehension that retards academic performance and ends up confirming the stereotype.[28]

A second class of explanations for the educational benefits of co-ethnic teachers are the *active teacher effects*, these include discrimination against students with different ethnicity. In the Indian context, there is significant evidence of discrimination against minority children by non-minority teachers. A recent report by the Human Rights Watch (2014) exposes the ways in which minorities are discriminated against in schools. For example, in case of tribal children, the report points to instances where the non-tribal teachers made the tribals sit separately from the rest of the class, they ridiculed the tribal children by saying that "their main aim is to come [to school] and eat [the mid-day meal], not to study" and commented on their appearance, remarking that "they are so dirty... they smell". Such discriminatory behavior is triggered purely by the student's ethnicity.

As discussed in Section 1, there have been several studies in developed countries and a few even in developing countries that examine the effect of a demographic match between student and teachers on different indicators of student performances. However, there is no study to my knowledge which has been able to distinguish between the two mechanisms behind these aggregate match effects. An understanding of the exact mechanism driving the ethnic-match effect is not only important to establish true causality but also crucial for policy design. Below I present a simple model which allows me to disentangle the mechanisms. In the model I am able to distinguish between the active effects or the effect of discrimination from having a teacher that does not belong

[28] Another related effect is the Pygmalion effect, which is a kind of self-fulfilling prophecy where initial expectations of a teacher influence ultimate performance by the student.

to your ethnicity[29] and the passive effects or increased complementarity[30] from having a co-ethnic teacher. The model is based on the assumption that the effect of discrimination by a teacher would be purely induced by a student's ethnicity, independent of student and teacher ability while the effect of increased complementarity due to the ethnicity-match would be correlated with both student and teacher ability. To help clarify the predictions I also simulate the model and find that I am able to disentangle the two mechanisms for a reasonable set of parameter values. At present, I am unable to test the predictions of the model as the required data did not come through in time. I hope to get the data this summer. Hence, in the empirical section we restrict our attention to estimating the aggregate effects of having a co-ethnic teacher.

3.1 Setting

In the model, I derive predictions based on a one-to-one match between a teacher and a student. The parent decides whether to enroll the child in school or not and only cares about her child's learning. A teacher's value function depends on the learning of a student less the cost of effort in achieving that amount of learning, given her own ability and the ability of the student.

In terms of the timing in the model, first a teacher is allotted to a school-grade, parents (of children already enrolled in school) can observe this choice and all observable characteristics of the teacher including her ethnicity and choose whether to enroll their kids in that grade. This decision is taken every year based on the expected learning of their child. Once the enrollment process is complete, a teacher decides the amount of effort to exert for each student based on the student's fully observable ability and their own cost of effort.

The model does not capture the choice of school when the parents are first taking the decision to enroll their child. Empirically this is not a concern as most parents in rural India send their children to a school in the village of residence, in my sample 71 per cent of villages have up to one primary school (2013-14).[31] The model also abstracts from the public good nature of teaching by looking at a one-to-one match of teachers and students versus matching a teacher to a class of

[29] These can also be understood as "push factors" or the actions of demographically different teachers that push students away.

[30] These can also be thought of as "pull factors" or the factors that attract students to a demographically matched teachers.

[31] Additionally, 90 per cent of the villages in my sample have no more than two primary schools.

students.[32]

Household's problem

The household is assumed to consist of a parent and a child. The parent decides whether the child goes to school or not based on her learning in school relative to the opportunity cost of her time. This household's problem simply captures the trade-off between the benefit from human capital investments in the child and the opportunity cost of not having her help out at home or in the field. I expect the opportunity cost to be increasing in the age of the child, since an older child is likely to have greater access to child labor opportunities or be more capable of helping at home by looking after her younger siblings.

A household's problem takes the form:

$$enroll\ if\ L(a,e) > o(y)$$

where $L(a,e)$ is the learning of a student as a function of student's (exogenous) perfectly observable ability a and the expected teacher effort e. Parents are able to perfectly predict the learning of their child knowing that a teacher will choose her value maximizing effort level.[33] The opportunity cost $o(\cdot)$ is assumed to be an increasing function of age of the child y.

Teacher's problem

In the model, a teacher maximizes her value function equal to the learning of the student less her cost of effort:

$$V = L(a,e) - c(A)e = \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)e^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

[32] I am working on extending this model to (1) incorporate an effort budget constraint for the teacher, $\sum_{i=1}^{n} e_i = E$, where the sum of the effort exerted on each student must sum to the total effort a teacher can exert, for example E could denote the hours a teacher spends in school and (2) capture the public good nature of teaching where the effort exerted by the teacher for one child will yield learning for other children in the classroom. In the latter case, the preliminary results indicate that the results in addition to being a function of student and teacher ability, will also depend on the ability and ethnicity composition of the classroom.

[33] This assumes that any shocks to the teachers are perfectly observed by the parents, which is not unrealistic since over a third of the teachers in the data live in the same village as where the school is located and as noted above around 71 per cent of the villages in the sample only have one primary school. Shocks during the school year , i.e. after the parent has made her choice, are assumed to be zero in the model.

where $L(a,e)$ is the learning of a student as a function of student's perfectly observed ability a and the teacher's effort e. Exerting effort imposes a per unit cost of $c(A)$ that is a declining and concave function of teacher's own ability A. I assume a Constant Elasticity of Substitution (CES) learning production function,[34] where the two inputs, a and e, are gross complements i.e. $\sigma < 1$ indicating that ability of a child and teacher effort generate better learning as a combination than separately or that higher a should be matched with a higher e to be effective.[35] α and $(1-\alpha)$ represent the factor shares of student ability and teacher effort, respectively and $\alpha \in (0,1)$.

The learning function exhibits positive and diminishing marginal return in both student ability and teacher effor and there is direct complementarity[36] between the two inputs in the learning function.

Below, I start with the benchmark model and then separately introduce active effects in the form of discrimination and passive effects in the form of increased complementarity.

3.2 Benchmark model

First, I solve for the optimal effort exerted by the teacher and that combined with the exogenous ability of each student gives me the optimal learning for each student. Then based on this $L(a,e^*)$, a parent makes her choice.

The teacher's problem is:

$$Max_e V = \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)e^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

[34] I chose a CES production because of its previous use as the human capital production function (Cunha and Heckman 2007, Cunha and Heckman 2008) and its flexible functional form which nests different types of production functions, Cobb-Douglas ($\sigma = 1$), perfect substitutes ($\sigma \to \infty$) and Leontief ($\sigma = 0$).

[35] Technically, the inputs being gross complements means that the supply of e decreases in response to a increase in the price of a (say because parents start working longer hours which decreases the time they spend at home teaching the child which increases the opportunity cost or price of ability) holding the price of e and the quantity of a constant if and only if $\sigma < 1$, and vice versa (Acemoglu 2002).

[36] As defined by Cunha and Heckman (2008).

Solving the first order condition $\frac{\partial V}{\partial e} = 0$ I get the optimal effort level e^*:

$$e^* = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\left(\left(\frac{c(A)}{1-\alpha}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}} \tag{1}$$

In the benchmark case I find that higher ability teachers work harder in equilibrium and teachers exhibit reinforcing as opposed to compensating behavior by exerting more effort with higher ability students.[37] In terms of student learning, learning in equilibrium is higher for higher ability students, the learning generated by a higher ability teacher is greater than the learning generated by a lower ability teacher and the marginal learning from teacher ability increases with student ability in equilibrium.

3.3 Incorporating discrimination (active effects)

I now augment the model to incorporate active teacher effects in the form of discrimination by the teacher. Discrimination is introduced as a pure discount on the learning of a student who does not belong to the teacher's ethnic group. The intuition behind this structure is that if a teacher discriminates against a student based on their ethnicity, then the ability of the child would not matter and she would attach a lower weight to the learning of that student in her value function. Hence, with discrimination the teacher's value function takes the following form:

$$V_\lambda = \lambda L(a,e) - c(A)e = \lambda\left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)e^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

where $\lambda \in (0,1)$ is the discrimination parameter. $\lambda = 1$ in the benchmark case with no discrimination and λ keeps decreasing ($\rightarrow 0$) with an increase in discrimination. Solving teacher's maximization problem $\frac{\partial V_\lambda}{\partial e} = 0$ yields the optimal effort with discrimination e^*_λ:

$$e^*_\lambda = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}} \tag{2}$$

In the case with discrimination I have:

[37] The latter is consistent with the evidence found by Duflo et al. (2011) in Kenya.

1. Teacher effort

   - e^*_λ is decreasing in discrimination ($\frac{\partial e^*_\lambda}{\partial \lambda} > 0$).
   - When matched on the basis of ethnicity, the maximum gain from the decline in discrimination accrues to the *highest* ability co-ethnic students ($\frac{\partial^2 e^*_\lambda}{\partial a \partial \lambda} > 0$).
   - Given high enough elasticity of substitution between teacher effort and student ability, higher ability teachers reduce their effort more when they discriminate than lower ability teachers for any given student ability in equilibrium ($\frac{\partial^2 e^*}{\partial A \partial \lambda} > 0$ for $\sigma \geq \frac{1}{2}$). This could be just because the lower ability teachers any way exert such low effort in the benchmark case that they have limited scope to further reduce their effort.

2. Student learning

   - As expected the marginal learning of student ability is decreasing with discrimination ($\frac{\partial^2 L^*_\lambda}{\partial \lambda \partial a} > 0$) i.e. marginal learning for a student of any ability is less when she is discriminated against because of the lower teacher effort received in equilibrium.
   - The marginal learning of teacher ability is decreasing with discrimination ($\frac{\partial^2 L^*_\lambda}{\partial \lambda \partial A} > 0$) or that a teacher of any ability generates less marginal learning when she discriminates against a student of ability a relative to when she does not discriminate.
   - The marginal learning for any teacher and student ability combination increases with an decrease in discrimination ($\frac{\partial}{\partial \lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A \partial a}\right) > 0$). Intuitively, the interaction effect between student and teacher ability on learning *increases* with a decrease in discrimination. While analytically I could only find a sufficient condition for this to be true (detailed in the Mathematical Appendix), numerically I find that it holds for a wide range of values of σ.

Prediction 1: $\frac{\partial}{\partial \lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A \partial a}\right) > 0$ ***implies that if the main mechanism behind the positive match effect is a decrease in discrimination then I should see a positive interaction effect between student and teacher ability for co-ethnic teachers on enrollment.***

To help clarify the predictions I simulate the model. For this I assume that both a and A lie in the range $[0,3]$ with lower values indicating lower ability. The cost function is of the form

$c(A) = -A^2 + b$, where b is a constant. $c(\cdot)$ is decreasing and concave for all values of A. We set $b = 10$. Based on these values, I find that Prediction 1 holds for a range of values of σ for every α.

In Figure 2 I plot the marginal learning of teacher ability as discrimination decreases ($\uparrow \lambda$). First, the reader should note that $\frac{\partial L}{\partial A}$ takes on positive values and it has a positive slope which shows that it is increasing with an decrease in discrimination ($\frac{\partial^2 L^*_\lambda}{\partial \lambda \partial A} > 0$). The gains from a decrease in discrimination are greater for higher ability students than for lower ability students, as is captured by the steeper slope for the high ability students compared to the low ability students. This change in slope is $\frac{\partial}{\partial \lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A \partial a}\right)$ as highlighted in Prediction 1.[38] I find the same result for a range of values for σ for every α. For example, for a low α such as $\alpha = 0.1$ (as in Panel (a) in Figure 2) I can distinguish between discrimination and complementarity as long as $\sigma \in (0.33, 0.5]$ and for a high α like $\alpha = 0.9$ (as in Panel (b) in Figure 2) the range is $\sigma \in [0.49, 0.82]$.

3.4 Adding complementarity (passive effects)

For the passive effects, we focus on increasing the marginal return to effort for the teacher. Specifically, I add an effort augmenting technology (ϕ) to the learning production function such that a any effort e by a teacher generates learning worth ϕe. With the complementarity factor $\phi > 1$ the value function of the teacher takes the following form:

$$V_\phi = L(a, \phi e) - c(A)e = \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(\phi e)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

where I have the benchmark case if $\phi = 1$. As ϕ increases the marginal return from teacher effort increases. I am able to draw clear predictions from the model in the case where σ is small, specifically, σ satisfies $\frac{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - (1-\alpha)}{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}} - \sigma > 0$.[39] With a low elasticity of substitution, I have the following expression for e^*_ϕ:

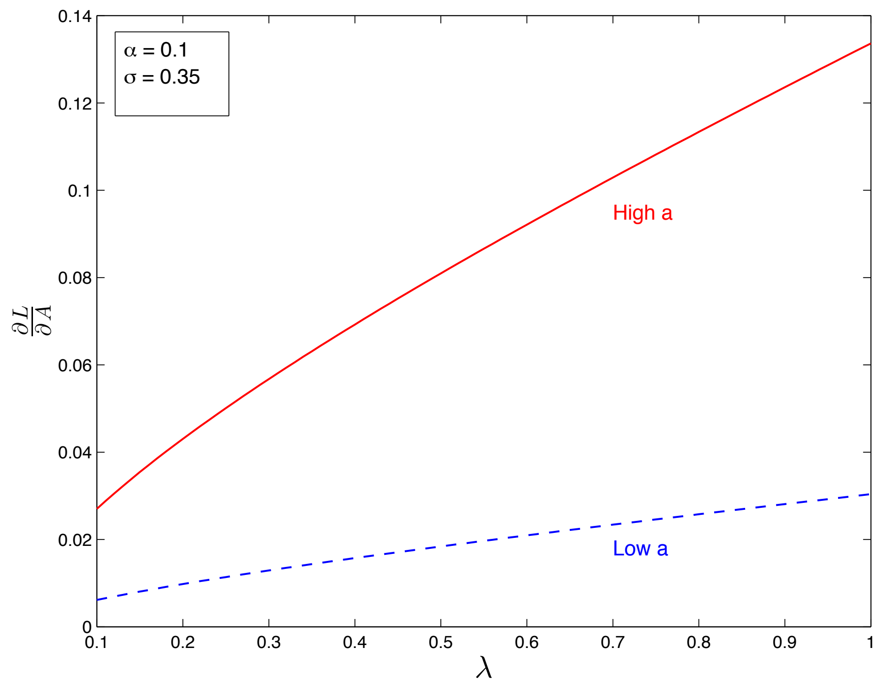
$$e^*_\phi = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\phi\left(\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}}$$

[38]Please see Figure A1 in the Appendix for the graph for $\frac{\partial}{\partial \lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A \partial a}\right)$ for the same parameter values.

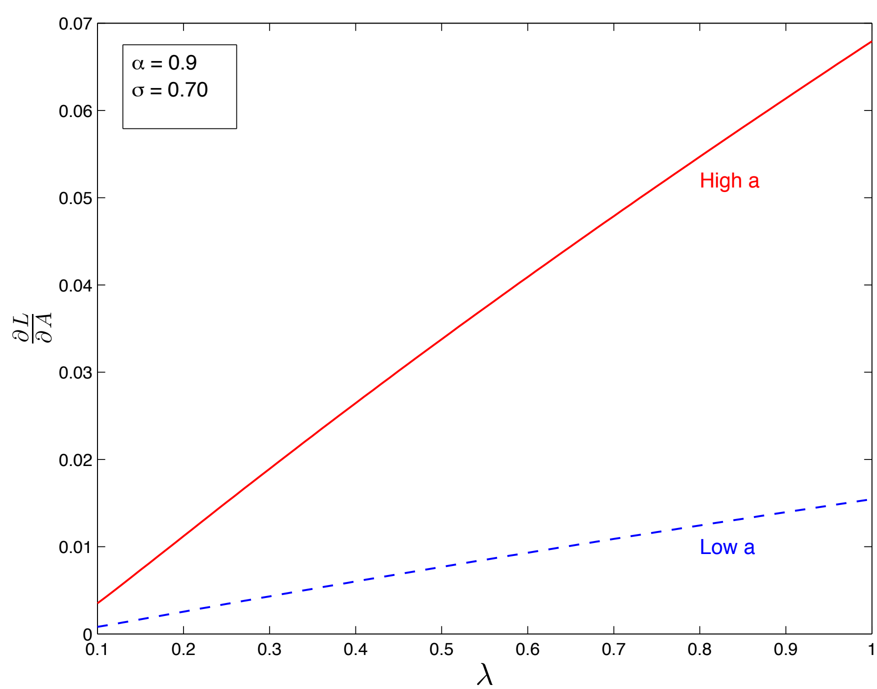
[39]Please refer to the Mathematical Appendix for the computation.

Figure 2: Change in the marginal learning of teacher ability with a decrease in discrimination for different levels of student ability

(a) The effects with a low factor share of student ability (Low α)



(b) The effects with a high factor share of student ability (High α)



The extreme case of $\sigma = 0$ where student ability and teacher effort are perfect complements, provides good intuition for the comparative statics to follow. When $\sigma = 0$ the learning function is a Leontief production function: $L(a, \phi e)$=$min\{a, \phi e\}$ and $e^*_\phi = \frac{a}{\phi}$. In such a case increasing the marginal return to effort only makes teachers exert lesser effort in equilibrium (for a given a) because the benefit from exerting effort is fixed at a.

In the case with increased complementarity and a small enough σ I have:

1. Teacher effort

    - e^*_ϕ decreases with an increase in complementarity ($\frac{\partial e^*_\phi}{\partial \phi} < 0$). This result can be understood by observing that effort is costly for teachers and they exert it only to get the benefit of increased learning. With a very low elasticity of substitution the learning is capped by student ability, so when the marginal return to effort is increased since they cannot increase learning beyond a limit, they reduce effort to save themselves the cost.
    - The marginal effort of student ability is decreasing with complementarity ($\frac{\partial^2 e^*_\phi}{\partial a \partial \phi} < 0$) i.e. with an increase in the complementarity between a student-teacher pair, the *highest* ability students lose the most, in an absolute sense, in terms of reduced teacher effort and are in this sense "taught below their ability".
    - The marginal effort of teacher ability is decreasing with complementarity ($\frac{\partial^2 e^*_\phi}{\partial A \partial \phi} < 0$) such that with an increase in complementarity all teachers exert lesser effort irrespective of their ability level but the decrease in effort is the maximum for highest ability teachers in absolute terms. As before, this could be just because the lower ability teachers exert such low effort in the benchmark case that they have limited scope to further reduce their effort.

2. Student learning

    - The marginal learning of student ability decreases with an increase in complementarity ($\frac{\partial^2 L^*_\phi}{\partial \phi \partial a} < 0$). An intuitive way to think about this is that when two inputs are almost perfect complements, increasing the marginal return of an input reduces the use of that input up to a level where the two inputs are equal ($\phi e^*_\phi = a$) and since learning is increasing in both inputs, this decrease in optimal effort causes a decline in learning.

In the perfect complements case learning is fixed at a and hence this derivative would be zero.

- The marginal learning of teacher ability decreases with an increase in complementarity ($\frac{\partial^2 L_\phi^*}{\partial\phi\partial A} < 0$). This effect is coming from the fact that the highest ability teachers reduce their effort the most when the marginal return to effort increases.
- The marginal learning for any teacher and student ability combination decreases with an increase in complementarity ($\frac{\partial}{\partial\phi}\left(\frac{\partial^2 L_\phi^*}{\partial A\partial a}\right) < 0$) when σ is small enough. Intuitively, the effect of the interaction between student and teacher ability on learning *decreases* with an increase in complementarity.

Prediction 2: $\frac{\partial}{\partial\phi}\left(\frac{\partial^2 L_\phi^*}{\partial A\partial a}\right) < 0$ ***implies that if the main mechanism behind the positive match effect is an increase in complementarity then I should see a negative interaction effect between student and teacher ability on enrollment, as long as*** $\frac{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - (1-\alpha)}{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}} - \sigma > 0$.

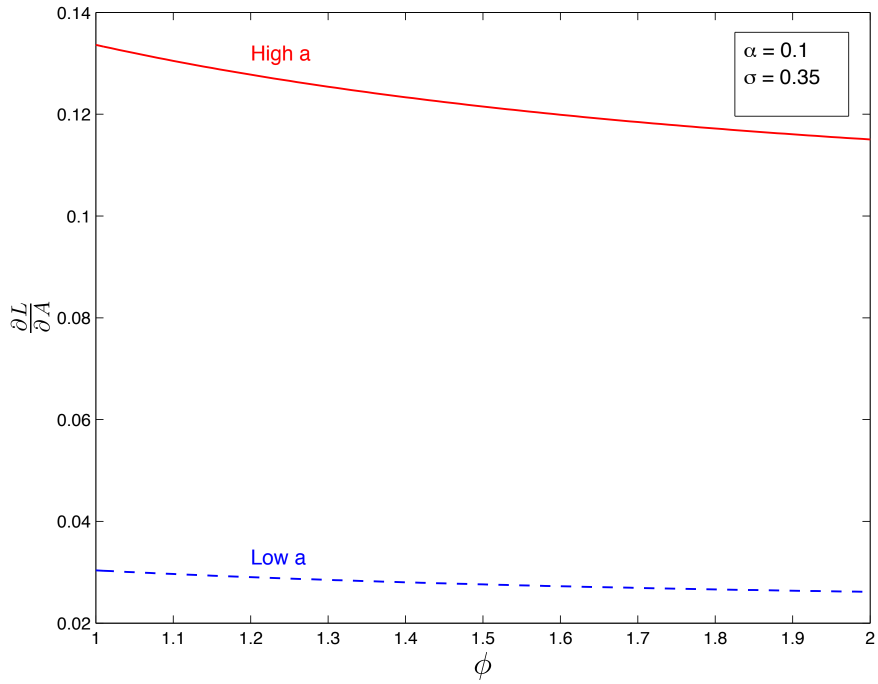
For the values outlined above, I plot the marginal learning of teacher ability as complementarity increases ($\uparrow \phi$) in Figure 3. First, the reader should note that $\frac{\partial L}{\partial A}$ takes on positive values and has a negative slope which shows that it is decreasing with an increase in complementarity ($\frac{\partial^2 L_\phi^*}{\partial\phi\partial A} < 0$). This change in the marginal learning of teacher ability with complementarity is greater for higher ability students compared to lower ability students. Hence, I see that higher ability students lose out more than lower ability students as complementarity increases, as seen by the more negative slope for high ability students compared to the low ability students. This change in slope is $\frac{\partial}{\partial\phi}\left(\frac{\partial^2 L_\phi^*}{\partial A\partial a}\right)$ as highlighted in Prediction 2.[40] I find the same result for a range of values for σ for every α. For example, for a low α such as $\alpha = 0.1$ (as in Panel (a) in Figure 3) I can distinguish between discrimination and complementarity as long as $\sigma \in (0.33, 0.5]$ and for a high α like $\alpha = 0.9$ (as in Panel (b) in Figure 3) the range is $\sigma \in [0.49, 0.82]$.

Overall, the model predicts that we should be able to distinguish between active effects and passive effects based on the sign of the cross-partial between student and teacher ability with and without an ethnic match. If the cross-partial is positive with an ethnic match then the main mechanism is a decrease in discrimination from switching to a co-ethnic teacher. On the other

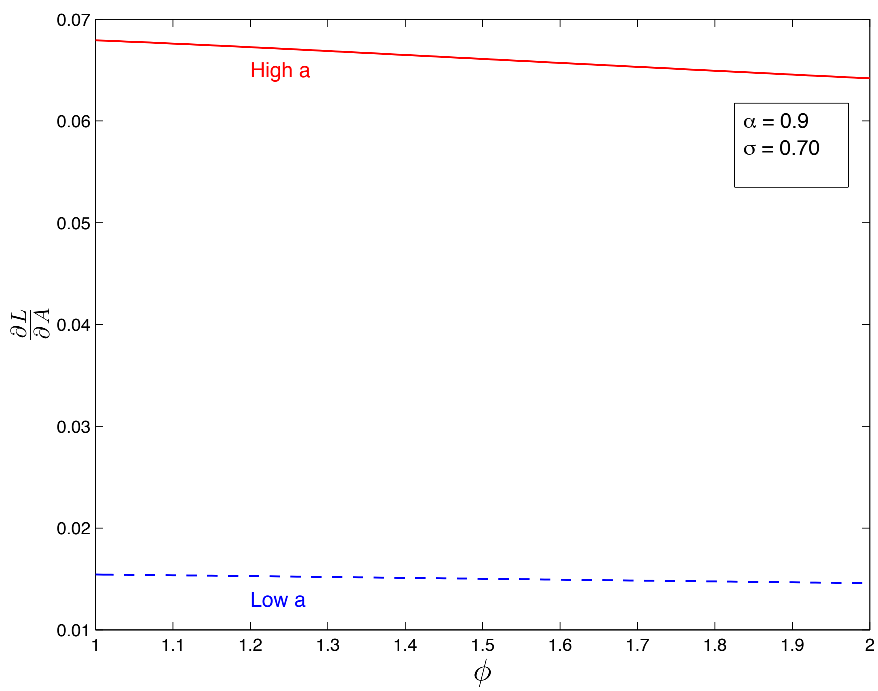
[40]Please see Figure A2 in the Appendix for the graph for $\frac{\partial}{\partial\phi}\left(\frac{\partial^2 L_\phi^*}{\partial A\partial a}\right)$ for the same parameter values.

Figure 3: Change in the marginal learning of teacher ability with an increase in complementarity for different levels of student ability

(a) The effects with a low factor share of student ability (Low α)



(b) The effects with a high factor share of student ability (High α)



hand, if the cross-partial is negative with an ethnic match then the main mechanism is an increase in complementarity from being taught by a co-ethnic teacher. To test the model empirically, I measure student ability using subject-wise student test scores at the school-grade level by ethnicity. The student test scores by ethnicity have been collected exclusively for this project in 12 of the 24 districts in Jharkhand by the Jharkhand Education Project Council. I will measure teacher's ability using teachers' score in the statewide Teacher Eligibility Test, a requirement by the RTE for candidates to be hired as regular teachers in government schools.[41] The first comprehensive test was conducted by the Jharkhand Academic Council in 2012. I received this data from the Jharkhand Academic Council this summer and I am at present working on testing the predictions of the model empirically.

4 Data

For this study, I use three data sets. The primary source of data for this paper is the Unified District Information System for Education (U-DISE) that is collected at the school level for all schools in India. From U-DISE I have data on all schools in Jharkhand for five years from 2009-10 to 2013-14.[42] While the Indian government makes U-DISE data widely available, during my field visits I collected detailed information on teachers that is not publicly available. My final data set includes grade-wise enrollment of students or the number of students by gender and ethnicity, measures of school infrastructure, information on teachers including the category of grades they teach (primary, upper primary, secondary or higher secondary), demographic information,[43] their educational and professional qualifications, experience and location of residence.

While schools move in and out of the sample in some years, I have been able to construct a continuous panel of schools for school years 2009-10 to 2013-14 in Jharkhand.[44] For simplicity,

[41] All teacher aspirants, who fulfill the basic eligibility criteria of age and educational qualifications, appear for Teachers' Eligibility Test (TET) conducted by the Jharkhand Academic Council. After clearing the TET, candidates appear for the teachers' entrance examination conducted by Jharkhand Public Service Commission. And finally, the candidates who clear the JPSC entrance exam are recruited as regular teachers in government schools. The TET has been held annually in Jharkhand since 2011.

[42] All schools that provide formal education come under the purview of U-DISE. This includes government schools (97 per cent of my estimation sample) and recognized unaided private schools, aided schools, unrecognized schools, recognized Madrasas and unrecognized Madrasas.

[43] Such as their gender, ethnicity and marital status.

[44] One of the major data-cleaning challenge was in matching school codes to create a panel dataset. As noted by

from now on I will refer to school years by the year in which the fall term occurs (e.g., school year 2009-10 is 2009).

Though it is logistically challenging to collect data from all schools in the country, the U-DISE data reflect a careful multi-state data-collection process. First, school headmasters answer a nationally-standardized survey-questionnaire in the month of September every year. Second, cluster officials verify responses for completeness and accuracy followed by checking at the block level. Third, district officials aggregate the data and check it for computational and consistency errors. Fourth, state-level officials conduct further consistency checks. In a final step, each state is responsible for hiring external agents to conduct post-enumeration audits and cross-check data with site visits.

In the empirical analysis I focus on primary and upper primary schools for which I have five years of data. U-DISE data contains 47,408 schools in Jharkhand, of these over 38,000 schools (82 per cent) have information for all five years. Within this set of schools with a complete panel I look as co-educational primary and upper primary schools (97 per cent). My final sample consists of 31,621 schools that have information on the ethnicity of the teachers (over two-thirds of all the schools in Jharkhand).

My second source of data is the 2011 Census of India. The Census contains village level information about demographics (such as population by ethnicity and gender, literacy rate by gender), wealth (in the form of variables like average household assets)[45] and labor force participation (for example, rate of employment by gender and sector of employment). These variables provide good controls for the composition of the local population for each school that are likely to affect enrollment.

Finally, I use geospatial data on the location of all the villages and schools in Jharkhand. This data was obtained from the Jharkhand Space Applications Center. This data enables matching school level data from U-DISE to the village level data from the Census. Figure 4 shows the ethnic

previous studies using U-DISE data (Adukia 2014) the given school codes do not necessarily match the same school from one year to the next. Each digit of the school code corresponds to geographic info about the school such as its state, district, block, and village. The school code changes when boundaries between villages, blocks and districts shift. I matched the schools using an algorithm which identifies schools based on their name, location and the year of change in the school code.

[45] I use principal component analysis to create a wealth index for each village based on the description of the housing unit such as the roofing material, wall materials, source of drinking water, access to electricity, type of laterine, fuel used for cooking, assets in the house etc.

distribution of teachers and students in all the schools in the estimation sample. In Panel (a), we can see that, as expected, tribal teachers are found more in the areas which have a high proportion of tribals (as seen in Figure 1) but even in these areas the proportion of tribal teachers in schools varies from around 0.4 to 1. Together with Panel (b) we can see the strong positive correlation between the spatial spread of tribal students and tribal teachers.

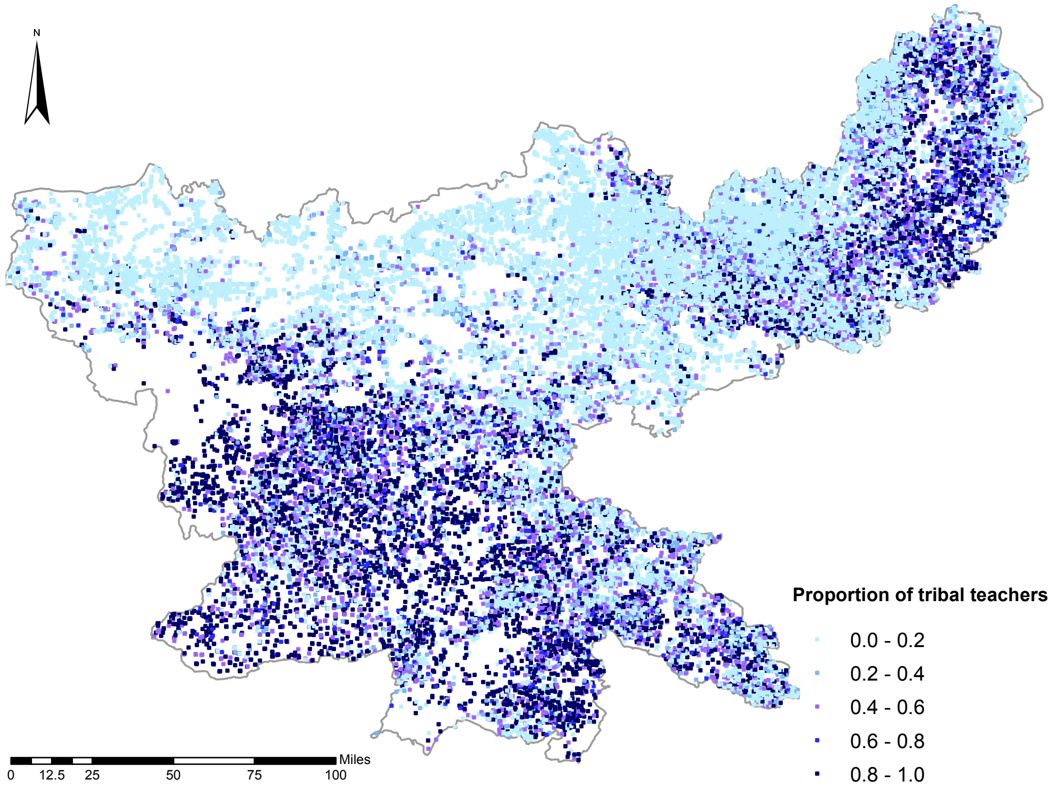
In Table 1 I present the summary statistics from the final estimation sample. Panel A describes the enrollment of students. There are on average 12 students in each school-grade that consist of an almost equal number of boys and girls. In an average school-grade, non-tribal students are almost four times the number of tribal students, which is more than expected based on the ethnic composition of the state. As mentioned in Section 1, tribal students exhibit much higher drop out rates than the rest of the population and this is visible in a crude sense in the changes in the ethnic composition between the primary and the upper primary grades. In the primary grades the number of non-tribals is 3.4 times the number of tribal students while in upper primary grades this goes up to 4.75 times.

Panel B presents descriptive statistics for the 156 thousand teachers in the final estimation sample. Tribal teachers comprise 21 per cent of the teachers in Jharkhand's primary and upper primary schools. They are more likely to be female, older and unmarried compared to their non-tribal counterparts. Tribal teachers are more likely to be hired as short-term contractors as opposed to regular civil-service teachers.[46] Tribal teachers are less qualified than non-tribal teachers on average in that they are less likely to have completed high school, have a bachelors or a masters degree. They are also marginally less likely to hold the head teacher position or have a professional degree (equivalent of a Bachelors in Education or Masters in Education) and have spent fewer years on the job. Since teacher characteristics vary systematically by ethnicity, I will report the main results on the impact of matching teacher and student ethnicity both with and without controls for these additional teacher characteristics. Additionally, in light of these differences I also check the robustness of the key results by allowing these teacher's characteristics to have heterogeneous effects on student enrollment.

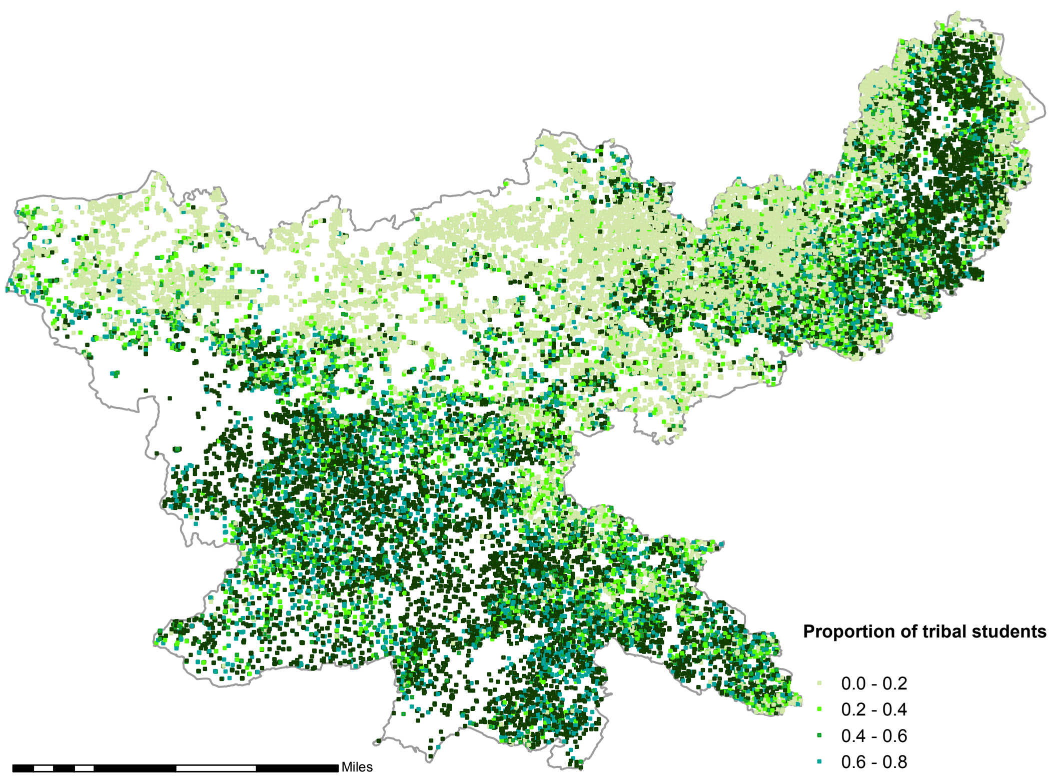
[46]Contract teachers (also known as para-teachers in India) are hired at the school level by school committees and have usually completed either high school or college but typically have no formal teacher training. Their contracts may be renewed annually and they are not protected by any civil-service rules (Muralidharan and Sundararaman 2013). Their typical salary is less than one-fifth of the average salary of regular government teachers when they first join a school in Jharkhand (National Study on Working Conditions of Teachers, Jharkhand, 2015).

Figure 4: Ethnic composition of students and teachers by school

(a) Average proportion of tribal teachers for each school in the estimation sample



(b) Average proportion of tribal students for each school in the estimation sample



Note: Each square represents a school in our sample.

Table 1: Summary statistics

	Observations	Mean	Non-tribal	Tribal	Non-tribal - Tribal
	Panel A: Student enrollment by school-grade				
Total	916,632	11.59	18.19	4.98	13.21***
		(17.665)	(21.395)	(8.898)	
Female	916,632	5.77	9.10	2.44	6.66***
		(9.068)	(11.037)	(4.527)	
Male	916,632	5.81	9.09	2.54	6.55***
		(9.204)	(11.179)	(4.799)	
Primary grades	779,250	10.40	16.07	4.74	11.33***
		(14.966)	(17.768)	(8.254)	
Upper primary grades	137,382	18.30	30.23	6.37	23.86***
		(27.543)	(33.059)	(11.818)	
	Panel B: Teachers				
Tribal	156,093	0.21			
		(0.407)			
Female	156,093	0.22	0.19	0.36	-0.17***
		(0.416)	(0.389)	(0.479)	
Age	156,093	38.70	38.47	39.55	-1.07***
		(10.166)	(10.333)	(9.468)	
Married	152,787	0.92	0.93	0.89	0.04***
		(0.265)	(0.249)	(0.312)	
Regular teacher	156,093	0.45	0.45	0.44	0.01***
		(0.497)	(0.497)	(0.496)	
Contract teacher	156,093	0.55	0.55	0.56	-0.01***
		(0.497)	(0.498)	(0.496)	
Head teacher	156,093	0.11	0.11	0.11	-0.00***
		(0.312)	(0.310)	(0.317)	
Completed high school	156,093	0.90	0.91	0.86	0.05***
		(0.295)	(0.280)	(0.343)	
Completed Bachelors	156,093	0.60	0.64	0.45	0.19***
		(0.490)	(0.480)	(0.497)	
Completed Masters	156,093	0.12	0.13	0.06	0.07***
		(0.324)	(0.341)	(0.238)	
Has a professional degree	156,093	0.09	0.10	0.07	0.03***
		(0.291)	(0.300)	(0.255)	
Experience	156,093	10.52	10.40	10.94	-0.54***
		(11.541)	(12.318)	(7.977)	
Lives in school village	156,093	0.37	0.38	0.34	0.04***
		(0.483)	(0.485)	(0.472)	
Lives in school block	156,093	0.73	0.74	0.70	0.04***
		(0.443)	(0.439)	(0.457)	

Note: (1) Standard deviations are given in parenthesis. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

5 Estimation and identification

5.1 Estimation strategy

To estimate the impact of having a tribal teacher on enrollment of tribals, the following equation is estimated using panel data for the period 2009-2013:

$$Enroll_{gst} = \beta_0 + \beta_1 PTT_{gst} + \beta_2 Tribal_{gst} + \beta_3 PTT_{gst} \times Tribal_{gst} + \varepsilon_{gst} \tag{3}$$

Each observation is a school-grade-year cell with g indexing the grade, s school and t year. $Enroll_{gst}$ is the number of students enrolled in grade g in school s at time t.[47] We pool the enrollment of tribals and non-tribals to enable easy comparison. PTT_{gst} is the current probability of a tribal teacher teaching the grade, as mentioned in the data section I do not observe a one-to-one match of teachers to grades in the data but the pool of teachers assigned to a category of grades for example primary or upper-primary. Therefore, PTT_{gst} is calculated as the proportion of tribal teachers in the pool of teachers who can be teaching grade g at time t. $Tribal_{gst}$ is an indicator equal to 1 if I am looking at tribal enrollment and zero for non-tribal enrollment. ε_{gst} is the unobserved error term.

The estimating equation allows me to calculate the marginal impact of changing each component of the student-teacher ethnicity combination relative to non-tribal students taught by non-tribal teachers ($PTT_{gst} = 0, Tribal_{gst} = 0$).

β_3 in the above equation captures the *relative* enrollment effect of having a pool of only tribal teachers in the grade on tribals vs. non-tribals. β_3 can be thought of as a difference-in-difference estimate. The coefficient on the interaction term reflects the sum of the relative advantage of matching tribal teachers perfectly with tribal students (vs. non-tribal students) and the relative disadvantage of matching non-tribal teachers perfectly with tribal students (vs. non-tribal students).[48] It is important to note that β_3 is symmetric by construction in that it also measures the relative

[47] I also estimated proportional effects using $log(Enroll_{gst})$ as the dependent variable and the results are robust. I prefer the estimation in levels because it identifies the effect in terms of the actual number of kids and I report practical size effects in SD terms in the discussion to compare results across specifications.

[48] In the classic difference-in-difference framework β_3 = [(tribal teachers, tribal students $(\beta_0 + \beta_1 + \beta_2 + \beta_3)$ - tribal teachers, non-tribal students $(\beta_0 + \beta_1)$) $-$ (non-tribal teachers, tribal students $(\beta_0 + \beta_2)$ $-$ non-tribal teachers, non-tribal students(β_0))].

effectiveness of non-tribals teaching non-tribal students (rather than tribal students) compared to tribals teaching non-tribal students (rather than tribal students). Additionally, a positive β_3 does not necessarily imply that both tribals and non-tribals have better outcomes when they have a co-ethnic teacher.[49] I refer to this coefficient on the interaction term as the match effect.

β_1 in equation 3 is the difference in enrollment for non-tribals taught by all tribal vs. all non-tribal teachers (i.e. all tribal teachers teaching non-tribal students - all non-tribal teachers teaching non-tribal students). Similarly, β_2 is the extent to which I see a difference in enrollment of tribal vs. non-tribal students when taught by all non-tribal teachers (or tribal students with all non-tribal teachers - non-tribal students with all non-tribal teachers).

The marginal effects of interest can be computed based on a combination of the parameters in equation 3:

1. β_1 measures the gain or loss to non-tribals from being paired with a uniform pool of tribal teachers ($\beta_0 + \beta_1$) relative to when paired with a pool of only non-tribal teachers (β_0)

2. $\beta_1 + \beta_3$ measures the effect on tribal students when matched with a group of tribal teachers ($\beta_0 + \beta_1 + \beta_2 + \beta_3$) relative to when paired with non-tribal teachers ($\beta_0 + \beta_2$). As mentioned previously, it is possible that tribal teachers are relatively more competent at teaching tribals than non-tribal teachers (a positive interaction term or $\beta_3 > 0$), but maybe overall less effective ($\beta_1 < 0$) resulting in tribals being better off with a pool of non-tribal teachers despite the absence of gains based on co-ethnicity ($\beta_1 + \beta_3 < 0$). I refer to this sum as the total effect.

5.2 Threats to identification

There are two main identification challenges in interpreting the coefficients in 3 causally. The first is non-random assignment of teachers based on ethnicity. There might be non-random assignment of tribal teachers across districts, blocks or villages with more tribal teachers found in areas where tribals have a lot of political clout. In such areas, parents would find it more profitable to keep their kids in school because of, say, assured employment opportunities in the future. I address this concern by augmenting equation 3 with district fixed effects and village level controls such as

[49]For example, a positive β_3 could co-exist with a situation where all students are better off with a tribal (or non-tribal) teacher because of the generally higher effectiveness of the tribal (or non-tribal) teacher for all students. This is similar to the reasoning provided by Muralidharan and Sheth (2015).

the proportion of tribal population in a village, a relative rank of the village in the state's wealth distribution, the male and female literacy rate, the employment rate in different sectors. In this regression β_3 measures the impact of having a co-ethnic teacher on enrollment relative to the average enrollment in a district conditional on certain village characteristics. Tribal teachers also may not be randomly assigned to schools, with more tribal teachers found in schools in areas with greater tribal education levels. Thus in such schools tribals would not drop out irrespective of the teacher's ethnicity. Such an omitted variable would bias β_3 upwards. I address this concern by adding school fixed effects and estimating the effect of a co-ethnic teacher on enrollment relative to the school's overall average enrollment. Finally, it may be the case that teachers are not randomly assigned to grades within schools and I would have a similar omitted variable bias if tribal teachers are assigned to grades with more tribal students for example, in lower grades, since I know that tribal students drop out earlier than non-tribal students. The latter is true nationally[50] and in my data. Table A1 shows that tribal enrollment decreases as the grade level increases. Controlling for cohort fixed effects and school fixed effects I find that the ratio of tribal to non-tribals in a grade decreases from 0.42 in Grade 1 to 0.29 in Grade 5 and by Grade 8 most of the tribal students have dropped out of school. To circumvent this potential non-random assignment I include school-grade fixed effects. This controls for average enrollment in a grade in a school, hence taking into account the differential enrollment by ethnicity across grades. Given the panel structure of my data, after the inclusion of school $\times$ grade fixed effects the identifying variation is coming from the differential effect of teachers (by ethnicity) on enrollment of tribal students vs. non-tribal students relative to the mean enrollment of students in that school-grade cell over the five years of the data.

A second threat to identification is reverse causality, in that it is not the students who are responding to changes in their teacher's ethnicity but teachers are in fact being moved to schools or grades within a school on the basis of the ethnic composition of the cohort. This does not seem to be the case. 97 per cent of the sample is government schools and here teachers are allocated to schools on the basis of existing need of teachers required to maintain the Pupil Teacher Ratio (30:1) mandated by the RTE 2009.[51] These vacancies are identified by the Block Extension Education Officer and consolidated to the block level. The final decision regarding the number of teachers

[50] Over 33 per cent of tribal children drop out of school before they complete Grade 5, compared to a country-wide drop out rate of 20 per cent (Statistics of School Education 2010-11).

[51] National Study on Working Conditions of Teachers: State Report for Jharkhand, February 2015.

to be recruited is taken by the Human Resource Development officials in consultation with the Finance Department, Ministry of Primary and Secondary Education and the Chief Minister. There is, however, a norm that teachers are placed in their home block[52] for example in the data over 72 per cent of the teachers are from the same block as where the school they work in is located. Hence, teachers do not seem to be assigned to schools on the basis of the ethnicity of the students.[53] Additionally, hardly any teachers switch schools. In Jharkhand, less than 1 per cent of 156 thousand teachers ever switch schools. This is because teachers are required to spend a minimum of 3 years at a given post before transfer can be obtained.[54] Transfers are usually done in the teacher's home or neighboring blocks.[55] Finally, with the widely followed practice of multi-grade teaching (in the data a teacher teacher an average of 2.5 grades per year) it is unlikely that teachers are being assigned to grades with more co-ethnic children.

6 Results

6.1 Main specification

The main results of the paper, based on estimation equation 3, are presented in Table 2. The columns show increasingly restrictive identification assumptions, to address the non-random assignment of teachers, as explained above. Starting with no controls (column 1) I include district fixed effects and village level controls in column 2, in column 3 I add school fixed effects and finally in column 4 I have school-grade fixed effects and teacher characteristics. The teacher covariates are added to account for the differences in demographic, educational and locational characteristics between tribal and non-tribal teachers; this helps me parse out a pure ethnicity effect from effects that are caused by teacher characteristics associated with teacher ethnicity. By controlling

[52]Interview with the District Superintendent of Education, District of Hazaribagh, Jharkhand, December 2014.

[53]Though when I tested the difference in observable school characteristics between tribal and non-tribal teachers (conditional on district fixed effects and with village level controls) I found that tribal teachers relative to non-tribal teachers work in schools with lesser facilities. Tribal teachers relative to non-tribal teachers are more likely to be in schools with no boundary wall, with no access to drinking water, with fewer *pucca* classrooms, with fewer functioning toilets, with a lower probability of having a separate girls toilet, with no computers, with lesser likelihood of having desk and chairs for all teachers and students. I found no difference between tribal and non-tribal teachers in the probability of working in a government school and in schools that have a playground.

[54]In some special circumstances this can be reduced to 2 years.

[55]National Study on Working Conditions of Teachers: State Report for Jharkhand, February 2015.

for teacher characteristics I am able to estimate the impact of switching a teacher's ethnicity from non-tribal to tribal holding other observable characteristics constant. This contrasts with the effect of simply replacing a non-tribal teacher with a tribal teacher where there is confounding of the teacher's ethnicity effect and the average characteristics correlated with that ethnicity. As can be seen from Table 2, the results are remarkably stable and robust under various specifications. Below I focus on discussing the results in column 4, unless specified otherwise.

The positive and highly significant β_3 shows us that tribal teachers are more effective in teaching tribal students relative to non-tribal teachers. The *relative* enrollment effect of having a pool of only tribal teachers in the grade on tribals vs. non-tribals is 28, which is a positive match effect of 0.44 SD on enrollment. Since β_3 is symmetric, this captures the relative advantage of a teacher at teaching co-ethnic students. Looking specifically at tribals, the enrollment of tribal students increases when they are more likely to be taught by tribal teachers relative to non-tribal teachers. 14 more tribals enroll in a grade on average ($\beta_1 + \beta_3$) if the probability of being taught by a tribal teacher increases from 0 to 1. In terms of practical effect size,[56] a 1 SD increase in the probability of having a tribal teacher leads to an increase of 0.57 SD in tribal enrollment per year. However, there is a strong negative effect on non-tribals from increased likelihood of being taught by a tribal teacher relative to non-tribal teachers, such that the enrollment of non-tribal students decreases by 14 when the probability of having a tribal teacher increase by 100 per cent. While the absolute effect seems equal and opposite to the effect of having a tribal teacher on tribal students, the practical effect is -0.24 SD, less than a half of the latter. This result highlights the flaws in the typical policy response of recruiting more tribal teachers if you want more tribal students to enroll in school, because it ignores the unintended and undesirable consequences of harming students who do not share the teacher's demographic traits. In a following section we examine whether this decline in enrollment is coming from non-tribal students dropping out or re-locating to another school in the neighborhood.

Next, I examine whether this effect is short-term or if it persists over time in Table 3. For this I include the lagged values of the probability of having a tribal teacher for each grade. Learning accumulates over time, by adding lagged values of the ethnicity of the teacher I attempt to capture the stock effect of having a tribal teacher on enrollment. As expected the contemporaneous effect

56 $\frac{(\beta_1+\beta_3)\times SD\ of\ Probability\ of\ having\ a\ tribal\ teacher}{SD\ of\ tribal\ enrollment}$

Table 2: Effect of having a tribal teacher on enrollment: pooled across tribal and non-tribal enrollment

	Enrollment			
	(1)	(2)	(3)	(4)
Prop. of tribal teachers (β_1)	-18.99***	-17.60***	-14.29***	-14.26***
	(0.218)	(0.221)	(0.234)	(0.194)
Tribal (β_2)	-19.28***	-19.27***	-19.28***	-19.28***
	(0.172)	(0.172)	(0.172)	(0.172)
Prop. of tribal teachers x Tribal (β_3)	**28.43*** **	**28.42*** **	**28.43*** **	**28.43*** **
	(0.251)	(0.251)	(0.251)	(0.251)
Constant	22.25***	20.84***	21.24***	24.38***
	(0.161)	(0.500)	(0.0976)	(0.236)
$\beta_1 + \beta_3$	**9.44**	**10.82**	**14.14**	**14.17**
F-test p-value ($H_0 : \beta_1 + \beta_3 = 0$)	**0.00**	**0.00**	**0.00**	**0.00**
Observations	916,632	916,504	916,632	916,632
R-squared	0.232	0.250	0.296	0.332
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

of having a co-ethnic teacher remains positive but decreases once I condition on previous grade's teacher composition. Having a uniform pool of tribal teachers teaching tribals vs. non-tribals increases enrollment by 16 relative to uniform pool of non-tribal teachers teaching tribals vs. non-tribals. This is equivalent to an increase of 0.19 SD. Similarly, there is a smaller increase in the number of tribal students when the likelihood of having a tribal teacher increases by 100 per cent, once I control for the lagged probability. It can be seen from Table 3, conditional on the ethnic composition of teachers in the previous grade, 8 more tribals enroll this period if the probability of being taught by a tribal teacher increases by 1. Or in SD terms, a 1 SD increase in current probability of having a tribal teacher increases tribal enrollment by 0.34 SD.

For the lagged effects, conditional on the current ethnic composition of the teachers, the *relative* enrollment effect of having a pool of only tribal teachers in the grade on tribals vs. non-tribals is 12, which is a 0.15 SD increase. Specifically for tribals, conditional on the ethnic composition of teachers in the previous grade, 6 more tribals enroll this period if the probability of being taught by a tribal teacher increases by 1. Or in SD terms, a 1 SD increase in current probability of having a tribal teacher increases tribal enrollment by 0.25 SD. I get similar results when I control for more lagged values of the ethnic composition of teachers (please refer to Table A2 in the Appendix). The fact that the effects are smaller when I condition on ethnic composition of teachers in previous grades indicates that the positive effect of a demographic match is indeed cumulative in nature. Intuitively, if the main mechanism for the demographic match are passive effects like role model effects then one can imagine that the first interaction with a co-ethnic teacher would have the maximum impact on enrollment and conditional on that past match any future matches would have a much smaller positive effect.

6.2 Heterogenous effects

In this section I explore the heterogeneity of effect by teacher characteristics, student gender and grade type.

Table 3: Persistence of the effect of having a tribal teacher on enrollment (over one year): pooled across tribal and non-tribal enrollment

	Enrollment at time t			
	(1)	(2)	(3)	(4)
Prop. of tribal teachers at time t	-11.35***	-11.15***	-8.691***	-8.048***
	(0.469)	(0.475)	(0.443)	(0.328)
Prop. of tribal teachers at time t-1	-7.346***	-7.398***	-6.115***	-6.354***
	(0.500)	(0.490)	(0.336)	(0.328)
Tribal	-21.02***	-21.01***	-21.02***	-21.02***
	(0.216)	(0.216)	(0.216)	(0.216)
Prop. of tribal teachers at time t x Tribal	**16.25*** **	**16.25*** **	**16.25*** **	**16.25*** **
	(0.550)	(0.550)	(0.550)	(0.550)
Prop. of tribal teachers at time t-1 x Tribal	**12.30*** **	**12.29*** **	**12.30*** **	**12.30*** **
	(0.586)	(0.586)	(0.586)	(0.586)
Constant	23.08***	27.12***	22.56***	24.51***
	(0.208)	(0.772)	(0.123)	(0.292)
Observations	409,320	409,236	409,320	409,320
R-squared	0.274	0.292	0.364	0.397
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Teacher characteristics

As seen in Table 1, tribal teachers are systematically different from their non-tribal counterparts. Hence it is possible that the β_3 estimated in equation 3 reflects not just the effect of an ethnicity match but also the effect of tribal students matching with teacher attributes that are consistently found more commonly in tribal teachers. In Table 4, I show a series of regressions where the main specification with school-grade fixed effects is augmented to include characteristics of the pool of teachers and the interaction of the characteristics with ethnicity of the student enrollment. The teacher characteristics include demographic characteristics such as gender, age, marital status, education, training, experience on the job and place of residence.

I present the results both without and with controls for teacher attributes in Panel A and Panel B of Table 4 respectively.[57] Based on the results, β_3 is extremely robust to allowing the slope to vary with characteristics other than the teacher's ethnicity and are similar to which is similar to the baseline results in column 1. These results corroborate the fact that the positive β_3 in Table 2 reflects an ethnicity match and is not accounted for by other characteristics that are systematically different between tribal and non-tribal teachers. In all cases (and across panels A and B), the estimated relative gain in enrollment from having a uniform pool of co-ethnic teachers (β_3) is around 28 or 0.43 SD and for tribal students the gain in switching from a non-tribal to a tribal teacher ($\beta_1 + \beta_3$) ranges from 0.55 SD to 0.59 SD per annum.

Student gender

Qualitative studies in sociology and anthropology have found that the tribals in India exhibit a lower degree of gender discrimination (Atal 2009, Mitra 2007) and demographic studies have provided evidence of great female agency among tribals (Maharatna 2005) compared to the rest of the population. In terms of sex ratios, a crude measure of gender discrimination, tribals in India have a more favorable overall sex ratio (101 males per 100 females amongst tribals versus 106 males per 100 females) as well as a more balanced child sex ratio[58] (104 males per 100 females

[57] This is similar to the analysis carried out by Muralidharan and Sheth (2015).

[58] The child sex ratio is the number of males per 100 females for children aged six years or younger. The child sex ratio is supposed to capture the sex selection activities carried out by parents.

Table 4: Heterogeneity of effects by teacher characteristics

Teacher characteristic	(1)	(2) Female	(3) Age	(4) Married	(5) Same village	(6) Head teacher	(7) Experience	(8) Contract teacher	(9) Bachelors	(10) B.Ed or M.Ed
	Panel A: No controls									
Prob. of grade having a tribal teacher (β_1)	-14.01***	-14.20***	-14.14***	-13.89***	-13.89***	-14.01***	-14.06***	-14.13***	-13.75***	-13.94***
	(0.194)	(0.194)	(0.194)	(0.196)	(0.194)	(0.194)	(0.194)	(0.193)	(0.192)	(0.193)
Tribal (β_2)	-19.28***	-18.80***	-12.15***	-15.17***	-17.72***	-19.29***	-18.19***	-22.25***	-16.74***	-18.93***
	(0.172)	(0.180)	(0.495)	(0.453)	(0.198)	(0.177)	(0.498)	(0.252)	(0.182)	(0.168)
Prob. of class having a tribal teacher x Tribal (β_3)	**28.43*** **	**28.79*** **	**28.68*** **	**28.14*** **	**28.18*** **	**28.43*** **	**28.52*** **	**28.45*** **	**27.52*** **	**28.33*** **
	(0.251)	(0.253)	(0.253)	(0.250)	(0.250)	(0.251)	(0.253)	(0.249)	(0.243)	(0.250)
Teacher characteristic		1.455***	0.0403***	1.225***	2.008***	0.144	0.0480***	-3.453***	1.149***	3.212***
		(0.235)	(0.00842)	(0.277)	(0.139)	(0.220)	(0.0172)	(0.141)	(0.141)	(0.328)
Teacher characteristic x Tribal		-2.655***	-0.188***	-4.251***	-3.855***	0.0984	-0.109**	4.867***	-4.203***	-4.640***
		(0.324)	(0.0136)	(0.450)	(0.236)	(0.358)	(0.0470)	(0.233)	(0.241)	(0.544)
Constant	21.18***	20.92***	19.67***	19.96***	20.37***	21.17***	20.70***	23.31***	20.49***	20.94***
	(0.0924)	(0.102)	(0.316)	(0.276)	(0.108)	(0.0958)	(0.192)	(0.141)	(0.109)	(0.0910)
$\beta_1 + \beta_3$	**14.42**	**14.59**	**14.54**	**14.25**	**14.29**	**14.42**	**14.46**	**14.32**	**13.77**	**14.39**
F-test p-value ($H_0 : \beta_1 + \beta_3 = 0$)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Observations	916,632	916,632	916,632	907,650	916,632	916,632	916,632	916,632	916,632	916,632
R-squared	0.331	0.332	0.334	0.331	0.335	0.331	0.333	0.337	0.335	0.332
	Panel B: With controls									
Prob. of grade having a tribal teacher (β_1)	-14.23***	-14.41***	-14.36***	-14.12***	-14.10***	-14.23***	-14.28***	-14.25***	-13.77***	-14.18***
	(0.196)	(0.197)	(0.197)	(0.196)	(0.196)	(0.196)	(0.197)	(0.195)	(0.193)	(0.196)
Tribal (β_2)	-19.21***	-18.75***	-12.11***	-15.17***	-17.60***	-19.21***	-18.12***	-22.18***	-16.65***	-18.86***
	(0.172)	(0.180)	(0.497)	(0.453)	(0.197)	(0.177)	(0.496)	(0.253)	(0.182)	(0.168)
Prob. of class having a tribal teacher x Tribal (β_3)	**28.36*** **	**28.71*** **	**28.61*** **	**28.14*** **	**28.09*** **	**28.36*** **	**28.45*** **	**28.39*** **	**27.44*** **	**28.26*** **
	(0.251)	(0.253)	(0.254)	(0.250)	(0.250)	(0.251)	(0.253)	(0.249)	(0.244)	(0.250)
Teacher characteristic		1.250***	0.0387***	1.609***	2.112***	0.0596	0.0523***	-3.350***	1.369***	3.170***
		(0.237)	(0.00877)	(0.278)	(0.137)	(0.223)	(0.0198)	(0.142)	(0.142)	(0.330)
Teacher characteristic x Tribal		-2.589***	-0.187***	-4.251***	-3.961***	0.0531	-0.108**	4.846***	-4.221***	-4.623***
		(0.323)	(0.0136)	(0.450)	(0.236)	(0.359)	(0.0469)	(0.235)	(0.241)	(0.548)
Constant	24.66***	24.43***	21.11***	22.64***	23.86***	24.66***	24.12***	26.14***	23.38***	24.49***
	(0.238)	(0.240)	(0.340)	(0.327)	(0.242)	(0.240)	(0.363)	(0.257)	(0.239)	(0.238)
$\beta_1 + \beta_3$	**14.13**	**14.30**	**14.25**	**14.02**	**13.99**	**14.13**	**14.17**	**14.14**	**13.67**	**14.08**
F-test p-value ($H_0 : \beta_1 + \beta_3 = 0$)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Observations	907,650	907,650	907,650	907,650	907,650	907,650	907,650	907,650	907,650	907,650
R-squared	0.331	0.332	0.334	0.332	0.335	0.331	0.333	0.337	0.335	0.332
School x Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: All standard errors are robust and clustered at the school level. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

for tribals versus 109 males per 100 females) compared to the rest of the population (2011).[59] In terms of labor force participation rates, a proxy for the bargaining power of women in a household, tribal women have higher labor force participation rates, the female participation rate is 46 per cent among tribal women compared to 30 per cent all over India (2010).[60]

Based on this evidence I test the hypothesis that tribal teachers (irrespective of their gender) exhibit lesser bias against female students (irrespective of their ethnicity) relative to non-tribal teachers who have been exposed a more patriarchal culture. I do not find this in my data - neither for a match effect nor for the total effect (Table A3).

Next, I looked at the differential effect of the ethnicity of the teacher on enrollment of students by ethnicity and gender. Results from Table 5 show that enrollment of students continues to increase with an increase in likelihood of having a co-ethnic teacher and does not change when I allow for heterogeneity by student gender. The match effect (β_3) of 28 additional students that is shown in column 4 of Table 2 appears to be equally shared by boys and girls (a SD effect each of around 0.22 for both). The total increase of 14 students that we saw in Table 2 is also equally shared by the two genders with the enrollment of each increasing by 7 students in case of tribals ($\beta_1 + \beta_3$) and non-tribals ($-\beta_1$). This is equivalent to a practical effect size of 0.52 SD (0.23 SD) for boys and 0.56 SD (0.23 SD) for girls in case of tribals (non-tribals).

Grade type

In Table 6 I examine whether the effect of having a tribal teacher varies across primary and upper primary grades. It is argued that when students are at a younger and more formative age then the role of a demographic match may be especially important.[61] During my visit to schools in Jharkhand I was also told by teachers that tribal parents seem to have a stronger preference for tribal teachers teaching their kids in lower grades. I find supportive evidence in my results both in terms of the match effect and the total effect. Having a uniform pool of tribal teachers teaching

[59]These ratios have, however, been worsening over time indicative of the "process of erosion of tribal tradition of gender equity" (Maharatna 2005)

[60]Nayar et al., 2012.

[61]Another channel that could go in the opposite direction is that when students are in upper primary grades they are around 10-14 years of age and that is the age range when child labor opportunities become widely available to them. The increase in opportunity cost makes it easier for the children to drop out and hence having a co-ethnic teacher at this stage could have a greater impact in this case than when the students are in primary grades.

Table 5: Effect of having a tribal teacher on enrollment by gender of students: pooled across tribal and non-tribal enrollment

	(1)	(2)	(3)	(4)
	Panel A: Enrollment of girls			
Prob. of grade having a tribal teacher (β_{1g})	-9.556***	-8.842***	-7.123***	-7.113***
	(0.108)	(0.109)	(0.117)	(0.0998)
Tribal (β_{2g})	-9.701***	-9.698***	-9.701***	-9.701***
	(0.0871)	(0.0870)	(0.0871)	(0.0871)
Prob. of class having a tribal teacher x Tribal (β_{3g})	**14.25*** **	**14.24*** **	**14.25*** **	**14.25*** **
	(0.126)	(0.126)	(0.126)	(0.126)
Constant	11.14***	10.54***	10.62***	11.87***
	(0.0814)	(0.251)	(0.0491)	(0.120)
$\beta_{1g}+\beta_{3g}$	**4.69**	**5.40**	**7.12**	**7.14**
F-test p-value ($H_0 : \beta_{1g}+\beta_{3g}=0$)	**0.00**	**0.00**	**0.00**	**0.00**
Observations	916,632	916,504	916,632	916,632
R-squared	0.223	0.241	0.279	0.314
	Panel B: Enrollment of boys			
Prob. of grade having a tribal teacher (β_{1b})	-9.435***	-8.762***	-7.165***	-7.148***
	(0.114)	(0.116)	(0.125)	(0.101)
Tribal (β_{2b})	-9.576***	-9.573***	-9.576***	-9.576***
	(0.0882)	(0.0881)	(0.0882)	(0.0882)
Prob. of class having a tribal teacher x Tribal (β_{3b})	**14.18*** **	**14.18*** **	**14.18*** **	**14.18*** **
	(0.129)	(0.129)	(0.129)	(0.129)
Constant	11.10***	10.29***	10.62***	12.50***
	(0.0830)	(0.259)	(0.0505)	(0.127)
$\beta_{1b}+\beta_{3b}$	**4.74**	**5.42**	**7.01**	**7.03**
F-test p-value ($H_0 : \beta_{1b}+\beta_{3b}=0$)	**0.00**	**0.00**	**0.00**	**0.00**
Observations	916,632	916,504	916,632	916,632
R-squared	0.211	0.228	0.269	0.305
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Table 6: Effect of having a tribal teacher on enrollment by grade type: pooled across tribal and non-tribal enrollment

	(1)	(2)	(3)	(4)
	Panel A: Enrollment in primary grades			
Prob. of grade having a tribal teacher (β_{1p})	-16.71***	-15.44***	-12.76***	-13.08***
	(0.178)	(0.187)	(0.188)	(0.186)
Tribal (β_{2p})	-17.25***	-17.24***	-17.25***	-17.25***
	(0.138)	(0.138)	(0.138)	(0.138)
Prob. of class having a tribal teacher x Tribal (β_{3p})	**25.93*** **	**25.92*** **	**25.93*** **	**25.93*** **
	(0.217)	(0.217)	(0.217)	(0.217)
Constant	19.88***	18.09***	18.98***	23.33***
	(0.124)	(0.376)	(0.0779)	(0.249)
$\boldsymbol{\beta_{1p} + \beta_{3p}}$	**9.22**	**10.48**	**13.17**	**12.85**
F-test p-value ($H_0 : \beta_{1p} + \beta_{3p} = 0$)	**0.00**	**0.00**	**0.00**	**0.00**
Observations	779,250	779,170	779,250	779,250
R-squared	0.254	0.273	0.323	0.346
	Panel B: Enrollment in upper-primary grades			
Prob. of grade having a tribal teacher (β_{1u})	-29.01***	-25.82***	-21.64***	-21.51***
	(0.780)	(0.797)	(0.638)	(0.641)
Tribal (β_{2u})	-29.53***	-29.53***	-29.53***	-29.53***
	(0.456)	(0.456)	(0.456)	(0.456)
Prob. of class having a tribal teacher x Tribal (β_{3u})	**43.71*** **	**43.69*** **	**43.71*** **	**43.71*** **
	(0.881)	(0.881)	(0.881)	(0.881)
Constant	34.00***	34.80***	33.04***	29.10***
	(0.447)	(1.467)	(0.236)	(0.745)
$\boldsymbol{\beta_{1u} + \beta_{3u}}$	**14.70**	**17.87**	**22.07**	**22.20**
F-test p-value ($H_0 : \beta_{1u} + \beta_{3u} = 0$)	**0.00**	**0.00**	**0.00**	**0.00**
Observations	137,382	137,334	137,382	137,382
R-squared	0.239	0.274	0.351	0.360
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

tribals (vs. non-tribals) leads to an increase of 26 (0.5 SD) in enrollment relative to a uniform pool of non-tribal teachers teaching tribals (vs. non-tribals) for primary grades and an increase of almost 44 (0.32 SD) in upper primary grades. Increasing the probability of having a tribal teacher by 100 per cent increases enrollment of tribal students by 13 in primary grades but and 22 in upper primary grades ($\beta_1 + \beta_3$). Similarly, increasing the probability of having a non-tribal teacher by 100 per cent increases enrollment of non-tribal students by 13 children in primary grades but and 21 kids in upper primary grades ($-\beta_1$). This means that an increase in the probability of having a co-ethnic teacher by 1 SD increase enrollment of tribal (non-tribal) students by 0.57 SD (0.27 SD) in primary grades and by 0.51 SD (0.18 SD) in upper primary grades.

6.3 Attrition vs. reallocation

In the above discussion I assume that a decline in enrollment is equivalent to students dropping out of the school. However, it could be that they are just relocating to another school. The two results have very different policy implications. To test how much of the change in enrollment is due to attrition of students and how much of it is due to reallocation I look at not only the regression at the school where the probability of having a co-ethnic teacher is changing but also the enrollment in all the schools in the catchment of that school.[62] I use different radii to define this catchment. The intuition behind this exercise is that if the changes in enrollment are only caused by attrition[63] then changing the catchment of the school should have the same effect that we found in Table 2. On the other hand, if all the students are actually just switching to another school then I should observe a zero effect from an increase in the probability of having an ethnically different teacher in a school's catchment. I assume that if a student drops out of one school grade then he or she enrolls in the same grade at another school in the catchment area.

Using the geospatial data on schools I calculate the total enrollment in a school-grade for catchments of 2 kilometers, 3 kilometers and 5 kilometers around each school in my sample. These catchment sizes are in line with the government norms; the RTE mandates that the limit of neighborhood within which a school has to be established has to be within a walking distance of

[62] Referred to as the "feeder area" by the Government of India.

[63] This means that students drop out of the school with an increase in the probability of having a teacher from a different ethnicity and do not enroll in any other school.

1 kilometer for children primary grades and within a walking distance of 2 kilometers for children in upper primary grades. I then estimate equation 3, with school-grade fixed effects and controls for average teacher characteristics, using this catchment enrollment. The results are presented in Table 7. Since the number of schools in a catchment increase with the size of the catchment, the absolute enrollment effects also increase. The results are easier to compare to the main results in SD terms. Looking at a 2km radius catchment a 1 SD increase in the probability of having a tribal teacher decreases the enrollment of non-tribals by 0.07 SD which is almost a third of the 0.24 SD decline in the main specification ($\beta_{1)}$. This means that a third of the non-tribals drop out while two-thirds actually move to other schools. The results are statistically no different when the radius is increased to 3 kilometers. The negative effect is a lot closer to those in the main specification when the radius is increased to 5 kilometers. But in this case we are picking up schools which are not really in the same consideration set and hence adding noise to the measure of catchment enrollment.

The match effect (β_3) is also smaller in case of the catchment enrollment because of the reallocation of students across schools. There is a relative enrollment effect of 0.15 SD (0.18 SD) from having a co-ethnic teacher in case of the 2 kilometer (3 kilometer) catchment, again one-third of the 0.44 SD effect in the main specification. The total effect on tribal enrollment ($\beta_1 + \beta_3$) remains almost the same, with a 1 SD increase in the probability of having a tribal teacher increasing tribal enrollment by 0.41 SD (0.53 SD) when the catchment has a 2 kilometer (3 kilometer) radius.

Overall, it seems that having a ethnically different teacher makes *some* students switch school and *some* students drop out.

6.4 Other types of demographic match

As detailed previously, the literature on the effects of having "a teacher like me" looks at the effect of students and teachers sharing gender, caste and ethnicity. In this section I estimate the effect of having female teacher on the gender gap in enrollment and the effect of having a teacher from the same caste on student enrollment. My results indicate a positive and significant effect in case of the gender match, but it is much smaller than the ethnicity match effect, which is expected given the salience of ethnicity in the sample population. I also find a small positive effect of a caste match on student enrollment.

Table 7: Attrition vs. Reallocation: effect of having a co-ethnic teacher on enrollment in the school's catchment area

	Enrollment		
Radius	2kms	3kms	5kms
	(1)	(2)	(3)
Prob. of grade having a tribal teacher (β_1)	-128.4***	-190.2***	-352.2***
	(4.597)	(5.139)	(6.961)
Tribal (β_2)	-259.6***	-390.6***	-745.4***
	(5.659)	(6.110)	(7.133)
Prob. of class having a tribal teacher x Tribal (β_3)	259.1***	383.4***	705.8***
	(8.252)	(9.109)	(11.87)
Constant	321.8***	483.3***	926.9***
	(5.106)	(5.578)	(7.001)
$\beta_1 + \beta_3$	130.63	193.24	353.68
F-test p-value ($H_0 : \beta_1 + \beta_3 = 0$)	0.00	0.00	0.00
Observations	906,534	906,534	906,534
R-squared	0.122	0.205	0.375
School x Grade FE	Yes	Yes	Yes
Teacher characteristics	Yes	Yes	Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Teacher's characteristics at the shcool-grade include the proportion of tribals, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Gender

In Table 8 I present the effect of having a teacher of the same gender on enrollment. In my preferred specification (column 4) I find a small but significant positive match effect of having a teacher of the same gender and a positive total effect. I observe a match effect of 0.658 (0.014 SD) from having a uniform pool of the same gender teachers. For the total effect, in absolute terms, an increase in the probability of having a female teacher of 100 per cent increases the enrollment of female students by 0.29 and decreases the enrollment of boys by 0.37. In terms of the practical effect size, an increase of 1 SD in the probability of having the same gender teacher increase increases enrollment of girls (boys) by 0.01 SD (0.01 SD). This is similar to the effect that Muralidharan and Sheth (2015) found in Andhra Pradesh, India on math and language test scores but greater than the

Table 8: Effect of having a female teacher on enrollment: pooled across female and male enrollment

	Enrollment			
	(1)	(2)	(3)	(4)
Prop. of female teachers (γ_1)	-0.143	0.232	-3.138***	-0.371**
	(0.163)	(0.167)	(0.281)	(0.169)
Female (γ_2)	-0.222***	-0.222***	-0.222***	-0.222***
	(0.0263)	(0.0263)	(0.0263)	(0.0263)
Prop. of female teachers x Female (γ_3)	**0.658*** **	**0.658*** **	**0.658*** **	**0.658*** **
	(0.0856)	(0.0856)	(0.0856)	(0.0856)
Constant	11.66***	10.87***	12.28***	14.85***
	(0.0745)	(0.492)	(0.0597)	(0.214)
$\gamma_1 + \gamma_3$	**0.51**	**0.89**	**-2.48**	**0.29**
F-test p-value ($H_0 : \gamma_1 + \gamma_3 = 0$)	**0.00**	**0.00**	**0.00**	**0.08**
Observations	916,632	916,504	916,632	916,632
R-squared	0.000	0.058	0.003	0.005
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of tribals, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

zero effect they found on student attendance.

Caste

Table 9 looks at the effect of having a SC teacher on enrollment of SC students relative to a non-SC teacher. There is a positive match effect, with a relative increase of 1.5 (0.01 SD) students in enrollment when students are matched with a pool of same caste teachers compared to different caste teachers. But while having a uniform pool of SC teacher increases the enrollment of SC students relative to a uniform pool of non-SC teachers, the total effect fluctuates a lot across specifications and the $\beta_1 + \beta_3$ equivalent of this caste regression ($\kappa_1 + \kappa_3$) has a F-statistic of 1.15. This means that the SC teachers are overall so ineffective ($\kappa_1 < 0$) that SC students end up being no better off with a pool of SC teachers than non-SC teachers despite the absence of gains from belonging to the same caste in the latter case.

Table 9: Effect of having a SC teacher on enrollment: pooled across SC and non-SC enrollment

	Enrollment			
	(1)	(2)	(3)	(4)
Prop. of SC teachers (κ_1)	6.031***	1.545**	-1.609***	-1.089**
	(0.704)	(0.671)	(0.555)	(0.451)
SC (κ_2)	-27.08***	-27.07***	-27.08***	-27.08***
	(0.234)	(0.234)	(0.234)	(0.234)
Prop. of SC teachers x SC (κ_3)	**1.552****	**1.533****	**1.552****	**1.552****
	(0.657)	(0.657)	(0.657)	(0.657)
Constant	30.43***	22.47***	31.17***	35.73***
	(0.248)	(0.878)	(0.128)	(0.401)
$\kappa_1 + \kappa_3$	**7.58**	**3.07**	**-0.06**	**0.54**
F-test p-value $(H_0 : \kappa_1 + \kappa_3 = 0)$	**0.00**	**0.00**	**0.91**	**0.28**
Observations	916,632	916,504	916,632	916,632
R-squared	0.199	0.249	0.293	0.331
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

7 Conclusion

My paper provides strong evidence that ethnic dynamics between a student and a teacher matter for student retention in school in a persistent manner.

The results suggest that teachers are more effective at the extensive margin when teaching co-ethnic students, that enrollment of tribals increases with an increase in probability of being taught by tribal teachers relative to non-tribal teachers. We find that non-tribals suffer adverse effects when taught by tribal teachers relative to non-tribal teachers, about a third of them drop out of school all together and the remaining switch to another school as the probability of having a tribal teacher increases. These results do not differ by student gender or grade type. These ethnicity match effects are stronger than the gender match and caste match effects. The identification concerns of non-random teacher allocation and reverse causality are allayed by providing qualitative evidence and by showing that the estimates do not change substantially under increasingly restrictive set of specifications.

Using a simple theoretical model I distinguish between active and passive effects underlying the empirically established aggregate match effect. The model predicts that we should be able to distinguish between these two effects based on the sign of the cross-partial between student and teacher ability with and without an ethnic match. If the cross-partial is positive with an ethnic match then the main mechanism is a type of active effect such as a decrease in discrimination from switching to a co-ethnic teacher. On the other hand, if the cross-partial is negative with an ethnic match then the main mechanism is a type of passive effect such as an increase in complementarity from being taught by a co-ethnic teacher.

A potential policy response to such an effect could be to hire more tribal teachers for tribal students which will enable India to achieve its goal of universal elementary education. However, this is not recommended for a variety of reasons. Firstly, in a diverse society like India matching students to teachers on demographic traits would amount to segregating schools on the basis of some salient characteristic[64] which is unconstitutional. Secondly, such a policy would, as the results indicate, have undesirable consequences for students who do not share the teacher's demographic traits, in this case the non-tribals. It is for these very reasons that identification of the

[64] Such as caste, religion or ethnicity.

exact mechanisms behind the aggregate match effect becomes important. If it is biases in teacher behavior that manifest themselves in overt discrimination against *out-group* students then the key policy recommendation would be to sensitize teachers to the consequences of their actions, to train them to cope with their biases in a way which would not harm the students and in the extreme case, setting out appropriate disciplinary measures for any discriminatory activities. On the contrary, if it is the positive role model effects that are at play, then the students will have to be educated on the role of a teacher and joint activities would have to be conducted at school to minimize the difference as perceived by the student between themselves and their teacher. Similarly, if it is stereotype threats that retard cognition and cause drop outs then the policy prescription would be to structure classroom interactions such that they inculcate a sense of self-worth in students, highlighting their potential and make them view any difficulty as a challenge as opposed to a cause for remediation (Steele 1997).

In essence, the most sustainable and promising way of closing achievement gaps is to improve the effectiveness of all teachers and achieve demographic neutrality in student-teacher interactions. This can be achieved through programs that educate and train teachers in the nuances of pedagogy and by instituting well-designed performance incentives.

8 References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. Journal of Labor Economics, 25(1), 95-135.

Acemoglu, D. (2002). Directed Technical Change. The Review of economic studies, 69(4), 781-809.

Adukia, A. (2013). Sanitation and Education. Cambridge, MA: Harvard Graduate School of Education.

Agrawal, T. (2012). Returns to education in India: some recent evidence. Journal of Quantitative Economics, 10(2).

Ahluwalia, M. S. (2011). Prospects and policy challenges in the twelfth plan. Economic and Political Weekly, 46(21), 88-104.

Aslam, M., & Kingdon, G. (2007). What can teachers do to raise pupil achievement? (No. 2007-14). Centre for the Study of African Economies, University of Oxford.

Atal, Y. (Ed.). (2009). Sociology and social anthropology in India. Pearson Education India.

Bapna, A., & Sharma, N. (2015). School Education in India: A Handbook.

Centre for Budget and Policy Studies, Bangalore, India (2015). National Study on Working Conditions of Teachers: State Report for Jharkhand.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. The American Economic Review, 104(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. The American Economic Review, 104(9), 2633-2679.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. Economics of Education review, 24(4), 377-392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26(6), 673-682.

Cunha, F., & Heckman, J. J. (2007). The Technology of Skill Formation. American Economic Review, 97(2):31–47.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Journal of Human Resources, 43(4), 738-782.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. Review of Economics and Statistics, 86(1), 195-210.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. American Economic Review, 158-165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. Journal of Human Resources, 42(3), 528-554.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. Industrial & Labor Relations Review, 48(3), 547-561.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2011). A community college instructor like me: Race and ethnicity interactions in the classroom (No. w17381). National Bureau of Economic Research.

Fisman, R., Paravisini, D., & Vig, V. (2012). Cultural proximity and loan outcomes (No. w18096). National Bureau of Economic Research.

Glewwe, P., Hanushek, E. A., Humpage, S., & Ravina, R. (2013). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. Education Policy in Developing Countries, 4(14,972), 13.

Hall, G. H., & Patrinos, H. A. (Eds.). (2012). Indigenous peoples, poverty, and development. Cambridge University Press.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. Journal of public economics, 95(7), 798-812.

Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher?. Labour Economics, 15(1), 37-53.

Human Rights Watch (2014). "They say we're dirty" Denying an education to India's marginalized.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina (No. w18624). National Bureau of Economic Research.

Jencks, C., & Phillips, M. (Eds.). (1998). The Black-White test score gap. Brookings Institution Press.

Jharkhand Government, Human Resource Department (2011). The Jharkhand Right of Children to Free and Compulsory Education Rules, 2011.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation (No. w14607). National Bureau of Economic Research.

Kingdon, G., & Rawal, S. (2010). Akin to my teacher: Does caste, religious or gender distance between student and teacher matter? Some evidence from India (No. 10-18). Department of Quantitative Social Science-Institute of Education, University of London.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. Journal of Public Economics, 92(10), 2083-2105.

Maharatna, A. (2005). Demographic perspectives on India's tribes. Oxford University Press.

Ministry of Tribal Affairs, Government of India (2014). Report of the High Level Committee on Socio- economic, Health and Educational Status of Tribal Communities of India.

Mitra, A. (2008). The status of women among the scheduled tribes in India. The Journal of Socio-Economics, 37(3), 1202-1217.

Muralidharan, K., & Sheth, K. (2015). Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers. The Journal of Human Resources (forthcoming)

Muralidharan, K., & Sundararaman, V. (2013). Contract teachers: Experimental evidence from India (No. w19440). National Bureau of Economic Research.

Nagavarapu, S., & Sekhri, S. (2014). Informal Monitoring and Enforcement Mechanisms in Public Service Delivery: Evidence from the Public Distribution System in India. Tech. rep., Mimeo, Brown University.

National Scientific Council on the Developing Child (2007). The timing and quality of early experiences combine to shape brain architecture: Paper No. 5.

Nayar, R., Gottret, P., Mitra, P., Betcherman, G., Lee, Y. M., Santos, I., Dahal, M. & Shrestha, M. (2012). "More and Better Jobs in South Asia," World Bank Publications, The World Bank, number 2391.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417-458.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American psychologist, 52(6), 613.

United Nations Development Program (2011). Jharkhand: Economic and Human Development Indicators. www.in.undp.org.

Winters, M. A., Haight, R. C., Swaim, T. T., & Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. Economics of Education Review, 34, 69-75.

9 Appendix

Table A1: Changes is enrollment by ethnicity and grade

		Enrollment		
		(1)	(2)	(3)
Grade 1	Tribal	-13.25***	-13.25***	-13.25***
		(0.156)	(0.156)	(0.156)
	Constant	18.82***	20.82***	22.83***
		(0.128)	(0.161)	(0.115)
Grade 2	Tribal	-11.32***	-11.32***	-11.32***
		(0.130)	(0.130)	(0.130)
	Constant	16.10***	18.21***	18.01***
		(0.107)	(0.137)	(0.0881)
Grade 3	Tribal	-11.18***	-11.18***	-11.18***
		(0.129)	(0.129)	(0.129)
	Constant	15.91***	17.46***	17.02***
		(0.107)	(0.133)	(0.0845)
Grade 4	Tribal	-10.66***	-10.66***	-10.66***
		(0.126)	(0.126)	(0.126)
	Constant	15.08***	16.36***	15.92***
		(0.106)	(0.132)	(0.0833)
Grade 5	Tribal	-10.22***	-10.22***	-10.22***
		(0.124)	(0.124)	(0.124)
	Constant	14.43***	15.45***	14.31***
		(0.107)	(0.135)	(0.0790)
Grade 6	Tribal	-25.99***	-25.99***	-25.99***
		(0.431)	(0.431)	(0.431)
	Constant	33.17***	37.49***	33.53***
		(0.399)	(0.575)	(0.309)
Grade 7	Tribal	-24.27***	-24.27***	-24.27***
		(0.424)	(0.424)	(0.424)
	Constant	30.68***	33.00***	29.02***
		(0.401)	(0.575)	(0.313)
Grade 8	Tribal	-21.32***	-21.32***	-21.32***
		(0.406)	(0.406)	(0.406)
	Constant	26.84***	25.87***	21.95***
		(0.392)	(0.522)	(0.275)
Cohort FE			Yes	Yes
School FE				Yes

Note: (1) The number of observations for all primary grade regressions is 155,850 and for the upper primary grades is 45,794. (2) All standard errors are robust and clustered at the school level. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Table A2: Persistence of the effect of having a tribal teacher on enrollment (over two years): pooled across tribal and non-tribal enrollment

	Enrollment at time t			
	(1)	(2)	(3)	(4)
Prop. of tribal teachers at time t	-11.17***	-10.99***	-8.484***	-6.998***
	(0.699)	(0.708)	(0.713)	(0.455)
Prop. of tribal teachers at time t-1	-1.637***	-1.653***	-2.066***	-2.024***
	(0.566)	(0.557)	(0.412)	(0.411)
Prop. of tribal teachers at time t-2	-7.001***	-7.053***	-5.675***	-5.712***
	(0.644)	(0.632)	(0.406)	(0.404)
Tribal	-21.34***	-21.33***	-21.34***	-21.34***
	(0.247)	(0.247)	(0.247)	(0.247)
Prop. of tribal teachers at time t x Tribal	14.64***	14.65***	14.64***	14.64***
	(0.786)	(0.787)	(0.786)	(0.786)
Prop. of tribal teachers at time t-1 x Tribal	3.517***	3.521***	3.517***	3.517***
	(0.687)	(0.687)	(0.687)	(0.687)
Prop. of tribal teachers at time t-2 x Tribal	11.56***	11.54***	11.56***	11.56***
	(0.762)	(0.762)	(0.762)	(0.762)
Constant	23.27***	23.31***	22.80***	24.28***
	(0.240)	(0.733)	(0.156)	(0.391)
Observations	223,138	223,090	223,138	223,138
R-squared	0.274	0.291	0.372	0.406
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

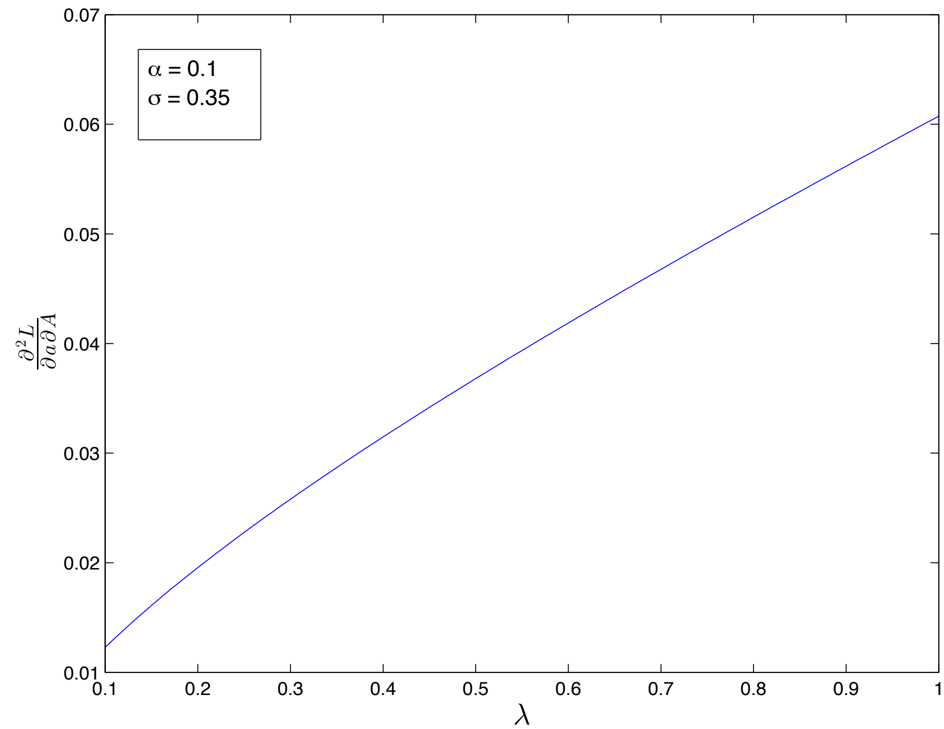
Table A3: Effect of having a tribal teacher on enrollment by gender: pooled across male and female enrollment

	Enrollment			
	(1)	(2)	(3)	(4)
Prob. of grade having a tribal teacher (ρ_1)	-4.688***	-3.706***	0.0174	0.0442
	(0.135)	(0.153)	(0.200)	(0.148)
Female (ρ_2)	-0.0465	-0.0467	-0.0465	-0.0465
	(0.0329)	(0.0329)	(0.0329)	(0.0329)
Prob. of class having a tribal teacher x Female (ρ_3)	**-0.179*****	**-0.178*****	**-0.179*****	**-0.179*****
	(0.0553)	(0.0553)	(0.0553)	(0.0553)
Constant	12.63***	9.556***	11.63***	14.76***
	(0.0832)	(0.436)	(0.0459)	(0.215)
$\rho_1 + \rho_3$	**-4.87**	**-3.88**	**-0.16**	**-0.13**
F-test p-value ($H_0 : \rho_1 + \rho_3 = 0$)	**0.00**	**0.00**	**0.41**	**0.36**
Observations	916,632	916,504	916,632	916,632
R-squared	0.022	0.064	0.000	0.004
Village level conrols		Yes		
District FE		Yes		
School FE			Yes	
School x Grade FE				Yes
Teacher characteristics				Yes

Note: (1) All standard errors are robust and clustered at the school level. (2) Village level controls include proportion of tribal population, a wealth index, employment rate in agriculture, household industry and cultivation for 6 months or more, male literacy rate, female literacy rate. (3) Teacher's characteristics at the shcool-grade include the proportion of females, the average age, average years of experience, proportion of contract teachers, proportion of teachers who have completed high school, college and masters separately, proportion of teachers with a professional degree (B.Ed or M.Ed equivalent) and the proportion of teachers who live in the village where the school is located. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Figure A1: Change in the cross partial of learning wrt student and teacher ability with a decrease in discrimination

(a) The effects with a low factor share of student ability (Low α)



(b) The effects with a high factor share of student ability (High α)

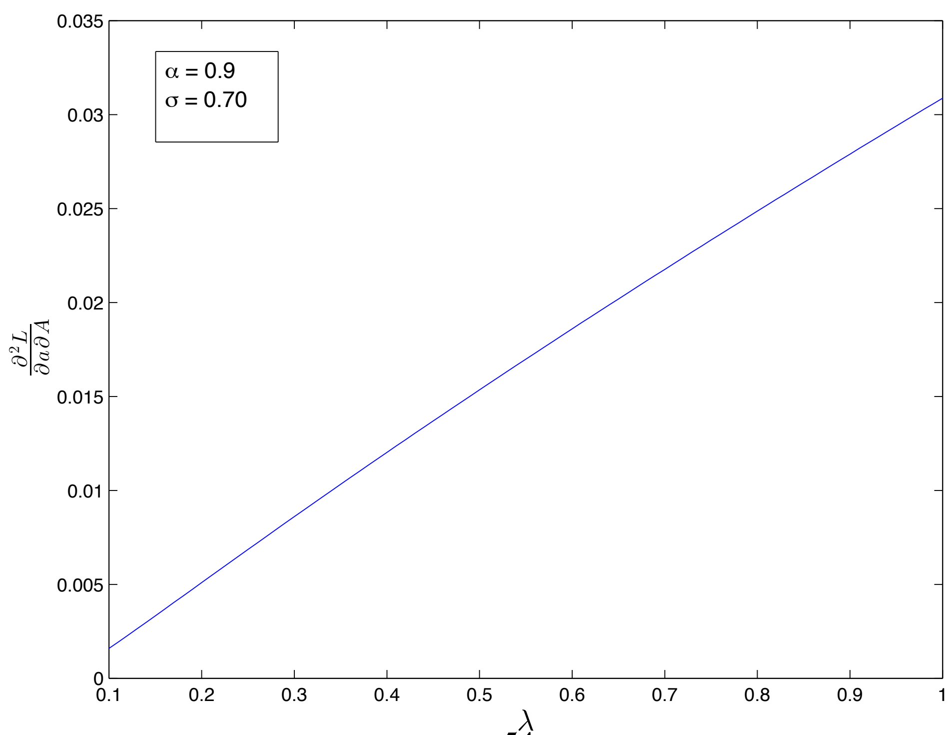
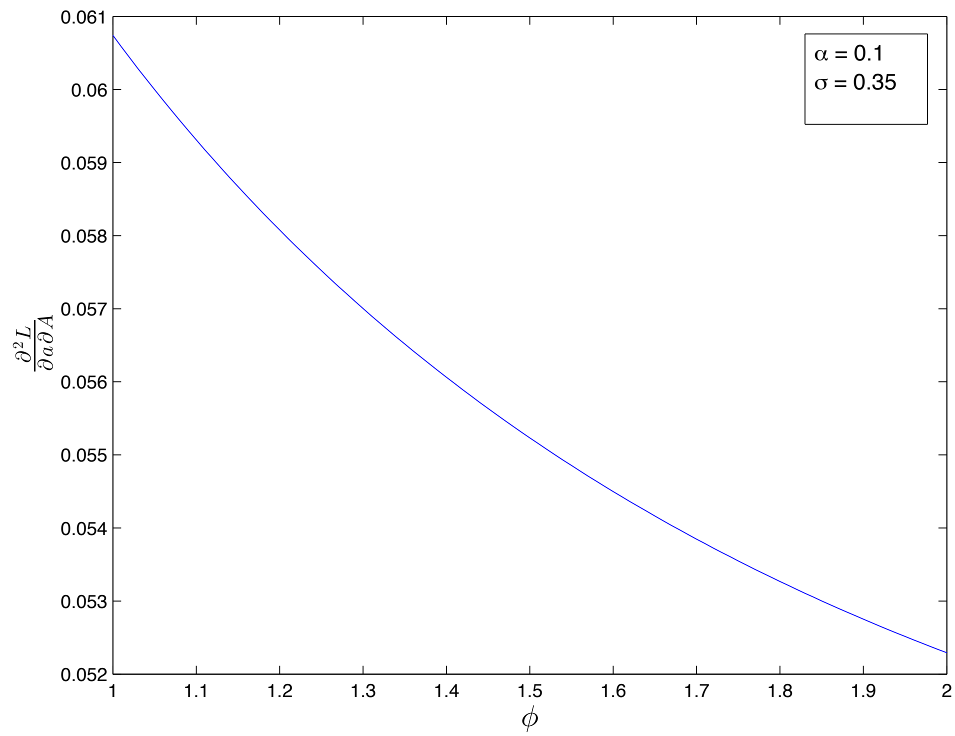
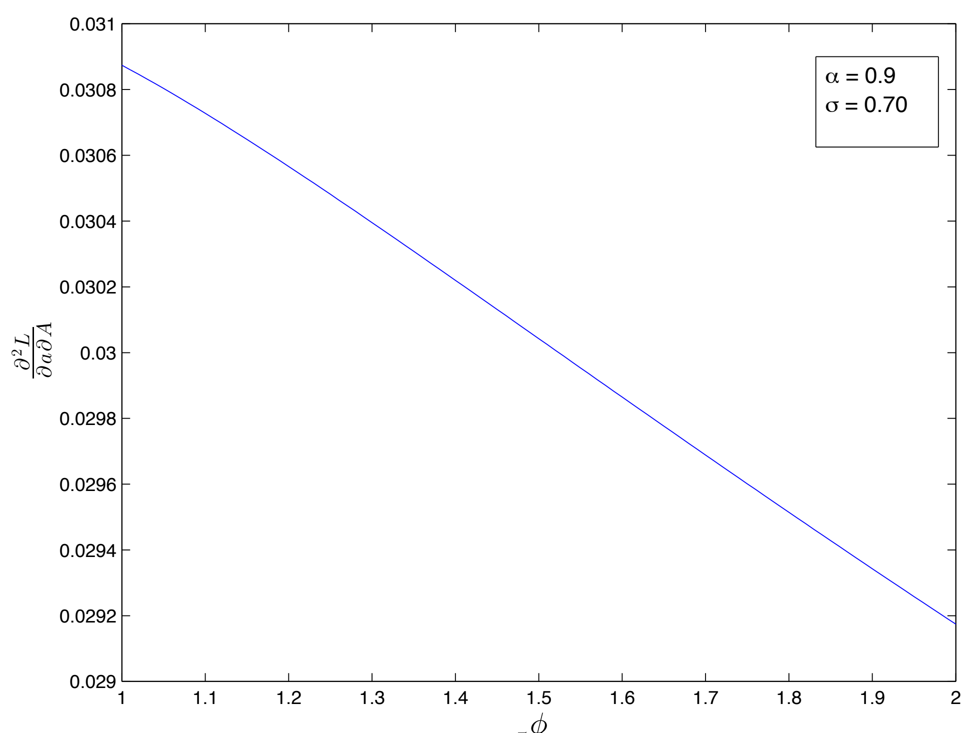


Figure A2: Change in the cross partial of learning wrt student and teacher ability with an increase in complementarity

(a) The effects with a low factor share of student ability (Low α)



(b) The effects with a high factor share of student ability (High α)



10 Mathematical Appendix

10.1 Benchmark case

The teacher's maximization problem is:

$$Max_e V = \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha) e^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

The FOC is:

$$\left[\frac{\partial V}{\partial e}\right]_{e=e^*} = 0 \Longrightarrow \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha) e^{*\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} = \frac{c(A)}{(1-\alpha)} e^{*\frac{1}{\sigma}}$$

solving for e^* gives us

$$e^* = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\left(\left(\frac{c(A)}{1-\alpha}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}}$$

Using this e^* I have the following comparative statics:

1. <u>Teacher effort</u>

   - $\frac{\partial e^*}{\partial a} = \frac{e^*}{a} = \frac{\alpha^{\frac{\sigma}{\sigma-1}}}{\left(\left(\frac{c(A)}{1-\alpha}\right)^{\sigma-1} - 1+\alpha\right)^{\frac{\sigma}{\sigma-1}}} > 0$

   - $\frac{\partial e^*}{\partial A} = -\frac{c'(A)\, a \alpha^{\frac{\sigma}{\sigma-1}} \frac{\sigma}{1-\alpha} \left(\frac{c(A)}{1-\alpha}\right)^{\sigma-1}}{\left(\left(\frac{c(A)}{1-\alpha}\right)^{\sigma-1} - 1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}} > 0$ given $c'(A) < 0$

2. <u>Student learning</u>

   - $\frac{\partial L^*}{\partial a} = \frac{\partial L(a,e^*)}{\partial a} = \underbrace{\frac{\partial L}{\partial a}}_{+} + \underbrace{\frac{\partial L}{\partial e^*}}_{+} \underbrace{\frac{\partial e^*}{\partial a}}_{+} > 0$

   - $\frac{\partial L^*}{\partial A} = (1-\alpha) e^{*\frac{-1}{\sigma}} \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha) e^{*\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \frac{\partial e^*}{\partial A} > 0$

- To derive $\frac{\partial^2 L^*}{\partial A \partial a}$ I use simple chain rule:

$$\begin{aligned} L^* &= L(a, e^*) \\ \frac{\partial L^*}{\partial A} &= \frac{\partial L}{\partial e^*}\frac{\partial e^*}{\partial A} \\ \frac{\partial^2 L^*}{\partial a \partial A} &= \frac{\partial^2 L}{\partial a \partial e^*}\frac{\partial e^*}{\partial A} + \frac{\partial L}{\partial e^*}\frac{\partial^2 e^*}{\partial a \partial A} \\ \frac{\partial^2 L}{\partial A \partial a} &= \frac{\partial^2 L}{\partial a \partial e^*}\frac{\partial e^*}{\partial A} + \frac{\partial L}{\partial e^*}\frac{\partial^2 e^*}{\partial a \partial A} \quad > \quad 0 \quad where \quad \frac{\partial^2 e^*}{\partial a \partial A} = a\frac{\partial e^*}{\partial A} > 0 \end{aligned}$$

10.2 With discrimination

With discrimination, the teacher's problem is:

$$Max_e V_\lambda = \lambda L(a,e) - c(A)e = \lambda\left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)e^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

where $\lambda \in (0,1)$. λ keeps decreasing ($\rightarrow 0$) with increases in discrimination. The FOC is:

$$\left[\frac{\partial V_\lambda}{\partial e}\right]_{e=e^*_\lambda} = 0 \Longrightarrow \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha){e^*_\lambda}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} = \frac{c(A)}{\lambda(1-\alpha)}{e^*_\lambda}^{\frac{1}{\sigma}}$$

solving for e^*_λ gives us

$$e^*_\lambda = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}}$$

Using this e^*_λ I have the following comparative statics:

1. Teacher effort

- $\frac{\partial e^*_\lambda}{\partial \lambda} = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a \sigma}{\lambda\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} > 0$

- $\frac{\partial e^*_\lambda}{\partial a} = \frac{e^*_\lambda}{a} = \frac{\alpha^{\frac{\sigma}{\sigma-1}}}{\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}} > 0$

- $\frac{\partial^2 e^*_\lambda}{\partial a \partial \lambda} = \frac{\alpha^{\frac{\sigma}{\sigma-1}} \sigma}{\lambda\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} > 0$

- $\frac{\partial e^*_\lambda}{\partial A} = -\frac{c'(A)\,\alpha^{\frac{\sigma}{\sigma-1}} a\sigma}{\lambda(1-\alpha)\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-2} > 0$ given $c'(A) < 0$

- $\frac{\partial^2 e^*}{\partial A\partial\lambda} = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a\sigma(\sigma-1)}{\lambda(1-\alpha)\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-2}\frac{c'(A)}{\lambda(1-\alpha)}\times$

$\left(1-\frac{\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}}{\left[\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right]}\frac{2\sigma-1}{\sigma-1}\right) > 0$ if $\sigma \in [\frac{1}{2},1)$

2. Student learning

- $\frac{\partial^2 L^*_\lambda}{\partial\lambda\partial a} = \frac{\partial^2 L}{\partial\lambda\partial a}\lambda L(a,e^*_\lambda) = \underbrace{\frac{\partial L}{\partial a}}_{(+)} + \underbrace{\left(\underbrace{\frac{\partial L}{\partial e^*_\lambda}}_{(+)} + \lambda\underbrace{\frac{\partial^2 L}{\partial e^*_\lambda\partial\lambda}}_{(-)}\right)}_{(+)}\underbrace{\frac{\partial e^*_\lambda}{\partial a}}_{(+)} + \lambda\underbrace{\frac{\partial^2 L}{\partial a\partial\lambda}}_{(+)} + \lambda\underbrace{\frac{\partial L}{\partial e^*_\lambda}}_{(+)}\underbrace{\frac{\partial^2 e^*_\lambda}{\partial a\partial\lambda}}_{(+)} >$
0
  - where $\frac{\partial^2 L}{\partial a\partial\lambda} = \frac{\partial L}{\partial e^*_\lambda}\frac{\partial^2 e^*_\lambda}{\partial a\partial\lambda} + \frac{\partial^2 L}{\partial e^*_\lambda\partial a}\frac{\partial e^*_\lambda}{\partial\lambda} > 0$
  - and $\frac{\partial L}{\partial e^*_\lambda} + \lambda\frac{\partial^2 L}{\partial e^*_\lambda\partial\lambda} = (1-\alpha)e_\lambda^{*\frac{-1}{\sigma}}\left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}$

$$\times\left(1-\underbrace{\frac{\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}}{\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-(1-\alpha)}}_{>1}\underbrace{\frac{\alpha a^{\frac{\sigma-1}{\sigma}}}{\alpha a^{\frac{\sigma-1}{\sigma}}+(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}}}_{\in(0,1)}\right) > 0$$

if $\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1} \geq \frac{1-\alpha}{1-\gamma}$ where $\gamma = \frac{\alpha a^{\frac{\sigma-1}{\sigma}}}{\alpha a^{\frac{\sigma-1}{\sigma}}+(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}}$

- $\frac{\partial^2 L^*_\lambda}{\partial\lambda\partial A} = \frac{\partial L}{\partial e^*_\lambda}\frac{\partial e^*_\lambda}{\partial A} + \lambda\left(\frac{\partial^2 L}{\partial a\partial e^*_\lambda}\frac{\partial e^*_\lambda}{\partial A} + \frac{\partial L}{\partial e^*_\lambda}\frac{\partial^2 e^*_\lambda}{\partial A^2}\right) > 0$ given that $c(A)$ is concave and with $\sigma \in [\frac{1}{2},1)$

- To derive $\frac{\partial}{\partial\lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A\partial a}\right)$ I again use simple chain rule:

$$\begin{aligned}
L^*_\lambda &= \lambda L(a,e^*)\\
\frac{\partial L^*_\lambda}{\partial A} &= \lambda \frac{\partial L}{\partial e^*}\frac{\partial e^*}{\partial A}\\
\frac{\partial^2 L^*_\lambda}{\partial a \partial A} &= \lambda\left(\frac{\partial^2 L}{\partial a\partial e^*}\frac{\partial e^*}{\partial A}+\frac{\partial L}{\partial e^*}\frac{\partial^2 e^*}{\partial a\partial A}\right)\\
\frac{\partial}{\partial\lambda}\left(\frac{\partial^2 L^*_\lambda}{\partial A\partial a}\right) &= \left(\underbrace{\frac{\partial^2 L}{\partial a\partial e^*_\lambda}+\lambda\frac{\partial^3 L}{\partial a\partial e^*_\lambda\partial\lambda}}_{(+)}\right)\underbrace{\frac{\partial e^*_\lambda}{\partial A}}_{(+)}\\
&\quad+\left(\underbrace{\frac{\partial L}{\partial e^*_\lambda}+\lambda\frac{\partial^2 L}{\partial e^*_\lambda\partial\lambda}}_{(+)}\right)\underbrace{\frac{\partial^2 e^*_\lambda}{\partial a\partial A}}_{(+)}\\
&\quad+\underbrace{\lambda\frac{\partial^2 L}{\partial a\partial e^*_\lambda}}_{(+)}\underbrace{\frac{\partial^2 e^*_\lambda}{\partial A\partial\lambda}}_{(+)}+\underbrace{\lambda\frac{\partial L}{\partial e^*_\lambda}}_{(+)}\underbrace{\frac{\partial^3 e^*_\lambda}{\partial a\partial A\partial\lambda}}_{(+)} \quad > \quad 0
\end{aligned}$$

where

$$\begin{aligned}
&\frac{\partial^2 e^*_\lambda}{\partial a\partial A} = a\frac{\partial e^*_\lambda}{\partial A} > 0 \text{ and}\\
&\frac{\partial^3 L^*}{\partial a\partial e^*_\lambda\partial\lambda} = \alpha(1-\alpha)\frac{1}{\sigma^2}\frac{\partial e^*_\lambda}{\partial\lambda}a^{\frac{-1}{\sigma}}e_\lambda^{*\frac{-1}{\sigma}-1}\left(\alpha a^{\frac{\sigma-1}{\sigma}}+(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}\right)^{\frac{2-\sigma}{\sigma-1}}\\
&\qquad\times\left(1-(\underbrace{2-\sigma}_{>1})\underbrace{\frac{(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}}{\alpha a^{\frac{\sigma-1}{\sigma}}+(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}}}_{\in(0,1)}\right) > 0\\
&\qquad\text{if} \qquad \frac{(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}}{\alpha a^{\frac{\sigma-1}{\sigma}}+(1-\alpha)e_\lambda^{*\frac{\sigma-1}{\sigma}}} \geq \frac{1}{2-\sigma}\\
&\frac{\partial^3 e^*_\lambda}{\partial a\partial A\partial\lambda} = \frac{\alpha^{\frac{\sigma}{\sigma-1}}\sigma(\sigma-1)}{\lambda(1-\alpha)\left(\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-2}\frac{c'(A)}{\lambda(1-\alpha)}\left(1-\frac{\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}}{\left[\left(\frac{c(A)}{\lambda(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right]}\frac{2\sigma-1}{\sigma-1}\right)\\
&\qquad\text{if } \sigma\in[\tfrac{1}{2},1)
\end{aligned}$$

10.3 Increased complementarity

With increased complementarity or increased marginal returns to effort, a teacher's maximization problem is:

$$Max_e V_\phi = L(a,\phi e) - c(A)e = \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(\phi e)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - c(A)e$$

where $\phi > 1$. Now the FOC is:

$$\left[\frac{\partial V_\phi}{\partial e}\right]_{e=e_\phi^*} = 0 \Longrightarrow \left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(\phi e_\phi^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} = \frac{c(A)}{\phi(1-\alpha)}(\phi e_\phi^*)^{\frac{1}{\sigma}}$$

solving for e_ϕ^* gives us

$$e_\phi^* = \frac{\alpha^{\frac{\sigma}{\sigma-1}} a}{\phi\left(\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}}$$

Using this e_ϕ^* I have the following comparative statics:

1. Teacher effort

- $\frac{\partial e_\phi^*}{\partial \phi} = -\frac{e_\phi^*}{\phi}\left[1 - \sigma\frac{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}}{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - (1-\alpha)}\right] < 0$

  if $\left[1 - \sigma\frac{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}}{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - (1-\alpha)}\right] > 0$

  or $\underbrace{\frac{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}}{\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - (1-\alpha)}}_{>1} < \frac{1}{\sigma}$

  therefore, if σ is small enough (in particular if $\sigma < \frac{x-(1-\alpha)}{x} < 1$) where $x = \left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}$

  then $\frac{\partial e_\phi^*}{\partial \phi} < 0$

- $\frac{\partial e_\phi^*}{\partial a} = \frac{e_\phi^*}{a} = \frac{\alpha^{\frac{\sigma}{\sigma-1}}}{\phi\left(\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1} - 1 + \alpha\right)^{\frac{\sigma}{\sigma-1}}} > 0$

- $\frac{\partial^2 e_\phi^*}{\partial a \partial \phi} = \frac{1}{a}\frac{\partial e_\phi^*}{\partial \phi} < 0$ because $\frac{\partial e_\phi^*}{\partial \phi} < 0$ from above.

- $\frac{\partial e^*_\phi}{\partial A} = -\frac{c'(A)\,\alpha^{\frac{\sigma}{\sigma-1}}\,a\sigma}{\phi^2(1-\alpha)\left(\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-2} > 0$ given $c'(A) < 0$

- $\frac{\partial^2 e^*_\phi}{\partial A \partial \phi} = -\frac{1}{\phi}\sigma\underbrace{(\sigma-1)}_{(-)}(1-\alpha)\underbrace{\left(\frac{c'(A)}{\phi(1-\alpha)}\right)}_{(-)}\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}\frac{1}{\left[\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}-(1-\alpha)\right]^2} < 0$

2. Student learning

- $\frac{\partial^2 L^*_\phi}{\partial \phi \partial a} = \frac{\partial^2 L(a,\tilde{e}^*)}{\partial \phi \partial a} = \underbrace{\frac{\partial^2 L}{\partial \phi \partial a}}_{(+)} + \underbrace{\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial a}}_{(+)} + \phi\underbrace{\left(\frac{\partial^2 L}{\partial \phi \partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial a} + \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial^2 e^*_\phi}{\partial \phi \partial a}\right)}_{(-)}$

  but for $\sigma < \frac{x-(1-\alpha)}{x}$ the $(+)$ terms outweigh the $(-)$ terms.

- To derive $\frac{\partial^2 L^*_\phi}{\partial \phi \partial A}$ I use simple chain rule:

$$L_\phi = L(a, \tilde{e}^*)$$

  where $\tilde{e}^* = \phi e^*_\phi$, now first differentiating wrt A

$$\frac{\partial L_\phi}{\partial A} = \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial \tilde{e}^*}{\partial e^*_\phi}\frac{\partial e^*_\phi}{\partial A} = \phi\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A}$$

  differentiating wrt ϕ

$$\frac{\partial^2 L_\phi}{\partial \phi \partial A} = \underbrace{\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A}}_{(+)} + \phi\underbrace{\left(\frac{\partial^2 L}{\partial \phi \partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A} + \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial^2 e^*_\phi}{\partial \phi \partial A}\right)}_{(-)}$$

  where $\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial e}{\partial A} = -c'(A)\,\sigma\alpha^{\frac{\sigma}{\sigma-1}}a\,\tilde{e}^{*\frac{-1}{\sigma}}\left(\alpha a^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(\phi e^*_\phi)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}$
  $\times\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-2}\frac{1}{\phi^2\left(\left(\frac{c(A)}{\phi(1-\alpha)}\right)^{\sigma-1}-1+\alpha\right)^{\frac{2\sigma-1}{\sigma-1}}}$, please note that σ appears in the numerator.
  but with a small σ I have

$$\underbrace{\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A}}_{(+)} < \phi\underbrace{\left(\frac{\partial^2 L}{\partial \phi \partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A} + \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial^2 e^*_\phi}{\partial \phi \partial A}\right)}_{(-)} \implies \frac{\partial^2 L_\phi}{\partial \phi \partial A} < 0$$

- To derive $\frac{\partial}{\partial \lambda}\left(\frac{\partial^2 L^*_\phi}{\partial A \partial a}\right)$ I use simple chain rule:

$$L_\phi = L(a, \tilde{e}^*)$$

where $\tilde{e}^* = \phi e^*_\phi$, now first differentiating wrt A

$$\frac{\partial L_\phi}{\partial A} = \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial \tilde{e}^*}{\partial e^*_\phi}\frac{\partial e^*_\phi}{\partial A}$$

differentiating wrt a

$$\frac{\partial^2 L_\phi}{\partial a \partial A} = \phi\left(\frac{\partial^2 L}{\partial a \partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A} + \frac{\partial L}{\partial \tilde{e}^*}\frac{\partial^2 e^*_\phi}{\partial a \partial A}\right)$$

and finally differentiating wrt the complementarity parameter ϕ

$$\frac{\partial}{\partial \phi}\left(\frac{\partial^2 L_\phi}{\partial A \partial a}\right) = \underbrace{\underbrace{\frac{\partial^2 L}{\partial a \partial \tilde{e}^*}\frac{\partial e^*_\phi}{\partial A}}_{(+)} + \underbrace{\frac{\partial^3 L}{\partial a \partial \tilde{e}^* \partial \phi}}_{(-)}\underbrace{\frac{\partial e^*_\phi}{\partial A}}_{(+)}}_{(-)} + \underbrace{\underbrace{\frac{\partial L}{\partial \tilde{e}^*}\frac{\partial^2 e^*_\phi}{\partial a \partial A}}_{(+)} + \underbrace{\frac{\partial^2 L}{\partial \tilde{e}^* \partial \phi}}_{(-)}\underbrace{\frac{\partial^2 e^*_\phi}{\partial a \partial A}}_{(+)}}_{(-)}$$

$$+ \underbrace{\frac{\partial^2 L}{\partial a \partial \tilde{e}^*}}_{(+)}\underbrace{\frac{\partial^2 e^*_\phi}{\partial A \partial \phi}}_{(-)} + \underbrace{\frac{\partial L}{\partial \tilde{e}^*}}_{(+)}\underbrace{\frac{\partial^3 e^*_\phi}{\partial a \partial A \partial \phi}}_{(-)}$$