# Ethnic Fragmentation and School Provision in India[*]

Chadha, Nishant

*India Development Foundation*

Nandwani, Bharti

*Indira Gandhi Institute of Development Research*

September 14, 2018

**Abstract**

A large number of studies in economics and political science contend that communities in ethnically diverse places, due to lower ability to act collectively, fare poorly in providing public services. This paper aims to present a more nuanced relationship between diversity and provision of public goods by incorporating the mechanism of provision of public goods in it. The argument is that the degree to which local community can influence public goods provision varies with different types of services and therefore we do not expect all types of public goods to be impacted in the same way. We support our argument by studying the provision of different types of schools in Indian districts. Schools which are financed by community (private and local government schools) are shown to be lower in number in diverse districts whereas public schools, for which community has little discretionary financing powers, are not impacted. However, we show that public schools rely on active community action for effective functioning of schools, and therefore they are found to be of poor quality. Extensive empirical tests have been performed to support the mechanism.

**Keywords:** Ethnic fragmentation,private schools, public schools, public good provision
**JEL Classification Numbers:H4, O2, I10**

[*]We would like to thank Rachel Gisselquist, Miguel Nio-Zaraza, Arne Bigsten, Alok Kumar, seminar participants at UNU-WIDER, DSE winter school and University of Gothenburg for helpful comments. We would also like to show our gratitude to UNU-WIDER, where a part of this research was conducted. The usual disclaimer applies.

# 1 Introduction

It has long been argued that provision of public goods and services like education, health facilities, roads and other infrastructure, etc aid in the long-term economic development of societies. Amongst many explanations that have been put forward to explain lack of such services, one which has received much scholarly attention is that ethnic diversity hampers public goods provision. Extensively tested both in developed and developing country context and for a variety of public goods (Alesina and La Ferrara, 2005; Alesina, Baqir and Easterly, 1999; Banerjee and Somanathan, 2007; Miguel and Gugerty, 2005), this relationship is so influential that it is regarded as “one of the most powerful hypotheses in political economy” (Banerjee, Iyer and Somanathan, 2005). The basic premise underlying this relationship is that communities in diverse regions fail to act collectively (broadly because of differing preferences (Alesina, Baqir and Easterly, 1999) or inability to impose social sanctions (Miguel and Gugerty, 2005)) to provide public services. However, this paper argues that public goods provision is not solely an outcome of local collective action. Multiple levels of government have critical financing roles to play and therefore the degree to which community action can influence these decisions varies with the type of public good. This suggests that not all types of public goods are expected to be affected by diversity in the same manner. We, therefore, hypothesise that the impact of diversity on public goods is sensitive to how public goods are actually provided and how important community action is in its provision.

This idea has been also been discussed by recent work(Gisselquist, Leiderer and Niño-Zarazúa, 2016; Lee, 2017) which has criticised the conventional relationship and argued that since ethnic fragmentation is essentially a local demographic characteristic, therefore, public goods which are not provided by the local community/government, might not be expected to be affected by local ethnic diversity. We formalise this idea by looking at the impact of ethnic diversity on the provision of different types of schools in Indian districts. To the best of our knowledge, this presents one of the first works to incorporate the mechanism of provision of public goods in the relationship between ethnic fragmentation and public goods.

We focus on schools in this analysis, even though our research question applies to other public goods like health facilities and roads, because the impact of human capital, in general, and education, in particular, on long-term development has been extensively established (Barro, 2001; Romer, 1990) pointing towards urgent need to carefully examine factors that explain lack of quality schools. Additionally, the nature of provision of schools in India is such that it allows us to study our research question. Schools are provided by many different entities; there are public schools provided by multiple levels of government (central, state or local) where students are not required to pay fees and all the decisions concerning infrastructure, teacher hiring, curriculum, etc are made by the respective governments. Then there are private unaided schools which are fees charging institutions provided by autonomous trust or society which makes decisions on teacher hiring, curriculum and other day to day operations. There are also a few private aided schools in India which, even though are named private because of a private management body, effectively depend on government for their funding and follow government rules for hiring teachers and are not allowed to charge fees[1].

Data on these schools are drawn from two large data sets namely, Indian population Census of 2011 and District Information System for Education (DISE) 2013-14 data. Census of 2011 is the first census to categorize schools into private and public allowing us to study the differential impact. However, school-level information is limited in the census data and hence we turn to a much richer school level dataset provided under the DISE[2]. DISE is the census of all the recognised schools in the country and we use the data for the year 2013-14 as it was one of the first attempts to combine the data on elementary and secondary/higher secondary schools. Also, it is temporally closest to the 2011 census data. To measure our main independent variable, ethnic fragmentation, we construct the Ethnic Fractionalisation Index (EFI), a popularly used measure of fragmentation in the literature, by making use of the population census of 1931 data. Details regarding why we use the historical census to construct EFI are provided in the Data section. For our purpose, we consider caste, an ascriptive category, as a separate ethnic group. Given that the Indian society is

[1] In addition to these private schools, there are schools which are not recognised by the government as they fail to meet the set basic minimum criteria. Because of lack of data availability on these kinds of schools, our analysis restricts itself to only recognised (both aided and unaided) private schools.

[2] Detailed description of the DISE data is provided in the data section

divided along the caste lines with deep social and economic cleavages, diversity based on caste also captures socio-economic divisions in the society.

We begin our empirical analysis by looking at the impact of fragmentation on the *number* of different types of school. Drawing on the existing evidence that it is the lack of collective action that affects public goods provision, we hypothesise lower provision of private unaided and local government schools and no impact on the number of state or central public schools in diverse areas. This is because national/state governments are guided by various legal provisions to provide access to education for all and therefore we believe the local community can exercise only limited discretionary powers in terms of provision. Local government schools and private schools, on the other hand, are an outcome of community members mobilising funds, making collective action among community a critical element in provision of these schools. Results support our hypotheses suggesting that the impact of ethnic fragmentation is sensitive to the mechanism of provision of public good.

Our empirical analysis also examines the quality of schools, an equally important determinant of the extent of human capital formation, to get a complete sense of the impact of fragmentation on the schooling system. Even though public schools do not depend on local community for provision of schools, the government devolves maintenance of quality to the local community and encourages collective action for building school development programs. We, therefore, expect diversity to negatively impact the quality of public schools in diverse areas. Private schools which are even though provided and maintained by the community, it is difficult to a priori say anything about the effect since we observe a truncated distribution of private schools in fragmented districts. Our results, again consistent with the hypothesis, indicate that public schools have poor quality in diverse places and private schools seem to be no different as compared to homogeneous districts.

In addition to documenting a differential negative impact on schools in fragmented districts, we provide appropriate empirical tests to show that it is negative impact on collective action that is driving this differential impact. To support our claim that a weak community network in diverse places makes mobilisation of funds difficult, we provide evidence of a negative impact of ethnic fragmentation on the number of self-help groups and agricultural societies. These organisations represent a platform where community members voluntarily come together and make monetary contributions which are used towards collective goals. Presence and effective functioning of these groups is an indication of collective action that can be used to raise money. We also show a positive correlation between our indication of mechanism of voluntary payments and provision of private schools suggesting that better social network makes it easier to mobilise resources.

For public schools, to show that lack of collective action amongst local population affects maintenance and quality of public schools, we provide evidence that fragmentation negatively affects the participation of community in Parent Teacher Association (PTA) and School Management Development Committee (SMDC) meetings. These meetings, mandated under various legal provisions, represent a platform where community members work together with the school administration to build plans for school development. Inactive participation by community members in these meetings is suggestive of inefficient coordination amongst them to maintain school quality. In addition, we also show a positive correlation between community meetings and school quality.

It is possible that instead of the collective-action explanation suggested by us, other potential factors are driving the observed differential results. We, therefore, rule out other explanations like the level of economic activity; inter-group disparities; political environment; proportion of disadvantaged groups; segregation in preferences across groups, by adding appropriate controls. The rest of the paper is structured as follows: Section 2 provides the literature review, section 3 provides a brief discussion about private and public schools in India, section 4 discusses the data used and the methodology followed to perform our tests, in section 5 we discuss our results and present the mechanism in section 6. We conclude in section 7.

## 2 Literature review

The fragmentation-public goods literature is extensive and many cross country and country-specific studies[3] have examined the role of ethnic diversity in the provision of a range of public goods. In this review, for brevity, we don't provide an exhaustive survey of this literature and instead focus on work that comes closest to our paper. Within the conventional literature that has shown that diversity exerts a negative influence, there are papers which document that the negative impact of diversity also falls on private provision of public goods. Miguel and Gugerty (2005) show that difficulty in imposing social sanctions in diverse areas in rural western Kenya leads to lower primary private school funding and worse school facilities[4]. Chaudhary (2009) finds a lower provision of private primary schools in ethnically diverse districts in colonial India, whereas no impact is observed for schools provided by the provincial government or the local board [5]. This suggests that our result of a lower number of private schools in fragmented places is consistent with the existing work.

In addition to the conventional work in this literature, there exists a body of revisionist work which questions the nature of the established relationship between ethnic diversity and public goods provision (Gisselquist et al. (2016), Singh (2011), Lee (2017))[6]. One of the critiques put forward by this work is that the effect of diversity may be different at national versus subnational levels because of the different roles that different levels of government might play in determining the provision of public goods. Since our paper builds on this idea to introduce new dimensions to the established relationship, therefore, in this section, we also review the body of work which revisits the diversity-debit hypothesis (negative economic, social and political outcomes of ethnic diversity).

[3] See Alesina and La Ferrara (2005) for a literature review

[4] The schools considered in the paper depend on both community and government for their funding but the impact is only considered on private funding whereas our paper also looks at the impact on public schools.

[5] The differential impact that she observes is very close to the results of our paper. However, even though she argues that the negative impact on private schools could be due to low demand for schooling by disadvantaged groups or due to difficulty in mobilising resources in such areas, no formal results are presented in favour of either argument.

[6] The results of these papers broadly indicate that ethnic heterogeneity is not necessarily a deal breaker for development.

Singh (2011) shows that inspite of high levels of ethnic heterogeneity, Kerala, an Indian state, has managed to make impressive advancement in the social sector. Employing a multidimensional subjective feeling of integrity/division to capture ethnic diversity in her analysis, she shows diversity is not necessarily a deal breaker for development. Gisselquist et al. (2016) show that, in the Zambian context, the negative relationship between provision of public goods and ethnic heterogeneity does not necessarily hold at the sub-national level specifically when the public good is funded centrally. The idea of this paper is similar to ours that the implication of ethnic fragmentation for the provision of public good would be different if the good is funded centrally as opposed to being provided locally.

Another paper which comes very close to the idea of our paper is Lee (2017) which argues that the provision of most of the public goods involves multiple levels of government, usually much higher than the unit of study, rather than collective action amongst community. Involvement of non-local factors in the provision of public goods constitutes the core argument of the paper and the results show that villages which have a higher proportion of socially dominant groups see a higher provision of centrally funded goods. Importantly, the paper also finds that ethnic fragmentation has no impact on publicly provided primary schools, consistent with what we find. Since this paper uses a different data set than ours, this lends additional evidence to our finding that fragmentation does not necessarily negatively affect the provision of publicly provided goods.

Schündeln (2013) recognises that the impact of fragmentation can differ by the technology of provision of public goods and shows that diversity infact increases the willingness to privately contribute for provision of public goods. However, no explicit comparison is provided for the different technologies of public goods provision. Also, the finding of this paper contradicts ours probably because unlike this paper, we do not look at the willingness to pay but the actual provision of private and public schools. Thus even though some of the recent work has discussed the importance of incorporating the mechanism of provision of public goods while looking at the impact of fragmentation, we have not come across many papers which have formally done this. We attempt to do that to provide a more nuanced relationship between diversity and public goods provision.

# 3 Private and public schools in India

Any discussion that seeks to explain variation in the number and quality of schools in India is incomplete without explicitly considering the distinction between public and private schools given that in the recent years they have experienced a very different evolutionary path. While the number of private schools have grown by about 35%, public schools have only experienced a growth of 1% from 2010-11 to 2015-16 (Kingdon (2017)). The demand for private schools is much higher than public schools with public schools experiencing a fall in enrollment by 11 million whereas private schools registered an increase of 17.5 million (Kingdon (2017)). Given the rising spread of private schools in India, it is useful to understand the economic background of students that private schools cater to. Our measure of economic background of the households that students belong to is based on the distribution of consumption expenditure obtained using the 68th round of National Sample Survey (NSS) conducted in the year 2011-12. We call a household rich (poor) if its consumption expenditure lies in the highest (lowest) quartile of the consumption distribution and middle class if the expenditure lies in the middle two quartiles.

Table 1 which presents the results indicate that, 38.12% of total students go to private schools and out of these students, 40% of them belong to the rich category, whereas the corresponding number for public schools is 13%. This is not surprising given that rich can afford going to fees charging private schools which are also thought to be better in terms of quality than public schools. However, the fact to which we want to bring attention to is that a large number of middle class and poor households go to private schools even when free of cost public schools are available. About 47% of all those who attend private schools are middle class households and 12% belong to the poor category. A part of this could be driven by legislations like the Right to Education Act of 2009 which mandates private schools to provide 25% of their seats to economically and socially backward children. Nevertheless, this suggests that a large percentage of school going children, including those who come from poor families, depend on private schools. Given this and the discourse that private schools could be a substitute for low quality public schools, it is informative to know whether the negative impact that social divisions have been documented to have on schooling falls on public or private schools or both. Depending on which type of school bears the brunt of diversity, it would suggest the mechanism to rely on to efficiently provide schools in a diverse country like India.

# 4 Data and methodology

We work with a cross-section of 479 Indian districts in 18 major states[7] to conduct our empirical analysis. Two distinct datasets, population Census of 2011 and District Information System for Education (DISE) 2013-14, have been used to construct our dependent variables. The reason for using Census of 2011 is that, unlike previous population censuses, village and town directories in this Census provide a separate classification for private and public schools, which is what we need to study our research question. We add the number of schools in all the villages and towns in a district to arrive at the district level aggregate. However, the census does not have school level information like quality or enrollment data and hence we turn to a much richer data set provided under the DISE. To review the performance of schools in India and to monitor policies targeted towards them, information on all the registered schools started to be maintained, from 1995, under the software named DISE. For the DISE data, the schools are asked to fill detailed information on a number of school characteristics like infrastructure, enrollment, results, etc. We use the data collected in 2013-14 as it was one of the first attempts to combine the data on elementary schools with secondary and higher secondary schools[8]. It is important to note that even though DISE is the census of all schools in the country, it does not fully cover unrecognised private schools. A school gets recognition upon satisfying certain minimum conditions regarding infrastructure, teacher qualifications, salary etc laid down by the state education boards. Recognition of all private unaided schools became mandatory after the enactment of RTE, however, it is believed that a substantial number of unrecognised schools continue to be in operation suggesting the reported

[7]Which cover about 93.5% of the population of the country. We do not work with the seven north eastern states and the union territories.

[8]Before 2012-13, DISE data collected information on only elementary schools in India.

number of private schools in DISE might be an underestimate of the actual number. If this is indeed true, then the inference of our paper will hold true only for recogniesd private schools.

For constructing our main independent variable, ethnic fragmentation, we consider caste groups as separate ethnic groups. In India, the Hindu population (the major religious group) is divided into a number of castes with deep social cleavages which govern social and economic interaction between these caste groups[9]. Even though caste might differ in principle in many respects from ethnicity, it follows the following basic features about ethnicity that underlie the diversity-debit hypothesis. One, following Kanchan Chandra's definition of ethnic groups, ethnicity implies nominal membership to an ascriptive category like race in which membership is inherited, which is very well followed by the Indian caste system (Chandra (2007)). Two, there are visible socio-economic differences between the ethnic groups. There is ample evidence that disadvantaged castes in India are discriminated against in the labour market (Banerjee and Knight (1985);Madheswaran and Attewell (2007)), have lower representation in the political sphere(Pande (2003)), have worse education and health outcomes (Maitra and Sharma (2009); Thorat and Neuman (2012)). Therefore, it is reasonable to relate the study of caste fragmentation in the Indian context, generally, to a broad literature on the impact of ethnic fragmentation elsewhere in the world.

Like in other contexts, one could argue for looking at religious diversity besides caste fragmentation in India. However, the reason we focus on caste is that religious groups other than the majority Hindu religion make up only 20% of the population according to the 2011 census. Out of this 20%, majority (14%) consists of Muslims suggesting that we'll essentially be looking at the impact of relative shares of Hindu-Muslim population if we base our fractionalisation on religion. While this is an important question in itself, in this exercise, we are interested in heterogeneity rather than identity. One could also make a case for heterogeneity based on language, however since it is common for states in India to have their own official language, it leaves little variation in the language spoken within a state.

[9]Caste is also called *jati* and there are thousands of *jatis* in India. For our purpose, we include a caste in sample, if it constitutes more than 1% of the state population. This leaves us with about 185 caste groups.

Thus, we work with caste fragmentation and make use of census of 1931 to construct a measure of ethnic fragmentation[10]. There are two broad reasons for using historical data to conduct our analysis. One, there would also be severe endogeneity concerns if we used contemporaneous caste shares. Both the mobility of caste groups and provision of public goods is likely to be driven by variables that are difficult to capture like state institutions or policies. It is also possible that the availability of schools and other public goods itself could result in a particular distribution of caste groups. Thus, historical dataset does away with this concern as it is highly unlikely that the caste composition of districts in 1931 would be affected by factors influencing the provision of schools provided much later in time. The other reason is that all the census that were conducted after independence (in 1947) do not report disaggregated caste data and instead have information on population proportion of the three broad social groups, Scheduled Castes (SC), Scheduled Tribes (ST), and the rest of the population. Broadly, all the disadvantaged castes are clubbed together in the census into a constitutional category called SC, the middle and upper castes are clubbed into a category named “others” and the disadvantaged tribes who are placed outside the Hindu caste system are put into a category named ST. These social categories are very broad with many layers of hierarchy within them and therefore these broad categorisation might not be able to capture the actual caste heterogeneity that operates at the ground level. We, therefore, make use of census of 1931 for detailed district-level information on population shares of different castes[11]. However, since 1931, a lot of new districts have been created, so for the districts that got formed after independence, we weight the caste figures from the original district according to the area of the new district which was taken from them, following (Banerjee and Somanathan (2007)). An important assumption we make when working with the census of 1931 is that caste proportions in 1931 have not changed much over time and therefore they are representative of the present day caste diversity. To justify this assumption, we rely on the existing evidence and claim that fragmentation based on

[10]Construction of ethnic fragmentation using historical data has been done earlier too (Banerjee and Somanathan (2007), Banerjee et al. (2005),Anderson (2011), Suryanarayan et al. (2017)).

[11]To get caste level data, some studies have used the Survey of Living Conditions data conducted by the World Bank for two Indian states for 1997 and 1998Lee (2017). However, since this paper studies all major Indian states, we make use of Census of 1931.

1931 proportions seems to be a good proxy.

One way in which our assumption would be violated is if there has been caste based migration. However, evidence suggests that caste based migration in India is quite low probably due to reliance on sub-caste networks of mutual insurance which rarely transgress village boundaries (Munshi and Rosenzweig (2009)). Though women migrate to different villages for the purpose of marriage, it almost always happens within the same caste. In general also, migration in India is low and whatever migration happens, about 62 % of it is intra-district, while 24 % is inter-district and 13 % is inter-state suggesting that migration should not substantially affect caste proportion within districts over time. In order to show that caste proportions have not substantially changed over time, Anderson (2011) matched the caste proportion obtained from the census of 1921 data with the caste proportion obtained from a recent data set that she collected in 1997-98 with the World Bank[12]. The district caste proportions using the two data sets turned to be very similar corroborating the representativeness of the 1931 proportions. Along with the existing evidence, we make use of India Human Development Survey (IHDS) data for the year 2011-12 which has data on present day caste proportions to check the correlation between the fragmentation measure based on IHDS data and the measure based on the 1931 proportions. We find a positive and highly significant correlation between them, further providing evidence that 1931 proportions would serve as a good proxy for present day caste diversity. However, the reason for not using IHDS data for caste level data in our main analysis is that this data is not representative at the district level, therefore the correlation is presented only as a robustness check.

As is quite standard in the literature, we measure ethnic diversity using the Ethnic Fractionalisation Index (EFI), given by $1 - \sum \beta_i^2$ where $\beta_i$ is the population share of the $i^{th}$ ethnic group. However, this measure is now being criticised for not being able to capture socio-economic divisions among the ethnic groups which is what is believed to have negative implications rather diversity per se (Singh (2011);Waring and Bell (2013)). We believe that this criticism should not

[12]Note that this exercise was done for two states, Bihar and Uttar Pradesh and the caste data for the year 1997-98 was collected for the seven castes. So essentially, caste proportions were matched for seven castes in two states over the two data sets.

exactly apply to our context as we consider caste groups which have long been established to have deep social and economic cleavages which are reflected in a number of spheres like labour market outcomes, political participation and social involvement, as mentioned before. Therefore, ethnic fragmentation based on caste should reflect diversity among socio-economically divided groups, which is increasingly getting recognition as a better measure of a fractured society.

Table A.1 in the appendix reports the average difference in school characteristics in high and low fragmented districts, where high (low) corresponds to the above (below) median fragmentation levels. The summary statistics indicate that, on average, highly fragmented districts have a lower number of private schools whereas, quite contrary to findings of the previous literature, the number of public schools is no different from less fragmented districts. In addition, fragmented districts see less number of teaching hours in public schools and less proportion of regular teachers in private schools. For other indicators, the unconditional average quality of schools seem better in fragmented districts.

Using the above mentioned datasets, we estimate the following regression equation to conduct our empirical analysis:

$$School_{ds} = \alpha_s + \beta_1 cfrag_{ds} + X^{'}_{ds}\delta + \epsilon_{ds} \tag{1}$$

where $d$ indexes the district and $s$ indexes the state. Dependent variable in the above regression equation indicates the number of private or public schools per thousand population. The main parameter of interest is $\beta_1$, the coefficient of $cfrag_{ds}$ which is the ethnic fractionalisation index. The existing literature, which has not differentiated between different types of public goods while estimating equation 1, has documented $\beta_1$ to be negative. The argument is that a failure of collective action among the local community is the channel responsible for lower provision of public services. Since we believe that local collective action is important in varying degrees for private and public schools, we make a distinction between these types of schools and expect the impact of ethnic fragmentation to be different for private and public schools. Private and local government schools which depend on local community for mobilising funds are expected to be negatively impacted by fragmentation whereas schools provided by higher levels of government where community action might not play an important role are not expected to be lower in number.

We include state dummies as policies and guidelines for the provision of schools could differ across states. $X_{ds}$ indicates district level controls, namely urbanisation rate, work participation rate and number of colleges in a district. Economic intuition for adding urbanisation rate is that the pattern and the rate at which a district gets urbanised might have some correlation with both the degree of fractionalisation (for e.g. more rural districts might be more fragmented) and the number and quality of schools (urban areas might see more private schools while rural more public). Hence, we do not want fragmentation to pick up the effect of urbanisation[13]. We add number of colleges and work participation rate in the regression because if a district, on an average, has a higher work participation rate and number of colleges, this might indicate better labour outcomes and in general higher preference for education in that district. Thus, these variables, to some extent, capture the demand for education in a district. Since our motive is to study the impact of ethnic fragmentation on the supply of schools, we control these variables in the regression to partially isolate the impact of ethnic fragmentation on the supply of schools from their demand. Data on all the district level controls come from the Indian population Census of 2011.

# 5 Results

The impact of ethnic fragmentation on the number of private and public schools obtained using the census data is reported in Table 2. The first panel reports the results without the control variables while the second one has all the controls. Consistent with what we expect, ethnic fragmentation seems to have a weak negative impact on *only private* elementary and secondary schools,[14] with very similar coefficients across two specifications. Coefficient of elementary private school indicates that if we were to move from a completely homogeneous district to a perfectly fragmented district, we would see a fall in private elementary schools (per '000 population) by 0.19.

[13] Apart from the urbanisation rate, we also performed our analysis with a dummy which indicates if the district has more than 50% of its population living in urban areas. All our results remain same with this control variable as well,

[14] Elementary schools have classes up to standard eight, in other words, they provide eight years of schooling. Students begin elementary school around the age of six and graduate around fourteen.

This amounts to a 68% decline in the number of private schools when compared with the average number of private elementary schools in our sample (0.276). Clearly, ethnic fragmentation has a detrimental impact on the provision of private schools. However, there is no evidence of a negative impact on public schools, suggesting a weak differential impact of ethnic fragmentation on different types of schools.

We now turn to a much richer school-level database, DISE, to perform the rest of our empirical analysis. We begin by re-estimating equation 1 to see if a similar differential effect exists with the DISE data too. Our finding with the census data gets substantiated, as reported in Table 3. Ethnic fragmentation lowers the number of private elementary and private secondary schools without having any impact on public schools. This finding of no impact on public schools, at first, might seem inconsistent with the literature that has shown negative linkages between provision of education and ethnic diversity. However, since the existing literature has not made any distinction between public and private provision of education, it could very well be possible that the reported negative impact in the literature was driven by lower private provision. In any case, we think that this result is reasonable not only in the Indian setting but also elsewhere where provision of most of the public schools is the responsibility of either the state or the central government[15]. Unlike private players or local councils, the state does not rely on the ability of communities to act together to raise funds for schools, a mechanism commonly suggested for the negative association between public provision of schools and fragmentation. In addition, governments are bound by equity considerations to provide public education for all. For example, many policies and programs in India (Sarva Shiksha Abhiyan, Right to Education act) aim at universalisation of elementary education in the country, which probably leaves little variation in the number of public schools across districts.

In addition to testing our hypothesis for private and public schools, we now study the impact of ethnic fragmentation on a finer classification of schools. Public schools in India are normally provided by three different levels of government; local, state and national suggesting that we can think of public schools as belonging to three distinct categories. Likewise, private schools can be

[15]Barring the ones provided by the local government which are very few in number. In India, local government schools constituted only 3% of total schools in 2011-12

divided into two distinct categories, those that receive aid for their operation from the government and those that are unaided private schools. The impact on this sub-classification of schools is reported in Table 4. Panel A reports the results for total number of these schools while panel B looks at the number at the elementary level. We look at only elementary level because at secondary or higher secondary levels majority of public schools are run by the state government and majority of private schools are unaided leaving little variation to exploit at higher levels. Results indicate that while local government and unaided schools are negatively affected by ethnic fragmentation, there is no effect for either aided schools or schools that are run by state/central government. This elucidates that even within seemingly homogeneous categories such as public and private schools, the impact of fragmentation is not same for different categories. These results also suggest that only those schools which depend on participation of community members for mobilisation of funds are negatively impacted which is an additional evidence in favour of our hypothesis.

## 5.1 Impact on quality of schools

Education in India is characterised by a large number of poor quality schools, lack of basic amenities, poor quality teachers and below average student test scores. Given this dismal state, discussion about provision of schools in fragmented areas would only partially be complete if we do not analyse the quality of the existing schools. This is the reason why apart from looking at the number of schools, we want to understand the impact of ethnic fragmentation on school quality[16]. We look at hiring of teachers and other teacher-quality variables to measure school quality. The reason for explicitly focusing on teacher quality is that earlier literature has suggested a significant impact of community involvement in school management on teacher performance(Pandey, Goyal and Sundararaman, 2009; Duflo, Dupas and Kremer, 2015). Since, we argue that there is lack of collective action in fragmented areas, we should see a direct impact on teacher variables if we expect school

[16]Note that in this exercise we are going to be looking at the truncated distribution of private schools in fragmented districts.

quality to be negatively impacted. We estimate the following equation to perform our test:

$$School_{ids} = \gamma_s + \beta_2 cfrag_{ds} + X^{'}_{ids}\delta_1 + X^{'}_{ds}\delta_2 + \epsilon_{ids} \tag{2}$$

where $i$ indexes the school, $d$ indexes the district, and $s$ indexes the state. The DISE data has school-level information on a number of variables like the number of teachers, classrooms and enrollment and so we estimate the relationship between caste fragmentation and aspects of school quality at the school level. As in equation 1, we include state dummies and district level controls here. In addition, we have a number of school-level controls ($X_{ids}$), namely, age of the school, $Urban$, a dummy indicating if the school is located in urban areas and, $Roadaccess$, a dummy indicating if the school is approachable by all weather roads. Data on these controls have been taken from the DISE data. In these tests, we allow the possibility that shocks to school provision and quality could be correlated within a district by clustering the standard errors at the district level.

The coefficient $\beta_2$, gives us the impact of fragmentation on public and private school quality. Public schools which even though are funded by state/national governments, rely on community participation for maintenance of schools and building of school development plans. Infact, a number of legislations have now mandated public schools to involve the local community in school maintenance, we, therefore, expect $\beta_2$ to be negative for public schools. On the other hand, even though private schools are also managed by the local community, it is difficult to apriori say anything about their quality. The reason is that existing private schools in fragmented places are a result of successful collective action resulting in mobilisation of funds. This successful collective action, which is responsible for the provision of schools, could also lead to better teacher-quality in fragmented places.

Results, reported in Table 5, suggest that in line with what we expect, public schools have worse quality in fragmented districts. They have a lower proportion of regular teachers and lower number of teachers per classrooms in fragmented districts. Teaching hours put in by public school teachers are also lower (even though the coefficient is negative, it is not significant). Additionally, there are more number of single teacher public schools in fragmented districts as compared to homogeneous districts, again indicating lower quality. In the last two columns, we use the four quality variables to create an index of school quality (for both public and private schools) using Principal Component Analysis. The coefficient of the public school quality index is negative, consistent with previous columns. On the other hand, we find that school quality in the private schools is not inferior as compared to homogeneous districts. However, we believe that these coefficients are estimated with a bias since we observe a truncated distribution of private schools in fragmented districts. The existing private schools are a result of successful collective action and so, these results are likely to be contaminated with selection bias. However, in the absence of information on school characteristics on missing private schools, we cannot do much to correct this bias.

# 6 Mechanism

We now present formal tests to support our argument that it is the lack of collective action in diverse areas that lowers the number of private schools and negatively impacts the quality of public schools. Lack of collective action affects the two types of schools differentially because we hypothesise that collective action is important at different stages in the provision of these schools. Public schools in India are funded by the government and the provision decision is guided by equity considerations and therefore local demographic structure might not play a role in their provision. However, public schools have now become increasingly dependent on local community to maintain schools. Various state governments like Madhya Pradesh, Karnataka and Uttar Pradesh have decentralised public schools monitoring by mandating the constitution of Parent-Teacher Association, School Development and Monitoring Committee and Village Education Committee, respectively. More recently and importantly, the enactment of the RTE act of 2009 and RMSA has mandated the constitution of School Management Committee (SMC), School Management Development Committee (SMDC) and Parent-Teacher Association (PTA) in all public and aided schools[17]. These are local community associations which are primarily responsible for monitoring school funds, teacher performance, student learning and school infrastructure, thus representing an institutional mechanism to encourage community participation in public school maintenance. This suggests that while failure of collective action among community might not impact the number of public schools, teacher performance and other parameters of school quality might be affected.

We provide evidence for the lack of collective action for the maintenance of public schools by looking at the impact of ethnic fragmentation on SMC, SMDC and PTA meetings. Results are reported in Table 6. Negative coefficients in columns 2 and 3 suggest that while the frequency of SMC meetings is not low, lower number of SMDC and PTA meetings are held in fragmented districts. In columns 4 to 7, we look at the extent of community participation in SMDC meetings by exploiting additional information on these meetings present in the data. We find that participation by local body members, parents and members of SC/ST groups in SMDC meetings is also low. Inactive participation of community members in these platforms is an indication of lack of teamwork and coordination amongst the local community to maintain school and build school plans.

We now consider the impact of lower frequency of these meetings on school quality in order to complete our argument. Panel A of Table 7 looks at the impact of PTA meetings and panel B looks at SMDC meetings[18]. The reported results show that while SMDC meetings seem to not have any impact on school quality, the results for PTA meetings are positive. This might be pointing towards the fact that community meetings which have wide representation from the stakeholder group (parents in this case) is most effective for public goods maintenance. The results of panel A suggest that more frequently the PTA meets, higher is the proportion of regular teachers in a school, more is the number of hours put in by teachers, there are more number of teachers per classrooms and there is less proportion of single teacher schools. These results which are based on a secondary data source are very well in line with earlier studies (Pandey, Goyal and Sundararaman, 2009; Duflo, Dupas and Kremer, 2015) which have relied on randomised control trials to show that parental

[17]SMC represent community participation platform for elementary schools while SMDC is meant for secondary and higher secondary schools. If the school has classes above elementary level, then it is suggested that SMC be the primary community participation platform.

[18]Since there is no reported impact of fragmentation on SMC meetings, we don't look at them in this table.

and community involvement in the maintenance of schools improves teacher performance and to some extent student learning outcomes, bolstering our claim. The results of this table which show that lack of community monitoring in fragmented places is likely to negatively impact the quality of public schools has important implications in the Indian context as community participation is now being seriously considered as an important way to improve the quality of schools. In addition to the mandate of RTE and RMSA, there are many other initiatives of the government to involve community like Lok Jumbish in Rajasthan, and the SDMCs of Karnataka. However, our results show that such decentralisation initiatives in education sector might not work in a fragmented society where social network is weak.

On the other hand, private schools are legally mandated to be set up by opening charitable trusts or societies. This requirement seems to have some bite in practice as a survey of schools conducted in Hyderabad finds around 86% of the private schools to be run by charitable trust Tooley et al. (2007). This might indicate that private schools are likely depend on voluntary contributions from the local community members to raise money to start/run a school atleast in the short run. So, our conjecture here is that collective action in fragmented districts is not strong enough due to which the mechanism of voluntary payments by the community members to raise private schools does not work well. We test the first part of the above-made assertion by looking at the impact of caste fragmentation on the number of self-help groups and agricultural credit societies[19]. Self-help Groups (SHGs) and agricultural credit societies represent one of the many ways by which community members voluntarily pool a part of their savings which is then used in the times of distress for mutual help. Even though, this is not the best way to test the mechanism of voluntary payments to raise schools in fragmented places, it gives us as an indication of the ability of the community to form groups in which members are connected to each other with the purpose of raising money. Results are reported in Table 8 and the two dependent variables indicate the proportion of villages in a district which have SHGs and Agricultural societies, respectively. Negative coefficients for

[19]We obtain the data on these two variables from the village directories of 2011 Census. These variables are reported as indicator variables, taking a value one if a village has a self-help group/agricultural credit society. We aggregate these variables upto the district level to conduct our analysis.

both the dependent variable suggests that community in fragmented districts has less tendency to form groups that can be used for mutual help. We argue that this is somewhat an indication of weak collective action that is required to raise funding for schools.

We also look at the correlation between our measure of collective action to raise funds and number of private and public schools in Table 9. If the two groups considered above truly capture collective action which can facilitate the mobiliation of funds, there should be some correlation between their presence and private schools and no correlation with public schools which doesn't require local resource raising. This is exactly what we find in the table which suggests that homogeneous districts with well functioning mechanism of voluntary payments, as indicated by higher number of self help groups and agricultural societies, see more number of private schools.

### 6.1 Alternative explanation

We show in the paper that it is the failure of collective action that leads to a negative differential impact on the provision of schools. However, in addition to the existing explanation, there could be other channels responsible for lower number of private schools and poor quality public schools in fragmented areas. We, therefore, test a number of potential explanations in Tables 10 (which looks at the number of schools) and 11 (which looks at the quality of schools). The first alternative channel could simply be low economic activity in fragmented places, given that there is evidence that ethnic fragmentation negatively impacts local economic growth (Easterly and Levine (1997)). Low local economic growth by suppressing the demand for fee-charging private schools and/or by lowering the capacity of the community to provide schools could lower the provision of private schools in fragmented areas. Additionally, it could also imply lower quality public schools if the community has to provide some monetary contributions to maintain public schools. Results, reported in the first sections of Tables 10 and 11, however, show that there is no change in our main results even after controlling for disrict GDP, indicating that the observed effect of ethnic fragmentation is not driven by low economic activity in fragmented places. Data for district GDP is obtained from the Open Government Database (OGD) of the Indian government. The reason for not controlling district GDP in our main test is that the data on district GDP for the year 2010-11 is available only for 12 out of the 18 states we consider. For the remaining six states, we use the latest available district GDP data, which leads to some inconsistency in the timing of district GDP data. For this reason, we consider other indicators of economic activity like urbanisation, work participation rate, both of which are highly correlated with district GDP, in our main tests.

In addition to average district domestic product being affected by the local ethnic diversity, there is evidence that ethnic diversity widens the distribution of consumption/income (Chadha and Nandwani (2018)). An unequal distribution of income could result in different preferences over the distribution of schools which might be responsible for low provision of private schools and poor quality public schools in diverse areas. Therefore, we control Gini coefficient of real per capita consumption expenditure obtained from the National Sample Survey round conducted in the year 2011-12. This allows us to capture the effect of diversity on schools over and above the effect which could be driven by an unequal distribution. Second part of Tables 10 and 11, shows that districts with higher inequality have low provision of public schools and better quality public and private schools. Importantly, results also show that inequality does not alter the reported relationship between fragmentation and provision of schools ruling out inequality as the potential channel.

As mentioned before, discrimination along the caste lines is widely prevalent in India which has created socio-economic divides between caste and social groups. If these group-based disparities are more pronounced in fragmented areas, then lower provision of private schools and poor quality public schools can simply be an outcome of the fact that economically weaker groups have a lower demand for education. To see if group disparities are driving the result, we control economic inequality between social groups in the third part of Tables 10 and 11 by adding Group-Gini coefficient constructed using the NSS data of year 2011-12. However, we find that the coefficients for private schools continue to be negative and significant and so is the coefficient of public school quality, ruling out group based inequalities to be the potential channel.

In the past three and a half decades or so, politics in India has become quite competitive and it is believed to be even more so in socially diverse regions which have seen the emergence of a number of caste based parties. Along with this apparent correlation with caste diversity, the nature of political competition ( in particular dominance of multi or two parties) has also been shown to matter for the provision of public goods (Chhibber and Nooruddin (2004)). Therefore, to make sure our results are robust to political make up in a district, we add political fragmentation in the election year 2009 constructed using vote shares of different political parties in a district[20]. The results, reported in the fourth section of the Tables 10 and 11, show that ethnic fragmentation continues to exert a negative influence on the provision of private schools and quality of public schools even when political competition is added. We also employ other measures of political environment, namely, vote share of the winner party, vote share of national and state parties and the proportion of voters per population in a district. However, our result remains robust to all the measures of political environment (results, not reported, can be obtained from the authors on request).

The effect of fragmentation can also be confounded if heterogeneous districts have a high proportion of disadvantaged castes who either do not have the resources to provide schools or have low preference for education. To check this, we add the population proportion of disadvantaged groups, Scheduled Castes (SC) and Scheduled Tribes (ST) in our regression specification reported in the fifth section of Tables 10 and 11. This will also check the robustness of our results to a potential explanation, suggested by Lee (2017), that central government might discriminate against areas inhabited by less powerful groups. However, the results don't change even after addition of disadvantaged groups, confirming that results of the paper are driven due to failure of collective action in heterogeneous societies rather than discrimination or low group preference for education.

Additionally, we test if members of different caste groups do not want to study together in the same school and if this is true, this can end up in segregation of schools by caste. In that situation, it might not be viable to provide schools (specifically by private players) for students belonging to

[20]The unit at which parliamentary elections happen, parliamentary constituencies, are different from districts, which is the unit of our analysis. We match constituencies to districts using delimitation commission reports of the Election Commission of India. Authors can be contacted for more details on the procedure of matching.

minority caste, a reason possible for driving low provision of schools rather than the mechanisms that we consider. This mechanism has also been suggested by Alesina et al. (1999) who point that neighbourhood segregation by ethnicity can negatively impact the provision of public goods. To test if this is the case, we create enrollment fragmentation (that is how heterogeneous is the enrollment by social groups in a given class) for each class by using information on enrollment from SC, ST, OBC and general category using DISE data. Since we consider very broad categories to construct our fragmentation measure of the school, our ethnic fragmentation measure for this test is also created using the broad social groups (SC, ST, others) reported in 2011 Census. The results reported in Table 14 show that districts which are fragmented seem to have a diverse class composition, with no negative differential impact for private schools. The results are only reported for classes 1, 5, 10 and 12 for brevity. This suggests that for the existing schools fragmentation does not lead to segregation in enrollment and therefore the impact of ethnic fragmentation does not seem to be driven by segregation in enrollment.

Apart from these formal tests, we also make sure that any particular state in our sample is not driving the results. We sequentially drop a state at a time from our sample but the results, however, do not indicate sensitivity to any particular state. Thus, overall, we rule out all the possible alternate explanations, that we can think of, which could be driving the reported results.

# 7 Robustness check: spatial correlation

The inference drawn in the paper which suggests differential impact of fragmentation on private and public schools relies on the assumption of independent and identically distributed error terms. This implies that we assume that non-diagonal terms on the variance-covariance matrix of the error terms are zero. However, there is a concern that unobservable factors (which have not been accounted for in the regressions) determining the provision of schools in districts are correlated across neighbouring districts leading to a spatially clustered distribution of schools. This concern arises from the possibility that for public schools it could be administratively easier for government (state/central) to provide them in nearby districts at a given point and private schools could respond to economic environment and competition in bordering districts. This implies that for districts which are closer in proximity, we would expect a high correlation between the error terms and this correlation would fade as the distance between the districts grow. This would mean that atleast some of the non-diagonal elements in the variance matrix would be non zero, thus causing us to potentially draw incorrect inference.

In this section, we test if there is spatial correlation in the data and also make sure that our results are robust to that. The test statistic that is widely used to detect spatial correlation is Moran's I statistic computed as $I = \frac{\sum_{i,j=1}^{n} w_{ij} r_i r_j}{\sum_{i=1}^{n} r_i^2} . \frac{n}{\sum_{i,j=1}^{n} w_{ij}}$, where $r_i$ is the residual $(y_i - \overline{y})$ in unit i, $w_{i,j}$ is an element of $nXn$ matrix of spatial proximity, W. By convention, $w_{i,i}$=0, and $w_{i,j}$ is positive and decreases as i and j become distant. To compute matrix W, we need to know the placement of units in the euclidean space so that distance between them can be computed. In the context of this paper, we use the data on latitudes and longitudes for the districts to compute the matrix W. As is clear from the formula, Moran's I statistic indicates the covariance of the residuals with themselves taken at the neighboring locations normalized by the population variance, which is exactly what we seek to test spatial correlation.

The value of the Moran's I statistic for private and public elementary school turns out to be 0.23 and 0.08, respectively. For both the types of schools, the statistic is highly significant (p-value less than 0.01), indicating that schools are spatially clustered, with neighbouring districts (clusters) having a high density of schools while others with a low density. This evidence of spatial correlation casts suspicion on the standard errors and consequently the inference drawn. To make sure our results are robust to spatial correlation, we compute Conley standard errors, following Conley (1999) which account for correlation in the unobservables among the neighbouring units.

The results are reported in Table 13. Conley standard errors which allow for both heteroskedasticity and spatial correlation are reported in column 3 and for comparison we have also reported errors that have been computed using IID assumption (column 1) and heteroskedasticity (column 2). Since parameters in the paper are estimated using OLS, the impact of allowing for spatial dependence is only on the errors and not on the point estimate. As we can see there does not seem to be any difference between standard errors in columns 1 and 2 indicating that errors do not seem to be heteroskedastic. However, when we compare columns 1 and 3, we can see a considerable difference between the IID standard errors and errors corrected for spatial correlation. The spatially corrected standard errors are much smaller in magnitude as compared to the scenario of independent errors. However, even after this correction, the qualitative result of the paper does not change. The inference that fragmentation exerts a negative impact on private schools and no impact on public schools can be made even more strongly after correcting for spatial correlation.

# 8 Conclusion

In this paper, we link an important aspect of social-demographic structure, ethnic divisions in the society, with a critical component of human capital accumulation, provision of schools in India. In general, there is an extensive literature which has established that ethnic diversity exerts a negative influence on the provision of public goods. We add to this established relationship by showing that while diversity indeed has a negative influence on the provision of schools, the impact falls only on those types of schools which rely on collective action by community for their provision. This paper, thus, shows that the established negative influence of diversity should not be taken as it is, as it critically depends on the mechanism by which public goods are provided. This has an important implication which suggests that diverse places can improve access to public goods and overcome the diversity curse by focusing on the type of public service which is impacted least by diversity. Our empirical analysis focuses on the provision of public and private schools in Indian districts and we show that the two types of schools are impacted differently by diversity. Private schools are lower in number but are not impacted in terms of their quality in fragmented places. On the other hand, state and central schools which do not rely on local caste composition for their provision seem to be not affected by caste diversity, at least on extensive margin. We show that the reason schools are affected in diverse areas is that ability to act collectively is low in fragmented districts. Since the two types of schools are provided with very different objectives, low collective action results in a differential effect for the two schools. The paper provides evidence that public schools are of lower quality in diverse districts in India and that private schools also have not been able to take up this space created by lower quality public schools. This clearly shows that among a range of social and economic problems that diverse places might face one of them is lower number of private and lower quality public schools. Given that schools represent an important component of human capital accumulation, this finding has implications for performance of a diverse country like India in terms of achieving desirable future education outcomes. The results also indicate that since not all mechanisms of provision of schools are impacted by diversity, thus, further research needs to be done to come up with the most efficient way to provide quality education facilities in diverse places.

# References

**Alesina, A., R. Baqir, and W. Easterly**, “Public Goods and Ethnic Divisions,” *The Quarterly Journal of Economics*, 1999, *114* (4), 1243–1284.

**Alesina, Alberto and Eliana La Ferrara**, “Ethnic Diversity Economic Performance,” *Journal of economic literature*, 2005, *43* (3), 762–800.

**Anderson, Siwan**, “Caste as an Impediment to Trade,” *American Economic Journal: Applied Economics*, 2011, *3* (1), 239–263.

**Banerjee, Abhijit and Rohini Somanathan**, “The political economy of public goods: Some evidence from India,” *Journal of development Economics*, 2007, *82* (2), 287–314.

**_ , Lakshmi Iyer, and Rohini Somanathan**, “History, social divisions, and public goods in rural India,” *Journal of the European Economic Association*, 2005, *3* (2-3), 639–647.

**Banerjee, Biswajit and John B Knight**, “Caste Discrimination in the Indian Urban Labour Market,” *Journal of Development Economics*, 1985, *17* (3), 277–307.

**Barro, Robert J**, “Human capital and growth,” *The American Economic Review*, 2001, *91* (2), 12–17.

**Chadha, Nishant and Bharti Nandwani**, “Ethnic fragmentation, public good provision and inequality in India, 1988–2012,” *Oxford Development Studies*, 2018, pp. 1–15.

**Chandra, Kanchan**, *Why ethnic parties succeed: Patronage and ethnic head counts in India*, Cambridge University Press, 2007.

**Chaudhary, Latika**, “Determinants of Primary Schooling in British India,” *The Journal of Economic History*, 2009, *69* (01), 269.

**Chhibber, Pradeep and Irfan Nooruddin**, “Do party systems count? The number of parties and government performance in the Indian states,” *Comparative Political Studies*, 2004, *37* (2), 152–187.

**Conley, Timothy G**, “GMM estimation with cross sectional dependence,” *Journal of econometrics*, 1999, *92* (1), 1–45.

**Duflo, Esther, Pascaline Dupas, and Michael Kremer**, “School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools,” *Journal of Public Economics*, 2015, *123*, 92–110.

**Easterly, William and Ross Levine**, “Africa’s growth tragedy: policies and ethnic divisions,” *The Quarterly Journal of Economics*, 1997, *112* (4), 1203–1250.

**Gisselquist, Rachel M, Stefan Leiderer, and Miguel Niño-Zarazúa**, “Ethnic Heterogeneity and Public Goods Provision in Zambia: Evidence of a Subnational Diversity Dividend,” *World Development*, 2016, *78*, 308–323.

**Kingdon, Geeta Gandhi**, “The private schooling phenomenon in India: A review,” 2017.

**Lee, Alexander**, “Ethnic Diversity and Ethnic Discrimination: Explaining Local Public Goods Provision,” 2017.

**Madheswaran, S and Paul Attewell**, "Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey," *Economic and political Weekly*, 2007, pp. 4146–4153.

**Maitra, Pushkar and Anurag Sharma**, "Parents and children: Education across generations in India," in "5th annual conference on economic growth and development, ISI Delhi, Delhi" 2009.

**Miguel, Edward and Mary Kay Gugerty**, "Ethnic diversity, social sanctions, and public goods in Kenya," *Journal of public Economics*, 2005, *89* (11), 2325–2368.

**Munshi, Kaivan and Mark Rosenzweig**, "Why is mobility in India so low? Social insurance, inequality, and growth," Technical Report, National Bureau of Economic Research 2009.

**Pande, Rohini**, "Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India," *The American Economic Review*, 2003, *93* (4), 1132–1151.

**Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman**, "Community participation in public schools: impact of information campaigns in three Indian states," *Education economics*, 2009, *17* (3), 355–375.

**Romer, Paul M**, "Human capital and growth: theory and evidence," in "Carnegie-Rochester conference series on public policy," Vol. 32 Elsevier 1990, pp. 251–286.

**Schündeln, Matthias**, "Ethnic heterogeneity and the private provision of public goods," *The Journal of Development Studies*, 2013, *49* (1), 36–55.

**Singh, Prerna**, "We-ness and welfare: A longitudinal analysis of social development in Kerala, India," *World Development*, 2011, *39* (2), 282–293.

**Suryanarayan, Pavithra et al.**, "When do the poor vote for the right-wing and why: Status inequality and vote choice in the Indian states," Technical Report, World Institute for Development Economic Research (UNU-WIDER) 2017.

**Thorat, Sukhadeo and Katherine S Neuman**, *Blocked by caste: economic discrimination in modern India*, Oxford University Press, 2012.

**Tooley, James, Pauline Dixon, and SV Gomathi**, "Private schools and the millennium development goal of universal primary education: A census and comparative survey in Hyderabad, India," *Oxford Review of Education*, 2007, *33* (5), 539–560.

**Waring, Timothy M and Adrian V Bell**, "Ethnic dominance damages cooperation more than ethnic diversity: results from multi-ethnic field experiments in India," *Evolution and Human Behavior*, 2013, *34* (6), 398–404.

Table 1: Distribution of schools in various consumption quartiles

| Cons quart | Public | Private | Total |
|---|---|---|---|
| 1 | 20.91 | 4.76 | 25.67 |
| 2 | 18.34 | 7.24 | 25.58 |
| 3 | 14.33 | 10.72 | 25.05 |
| 4 | 8.3 | 15.4 | 23.7 |
| Total | 61.88 | 38.12 | 100 |

*Notes:* This table has been constructed using the 68th round of National Sample Survey (NSS) conducted in the year 2011-12. *Cons quart* indicates the four quartiles of the monthly per capita consumption expenditure distribution. *Public* indicates that the school is public and *Private* denotes that school is private. Each cell $i, j$ (where $i$=1,..,4 and $j$=1,2) indicates the percantage of children belonging to consumption quartile $i$ and studying in school $j$.

Table 2: Effect of ethnic fragmentation on school provision (census data)

Panel A

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | G_element | P_element | G_second | P_second | G_highsecond | P_highsecond |
| Ethnic frag | 0.0948 | -0.1951$^{*}$ | -0.1347$^{+}$ | -0.0533$^{+}$ | -0.0337 | -0.0221 |
| | (0.875) | (0.083) | (0.140) | (0.102) | (0.539) | (0.283) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | No | No | No | No | No |

Panel B

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | G_element | P_element | G_second | P_second | G_highsecond | P_highsecond |
| Ethnic frag | 0.3330 | -0.1922$^{*}$ | -0.1224 | -0.0548$^{*}$ | -0.0274 | -0.0254 |
| | (0.537) | (0.086) | (0.169) | (0.090) | (0.613) | (0.195) |
| urbanisation | -1.0617$^{***}$ | 0.0570$^{+}$ | -0.0952$^{***}$ | 0.0273$^{**}$ | -0.0325$^{*}$ | 0.0262$^{***}$ |
| | (0.000) | (0.141) | (0.002) | (0.015) | (0.085) | (0.000) |
| workrate | 3.2785$^{***}$ | -0.1090 | 0.2183$^{**}$ | -0.0275 | 0.1746$^{***}$ | -0.0599$^{**}$ |
| | (0.000) | (0.425) | (0.045) | (0.486) | (0.009) | (0.013) |
| college_dis | -0.0000 | 0.0000 | -0.0000 | 0.0000$^{+}$ | -0.0000 | 0.0000 |
| | (0.792) | (0.511) | (0.212) | (0.140) | (0.505) | (0.286) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 405 | 405 | 405 | 405 | 405 | 405 |

*p*-values in parentheses

*Notes: G_element, G_second and G_highsecond* indicate the number of government elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. *P_element, P_second and P_highsecond* indicate the number of private elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. These variables are constructed at the district level using 2011 census data from village and town directories. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. Results presented in the first panel of the table does not any use any control variables whereas second panel has all the controls. *urbanisation* is the proportion of population living in urban areas, *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three control variables have been constructed using census data.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 3: Effect of ethnic fragmentation on school provision (DISE data)

Panel A

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | G_element | P_element | G_second | P_second | G_highsecond | P_highsecond |
| Ethnic frag | -0.0695 | -0.1265$^{**}$ | -0.0190 | -0.0485$^{**}$ | -0.0011 | -0.0011 |
| | (0.794) | (0.011) | (0.596) | (0.034) | (0.964) | (0.936) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | No | No | No | No | No |

Panel B

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | G_element | P_element | G_second | P_second | G_highsecond | P_highsecond |
| Ethnic frag | 0.0122 | -0.1586$^{***}$ | -0.0075 | -0.0572$^{**}$ | 0.0054 | -0.0072 |
| | (0.938) | (0.001) | (0.777) | (0.011) | (0.765) | (0.596) |
| urbanisation | -0.4299$^{***}$ | 0.0712$^{***}$ | -0.0202$^{**}$ | 0.0254$^{***}$ | -0.0025 | 0.0204$^{***}$ |
| | (0.000) | (0.000) | (0.031) | (0.001) | (0.698) | (0.000) |
| college_dis | 0.0675$^{***}$ | -0.0007 | 0.0104$^{***}$ | -0.0001 | 0.0071$^{***}$ | 0.0010$^{***}$ |
| | (0.000) | (0.532) | (0.000) | (0.821) | (0.000) | (0.009) |
| workrate | 1.1305$^{***}$ | -0.1202$^{**}$ | 0.0629$^{*}$ | -0.0805$^{***}$ | 0.0368$^{+}$ | -0.0530$^{***}$ |
| | (0.000) | (0.041) | (0.056) | (0.004) | (0.121) | (0.003) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 396 | 394 | 396 | 394 | 384 | 380 |

*p*-values in parentheses

*Notes:* G_element, G_second and G_highsecond indicate the number of government elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. P_element, P_second and P_highsecond indicate the number of private elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. These variables are constructed at the district level using the DISE data. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. Results presented in the first panel of the table does not any use any control variables whereas second panel has all the controls. *urbanisation* is the proportion of population living in urban areas, *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three control variables have been constructed using census data.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 4: Effect of ethnic fragmentation on school provision (DISE data)

| | Panel A | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Local | G | Aided | Unaided |
| Ethnic frag | -0.1153$^{*}$ | -0.3273 | -0.0115 | -0.1915$^{***}$ |
| | (0.101) | (0.125) | (0.702) | (0.004) |
| urbanisation | -0.1737$^{***}$ | -0.1999$^{***}$ | -0.0175$^{*}$ | 0.1337$^{***}$ |
| | (0.000) | (0.008) | (0.098) | (0.000) |
| college_dis | 0.0025$^{+}$ | 0.0770$^{***}$ | 0.0001 | -0.0008 |
| | (0.112) | (0.000) | (0.869) | (0.604) |
| workrate | 0.0472 | 0.2574 | -0.0768$^{**}$ | -0.1130 |
| | (0.588) | (0.330) | (0.040) | (0.165) |
| State FE | Yes | Yes | Yes | Yes |

| | Panel B | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Local_elementary | G_elementary | Aided_elementary | Unaided_elementary |
| Ethnic frag | -0.1121$^{+}$ | -0.2039 | -0.0004 | -0.1583$^{***}$ |
| | (0.117) | (0.251) | (0.986) | (0.000) |
| urbanisation | -0.1739$^{***}$ | -0.1933$^{***}$ | -0.0097 | 0.0810$^{***}$ |
| | (0.000) | (0.002) | (0.181) | (0.000) |
| college_dis | 0.0028$^{*}$ | 0.0661$^{***}$ | -0.0003 | -0.0004 |
| | (0.094) | (0.000) | (0.558) | (0.679) |
| workrate | 0.0459 | 0.1911 | -0.0713$^{***}$ | -0.0484 |
| | (0.604) | (0.385) | (0.006) | (0.358) |
| State FE | Yes | Yes | Yes | Yes |
| Observations | 396 | 396 | 396 | 396 |

*p*-values in parentheses

*Notes: Local* in panel A indicates the total number of schools provided by the local bodies, *ST* denotes the schools provided by national/state government, *Aided* is the total number of private aided schools and *Unaided* is the total number of private unaided schools. Dependent variables in Panel B measure the number of these schools at the elementary level. These variables are constructed at the district level using the DISE data. Independent variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. *urbanisation* is the proportion of population living in urban areas, *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three control variables have been constructed using census data.

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 5: Impact of fragmentation on teacher quality

| | (1) P_Regular teachers | (2) G_Regular teachers | (3) P_Teacher hrs | (4) G_Teacher hrs | (5) P_teachers/classes | (6) G_teachers/classes | (7) P_Single teach | (8) G_Single teach | (9) P_index | (10) G_index |
|---|---|---|---|---|---|---|---|---|---|---|
| Ethnic frag | 0.006 | -0.179*** | 0.131 | -1.354 | 0.272 | -0.177* | -0.009 | 0.189** | 0.057 | -0.953*** |
| | (0.942) | (0.000) | (0.902) | (0.259) | (0.273) | (0.052) | (0.847) | (0.028) | (0.836) | (0.001) |
| urban | 0.005+ | 0.028*** | 0.152** | 0.317*** | 0.159*** | 0.247*** | -0.012 | -0.062*** | 0.055** | 0.221*** |
| | (0.144) | (0.000) | (0.043) | (0.000) | (0.000) | (0.000) | (0.233) | (0.000) | (0.031) | (0.000) |
| Road access | -0.006 | 0.026*** | 0.307*** | 0.652*** | 0.172*** | 0.086*** | -0.040*** | -0.075*** | 0.107*** | 0.296*** |
| | (0.194) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| age | 0.001*** | 0.000 | 0.005* | 0.009*** | 0.003*** | 0.004*** | 0.000 | -0.002*** | 0.003*** | 0.005*** |
| | (0.000) | (0.161) | (0.076) | (0.000) | (0.000) | (0.000) | (0.888) | (0.000) | (0.001) | (0.000) |
| college_dis | 0.002 | -0.001 | 0.006 | -0.031 | 0.007 | -0.014*** | -0.000 | 0.008*** | 0.005 | -0.020*** |
| | (0.451) | (0.552) | (0.779) | (0.188) | (0.241) | (0.000) | (0.893) | (0.001) | (0.444) | (0.005) |
| workrate | -0.044 | -0.114** | -3.287*** | -5.741*** | -0.139 | -0.397*** | -0.013 | 0.180** | -0.567* | -1.422*** |
| | (0.614) | (0.047) | (0.003) | (0.000) | (0.676) | (0.000) | (0.883) | (0.037) | (0.072) | (0.000) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 123961 | 380755 | 125223 | 385719 | 125299 | 385731 | 125314 | 385752 | 123907 | 380729 |

*p*-values in parentheses

*Notes: P(G)_Regular teachers* is the proportion of regular teachers in a private (government)school. *P(G)_Teacher hrs* is the number of hours spent by teachers in a private (government) school. *P(G)_teachers/classes* is the number of teachers divided by total classes in a private (public) school. *P(G)_Single teach* is a dummy indicating the private (public) school has only one teacher. *P(G)_index* is the index of teacher quality constructed using the variables in the previous columns. All these are school level variables constructed using the DISE data. *Ethnic frag* is the fragmentation index constructed using 1931 census. *urban* is a dummy indicating if the school is in an urban area, *Road access* indicates if the school is connected by all weather roads, *age* indicates the age of the school. These are school level control variables constructed using the DISE data. *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three district level control variables have been constructed using census data. Standard errors are clustered at district level.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 6: Effect of ethnic fragmentation on community meetings

| | (1) SMC meeting | (2) SMDC meeting | (3) PTA meeting | (4) Parents | (5) Local | (6) Total males | (7) Total females |
|---|---|---|---|---|---|---|---|
| Ethnic frag | 0.9541 | -3.9876** | -0.7527*** | -2.4806** | -0.7619* | -3.5561+ | -1.9070 |
| | (0.452) | (0.030) | (0.001) | (0.041) | (0.103) | (0.124) | (0.274) |
| urban | -2.9355*** | -0.7611*** | 0.1539+ | -1.1471*** | -0.4970*** | -1.8917*** | -0.8801*** |
| | (0.000) | (0.000) | (0.135) | (0.000) | (0.000) | (0.000) | (0.000) |
| Road access | -0.6244*** | 0.0225 | 0.1291*** | -0.2335** | -0.0606 | 0.0133 | -0.0342 |
| | (0.000) | (0.754) | (0.000) | (0.027) | (0.187) | (0.924) | (0.802) |
| age | 0.0365*** | 0.0185*** | 0.0001 | 0.0254*** | 0.0107*** | 0.0391*** | 0.0208*** |
| | (0.000) | (0.000) | (0.817) | (0.000) | (0.000) | (0.000) | (0.000) |
| college_dis | 0.0335 | -0.0262 | -0.0030 | 0.0768 | 0.0321 | -0.0780+ | -0.0149 |
| | (0.342) | (0.209) | (0.513) | (0.211) | (0.248) | (0.112) | (0.635) |
| workrate | 4.8626*** | 0.8927 | -0.4037 | 3.3788*** | 1.1681** | 6.7213*** | 2.4711+ |
| | (0.001) | (0.610) | (0.429) | (0.009) | (0.012) | (0.002) | (0.118) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 487251 | 79747 | 446377 | 79746 | 79747 | 79747 | 79747 |

*p*-values in parentheses

*Notes: SMC meeting* is the number of School Management Committee meetings held in a school. *SMDC meeting* is the number of School Management Development Committee meetings held in a school. *PTA meeting* is the number of Parent Teacher Association meetings held in a school. *Parents* is the representatives of parents in SMDC meetings. *Local* is the local community representatives in SMDC meetings. These variables are constructed using the DISE data. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. *urban* is a dummy indicating if the school is in an urban area, *Road access* indicates if the school is connected by all weather roads, *age* indicates the age of the school. These are school level control variables constructed using the DISE data. *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three district level control variables have been constructed using census data. Standard errors are clustered at district level.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 7: Impact of community meetings on school quality

| | Panel A | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | Regular teachers | Teacher hrs | Teachers/classes | Single teach | Index |
| PTA meetings | 0.0008* | 0.3713*** | 0.0587*** | -0.0158*** | 0.0868*** |
| | (0.054) | (0.000) | (0.000) | (0.000) | (0.000) |
| State FE | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 374348 | 379340 | 379341 | 379361 | 374327 |

| | Panel B | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | Regular teachers | Teacher hours | Teachers/classes | Single teach | Index |
| SMDC meeting | 0.0020*** | 0.0076 | 0.0001 | 0.0004+ | 0.0042* |
| | (0.003) | (0.689) | (0.973) | (0.124) | (0.090) |
| State FE | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 48163 | 48701 | 48716 | 48718 | 48146 |

*p*-values in parentheses

*Notes:* Sample in panel A table consists of government schools. Sample in panel B consists of government and private aided schools for which constitution of SMDC is mandatory.*Regular teachers* is the proportion of regular teachers in a school. *Teacher hrs* is the number of hours spent by teachers in school. *Teachers/classes* is the number of teachers divided by total classes in a school. *Single teach* is a dummy indicating the school has only one teacher. *Index* is the index of teacher quality constructed using the variables in the previous columns. All these are school level variables constructed using the DISE data. *PTA meeting* is the number of PTA meetings in a school. *SMDC meeting* is the number of SMDC meetings in a school. *urban* is a dummy indicating if the school is in an urban area, *Road access* indicates if the school is connected by all weather roads, *age* indicates the age of the school. These are school level control variables constructed using the DISE data. *workrate* is the proportion of people currently working in a district, *college_dis* is the number of colleges (per thousand population) in a district. These three district level control variables have been constructed using census data. Standard errors are clustered at district level.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 8: Impact of fragmentation on self-help groups and agricultural credit societies

| | (1) | (2) |
|---|---|---|
| | SHgroup | Agrisociety |
| Ethnic frag | -0.5411$^{***}$ | -0.4178$^{**}$ |
| | (0.001) | (0.022) |
| urbanisation | 0.0068 | 0.0980$^{*}$ |
| | (0.879) | (0.052) |
| workrate | -0.1509 | -0.2948$^{+}$ |
| | (0.349) | (0.104) |
| litrate | 0.0340 | -0.1155 |
| | (0.708) | (0.258) |
| State FE | Yes | Yes |
| Observations | 274 | 274 |

*p*-values in parentheses

*Notes: SHgroup* indicates the proportion of villages having self help groups in a district. *Agrisociety*is the proportion of villages having agricultural credit societies in a district. These variables are constructed using village directories in the census data. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. *urbanisation* is the proportion of population living in urban areas, *workrate* is the proportion of people currently working in a district, *litrate* is the literacy rate in a district. These three control variables have been constructed using census data.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 9: Impact of self help groups and agricultural credit societies on school provision

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | P_element | G_element | P_second | G_second | P_highsecond | G_highsecond |
| | | | Panel A | | | |
| SHgroup | 0.0711*** | -0.3506*** | 0.0222*** | 0.0075 | 0.0069* | -0.0037 |
| | (0.000) | (0.000) | (0.000) | (0.157) | (0.071) | (0.238) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| | | | Panel B | | | |
| Agrisociety | 0.0491** | -0.2776*** | -0.0014 | -0.0012 | 0.0033 | 0.0097** |
| | (0.042) | (0.000) | (0.873) | (0.858) | (0.495) | (0.018) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 331 | 331 | 330 | 331 | 311 | 315 |

*p*-values in parentheses.

*Notes: G_element, G_second and G_highsecond* indicate the number of government elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. *P_element, P_second and P_highsecond* indicate the number of private elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. These variables are constructed at the district level using DISE data. *SHgroup* indicates the proportion of villages having self help groups in a district. *Agrisociety*is the proportion of villages having agricultural credit societies in a district. Controls include *urbanisation*(proportion of population living in urban areas), *work participation rate*(proportion of people currently working in a district), lagged net state domestic product and proportion of development expenditure in a state.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 10: Alternative explanations

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | G_element | P_element | G_second | P_second | G_highsecond | P_highsecond |
| Ethnic frag | -0.1214 | -0.1550*** | -0.0253 | -0.0495** | -0.0057 | -0.0012 |
| | (0.634) | (0.003) | (0.508) | (0.039) | (0.828) | (0.937) |
| Dist GDP (in 100 cr.) | -0.0006*** | 0.0000* | -0.0000** | 0.0000+ | -0.0000 | 0.0000 |
| | (0.000) | (0.053) | (0.028) | (0.106) | (0.308) | (0.208) |
| Ethnic frag | 0.0361 | -0.1496*** | -0.0020 | -0.0544** | 0.0078 | -0.0064 |
| | (0.814) | (0.002) | (0.939) | (0.016) | (0.665) | (0.642) |
| Gini | -0.4795*** | -0.0134 | -0.0626*** | 0.0054 | -0.0381*** | 0.0119 |
| | (0.000) | (0.726) | (0.003) | (0.765) | (0.009) | (0.277) |
| Ethnic frag | 0.1384 | -0.1893*** | 0.0041 | -0.0820*** | 0.0133 | -0.0189 |
| | (0.397) | (0.000) | (0.883) | (0.000) | (0.490) | (0.188) |
| Group Gini | -0.3442** | -0.0391 | -0.0383 | -0.0278 | -0.0186 | 0.0034 |
| | (0.050) | (0.457) | (0.203) | (0.259) | (0.369) | (0.824) |
| Ethnic frag | 0.0086 | -0.1566*** | -0.0079 | -0.0569** | 0.0052 | -0.0074 |
| | (0.956) | (0.001) | (0.766) | (0.011) | (0.775) | (0.591) |
| Political frag | 0.0430 | -0.0337 | -0.0017 | -0.0151 | 0.0040 | -0.0099 |
| | (0.700) | (0.323) | (0.930) | (0.347) | (0.758) | (0.310) |
| Ethnic frag | -0.0433 | -0.1307*** | -0.0181 | -0.0461** | -0.0024 | -0.0016 |
| | (0.780) | (0.007) | (0.492) | (0.036) | (0.894) | (0.904) |
| SC_prop | -0.1666 | 0.0842* | -0.0205 | 0.0220 | -0.0288+ | 0.0261* |
| | (0.293) | (0.089) | (0.446) | (0.327) | (0.123) | (0.061) |
| ST_prop | 0.3776*** | -0.0160 | 0.0232** | -0.0233** | 0.0093 | -0.0053 |
| | (0.000) | (0.439) | (0.040) | (0.014) | (0.235) | (0.375) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 378 | 375 | 378 | 376 | 366 | 360 |

*p*-values in parentheses

*Notes: G_element, G_second and G_highsecond* indicate the number of government elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. *P_element, P_second and P_highsecond* indicate the number of private elementary, secondary and higher secondary schools (per thousand population), respectively, in a district. These variables are constructed at the district level using DISE data. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. All the parts of the table have control variables except for the first part as it has District GDP which is highly correlated with all the other district level controls. Second part of the table has *Gini* coefficient of real per capita consumption expenditure computed using NSS round conducted in the year 2011-12. Third part has *Group Gini* coefficient capturing inequality between caste groups. Fourth part of the table has *Political frag*, which is the fractionalisation index created using the vote share of political parties in a district in the parliamentary elections held in the year 2009. Fifth part controls the population proportion of Scheduled caste (*SC_prop*) and Scheduled Tribe groups (*ST_prop*).

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 11: Alternative explanations: Quality

| | (1) | (2) |
|---|---|---|
| | P_index | G_index |
| Ethnic frag | 0.1273 | -0.9157*** |
| | (0.640) | (0.004) |
| GDP (in 100 cr.) | 0.0000 | 0.0001 |
| | (0.676) | (0.569) |
| Ethnic frag | 0.0268 | -0.9683*** |
| | (0.923) | (0.001) |
| Gini | 0.4762* | 0.9087*** |
| | (0.061) | (0.000) |
| Ethnic frag | -0.0483 | -1.1972*** |
| | (0.855) | (0.000) |
| Gini group | -0.0222 | 0.7945* |
| | (0.953) | (0.062) |
| Ethnic frag | 0.0680 | -0.9480*** |
| | (0.806) | (0.001) |
| Political frag | 0.0279 | -0.3104 |
| | (0.877) | (0.217) |
| Ethnic frag | 0.0811 | -0.8050*** |
| | (0.760) | (0.001) |
| SC_prop | -0.9419*** | 0.3907 |
| | (0.000) | (0.151) |
| ST_prop | -0.4007*** | -0.2490** |
| | (0.004) | (0.022) |
| State FE | Yes | Yes |
| Observations | 115681 | 362132 |

*p*-values in parentheses

*Notes: P(G)_index* is the index of private (public) school teacher quality constructed using *P(G)_Regular teachers*, *P(G)_Teacher hrs*, *P(G)_teachers/classes* and *P(G)_Single teach*. All these are school level variables constructed using the DISE data. Variable of interest, *Ethnic frag*, is the fragmentation index constructed using 1931 census. First part of the table has District GDP as the control variable. Second part of the table has *Gini* coefficient of real per capita consumption expenditure computed using NSS round conducted in the year 2011-12. Third part has *Group Gini* coefficient capturing inequality between caste groups. Fourth part of the table has *Political frag*, which is the fractionalisation index created using the vote share of political parties in a district in the parliamentary elections held in the year 2009. Fifth part controls the population proportion of Scheduled caste (*SC_prop*) and Scheduled Tribe groups (*ST_prop*). Standard errors are clustered at district level.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 12: Impact of fragmentation on distribution of students by caste

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | P_frag_1 | G_frag_1 | P_frag_5 | G_frag_5 | P_frag_10 | G_frag_10 | P_frag_12 | G_frag_12 |
| frag_cen | 0.4245*** | 0.2392** | 0.2678*** | 0.4425*** | 0.2565*** | 0.3549*** | 0.2607*** | 0.4125*** |
| | (0.000) | (0.012) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Enrol_c1 | -0.0002*** | 0.0002 | | | | | | |
| | (0.002) | (0.378) | | | | | | |
| urban | -0.0032 | 0.0319* | -0.0029 | -0.0047 | -0.0013 | -0.0006 | 0.0055 | -0.0077 |
| | (0.693) | (0.085) | (0.670) | (0.772) | (0.819) | (0.940) | (0.372) | (0.527) |
| Road access | 0.0154** | 0.0218+ | 0.0266*** | 0.0216** | 0.0081 | 0.0239*** | 0.0148 | 0.0106 |
| | (0.021) | (0.102) | (0.000) | (0.030) | (0.252) | (0.000) | (0.162) | (0.276) |
| age | 0.0003** | 0.0008*** | 0.0011*** | 0.0010*** | 0.0011*** | 0.0010*** | 0.0008*** | 0.0007*** |
| | (0.025) | (0.004) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| college_dis | 0.0011 | -0.0062* | -0.0037 | 0.0010 | -0.0026 | -0.0039** | -0.0027 | -0.0016 |
| | (0.674) | (0.055) | (0.222) | (0.853) | (0.207) | (0.032) | (0.283) | (0.441) |
| workrate | -0.4235*** | -0.3380* | -0.2728* | -0.0803 | -0.1669 | -0.4515*** | -0.2243* | -0.3466*** |
| | (0.001) | (0.069) | (0.065) | (0.723) | (0.182) | (0.000) | (0.075) | (0.003) |
| Enrol_c5 | | | -0.0001* | 0.0001 | | | | |
| | | | (0.079) | (0.410) | | | | |
| Enrol_c10 | | | | | 0.0001*** | 0.0002*** | | |
| | | | | | (0.000) | (0.000) | | |
| Enrol_c12 | | | | | | | 0.0001*** | 0.0002*** |
| | | | | | | | (0.000) | (0.000) |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 18734 | 4575 | 24296 | 9059 | 46432 | 35274 | 18503 | 11361 |

*p*-values in parentheses

*Notes:* $P(G)_frag_i$ is a school level variable measuring ethnic fragmentation in class i in a private (public) school constructed using the DISE data. Coefficient of *Frag_cen*, fragmentation index constructed using 2011 census, indicates the impact of fragmentation for public schools. *urban* is a dummy indicating if the school is in an urban area, *Road access* indicates if the school is connected by all weather roads, *age* indicates the age of the school, Enrol_$c_i$ indicates enrollment in class $i$ in a school. These are school level control variables constructed using the DISE data. Standard errors are clustered at district level.

$^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 13: Robustness check: Spatial correlation

| | Dependent variable: P_element | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Point estimate | IID SE | HET SE | Spatial SE |
| Ethnic frag | -0.126 | -0.049 | -0.049 | -0.025 |
| State FE | Yes | Yes | Yes | Yes |

| | Dependent variable: G_element | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Point estimate | IID SE | HET SE | Spatial SE |
| Ethnic frag | -0.069 | -0.252 | -0.252 | -0.085 |
| State FE | Yes | Yes | Yes | Yes |
| Observations | 394 | 394 | 394 | 394 |

*Notes:* First panel has $P\_element$ as the dependent variable and the second panel has $G_e lement$ as the dependent variable. The first column in both the panels report point estimate of the impact of fragmentation on the dependent variable. Second column presents the standard errors which are estimated with the assumption of independent and identically distributed errors, third column allows for heteroskedasticity and the fourth allows for both heteroskedasticity and spatial correlation.

# Appendix A Data concerns

Since we rely on the DISE data for most of our empirical analysis, in this section, we address some of the concerns related with this dataset to make sure we are not picking up any bias in the data. DISE data, which started to be collected from 1995-96, has been envisioned to be the census of all the existing registered schools. It relies on *self reported* information on a range of school characteristics like the medium of instruction, year of establishment, whether the school is approachable by all weather roads, funds granted under government schemes, information on instruction hours, teachers, enrollment, school results among others. So, it's a detailed data on school particulars which is compiled at the district level. One of the concerns with the DISE data that has started getting recognition is regarding the coverage of schools in the DISE data. If the schools which report information in the DISE data are systematically different from the ones which do not report information and if this misreporting has a correlation with the fragmentation index, then our results would be biased.

To see if the sample of schools covered under the DISE data is representative and there is no correlation of the under reporting of data with the fractionalisation index, we construct the difference between the number of schools covered by the census and the DISE. Positive difference would imply that census covers more schools than DISE and that divergence in reporting of schools between census and DISE is more. We regress this divergence on the fractionalisation index to check the correlation between them. Results are reported in Table A.2. We find the coefficient of the fractionalisation index to be negative and significant for two of the three dependent variables. This points that if anything, divergence in the reporting of data is less in diverse districts and therefore, the concern of under reporting of data is less, which seems to be a good news for us. Also, the negative coefficient seems reasonable as the gap between the two data sets is about three years. In a rapidly growing environment there would be more schools in DISE since it is data from later years.

The other concern with the DISE data is that it is self reported. It is, therefore, likely that schools misreport the information provided by them in order to project themselves as "good quality" schools. This is plausible given that grants and other government benefits to schools in India generally increase with enrollment and teachers hired. An implication of this would be that DISE captures quality which is better than what is true in the practice. However, this would mean that our results are an under estimate of the true effect of fragmentation, atleast for public schools, since the results indicate that schools are of poor quality in fragmented places. Therefore, schools projecting themselves as better quality schools should not be a big concern for us.

Table A.1: Summary statistcs

| Variable | Mean | Mean (highly fragmented) | Mean (less fragmented) | Difference | t value |
|---|---|---|---|---|---|
| G_element | 0.4907 | 0.5113 | 0.4525 | 0.059 | (-1.77) |
| G_second | 0.0451 | 0.0473 | 0.0418 | 0.006 | (-1.05) |
| G_highsecond | 0.0191 | 0.0187 | 0.0181 | 0.001 | (-0.17) |
| P_element | 0.1210 | 0.1015 | 0.1435 | -0.0420*** | (-5.93) |
| P_second | 0.0469 | 0.0365 | 0.0591 | -0.0226*** | (-7.58) |
| P_highsecond | 0.0236 | 0.0175 | 0.0296 | -0.0121*** | (-7.21) |
| G_Regular teachers | 0.861 | 0.866 | 0.0860 | 0.006*** | (7.50) |
| P_Regular teachers | 0.952 | 0.947 | 0.962 | -0.015** | (-13.7) |
| G_Teacher/classes | 0.559 | 0.566 | 0.545 | 0.021*** | (15.06) |
| P_Teacher/classes | 1.015 | 1.104 | 0.935 | 0.169*** | (34.16) |
| G_Teacher hrs | 2.346 | 2.242 | 2.498 | -.256*** | (-24.20) |
| P_Teacher hrs | 3.749 | 4.141 | 3.619 | 0.52*** | (27.20) |
| G_Single teach | 0.177 | 0.166 | 0.200 | -0.034*** | (-27.32) |
| P_Single teach | .037 | 0.028 | 0.045 | -0.016*** | (-14.700) |
| Urbanisation | 0.2477 | 0.2402 | 0.2578 | -0.018 | (-1.03) |
| College_dis | 3.3754 | 3.3575 | 3.3369 | 0.021 | (-0.06) |
| Workrate | 0.4117 | 0.4196 | 0.4025 | 0.0172* | (-2.42) |

*Notes:* The second column of the table reports the average value of the variables, constructed using DISE data. The third column presents the average value for districts which have fragmentation level above the median value in the sample. The fourth column presents the average value for districts which have fragmentation level below the median value in the sample. The fifth column presents the difference in the third and fourth column, and the sixth column reports the t value of the difference.

Table A.2: Appendix table: Impact of fragmentation on data divergence

| | (1) | (2) | (3) |
|---|---|---|---|
| | diff_cd_element | diff_cd_second | diff_cd_highsecond |
| Ethnic frag | -1196.3765 | -345.0466** | -131.5926+ |
| | (0.156) | (0.036) | (0.116) |
| State FE | Yes | Yes | Yes |
| Observations | 393 | 393 | 372 |

$p$-values in parentheses

*Notes:* Dependent variables in all the three columns are the number of schools at the elementary, secondary and the higher secondary level, respectively, obtained from the census data minus the number of schools at the three levels obtained from the DISE data. Thus, these variables indicate the difference in the number of schools across the two data sets. *Ethnic frag* is the fragmentation index constructed using 1931 census.

$^{+}\ p < 0.15$, $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$