Application and admissions rates into the RTE quotas in private schools

Mauricio Romero* Abhijeet Singh†

July 29, 2024

Abstract

We evaluate interventions to raise application rates to quota seats in private schools (under Clause 12(1)(c) of the Right to Education Act 2009) in Chhattisgarh. There are four distinct experiments. The treatment/interventions are the same, but they are conducted in different samples: (i) Urban areas; (ii) disadvantaged rural areas; (iii) rural areas with under subscription, and; (iv) "representative" or "typical" rural areas. Within each experimental group/sample, we first conducted an informational treatment. We randomly assigned communities to receive information about the RTE quota and the application process. In the second stage, we have a household-level randomized treatment that offers assistance in applying for quota seats. Our main outcome is RTE application rates for quota seats.

*ITAM; J-PAL; E-mail: mtromero@itam.mx
†Stockholm School of Economics; J-PAL; E-mail: abhijeet.singh@hhs.se

1 Background

Clause 12(1)(c) of the RTE Act stipulates that 25% of the seats in the elementary classes of all private schools shall be reserved for students from disadvantaged economic and caste backgrounds. The fees for these seats would be reimbursed by the state. In Chhattisgarh, this has been implemented through a centralized allocation system where students are assigned seats through school-level lotteries. The act aims to improve social equity in schooling and to enable access to private education for children from vulnerable sections.

In our previous study conducted in 2021-22 on RTE 12(1)C titled "The incidence of affirmative action: Evidence from quotas in private schools in India", we use lottery-based estimates to show that, although students admitted under the quota attend more expensive and preferred schools on average, the distribution of program benefits is very regressive. We find that there are still a large number of eligible students who do not apply for an RTE seat and that the relatively poorer households have a lower probability of applying. Instead, most of the RTE applicants belong to comparatively well-off households who would have sent their children to private schools anyway. In this study, using a randomized intervention implemented in Raipur district, we find that low application rates for poorer children are not driven by preferences and beliefs. Instead, information constraints and application frictions appear to be key.

2 Interventions

There are two treatments. First, between mid-January and mid-March 2024, we provided information about the RTE policy (e.g., what it is, who can apply, how to apply, what documents are needed for application, how to get these documents, and the application timeframe) to all households in randomly selected sites. During this period, a baseline survey was administered in both treated and control communities.

The second intervention was conducted between mid-March and mid-April 2024 — when the application portal for the RTE quota seats was open — and was randomized at the household level (across sites, regardless of whether they received the information treatment).[1] We offer application support to randomly selected households. If the household expressed interest in applying, we visited the household and helped parents submit an application for an RTE quota seat: specifically, surveyors helped them fill out the online form using internet-enabled tablets or the caregivers' devices, upload the required documents, and select the schools they are interested in. Note that forms can only be filled if all requisite documentation is available.

[1] We provided application support in 160 out of the 240 rural sites and provided it in all the urban sites.

3 Experimental Design

We are conducting the same interventions in 4 distinct experimental samples. Three of these samples are in rural areas, and one is in urban areas. Our goal is to identify the efficacy of these interventions in different samples (and potential gains from targeting these interventions).

3.1 Sample selection

3.1.1 Rural areas

Our sampling procedure in rural areas uses the following datasets as the base to create the sampling frame: (i) the Unified District Information System for Education (U-DISE), with information on school characteristics, including management and location; (ii) the SHRUG database with geo-located village information on poverty rates and proportion of Scheduled Castes and Scheduled Tribes populations (based on Census 2011); (iii) administrative data from 2023 on the RTE lottery applications obtained from the Government of Chhattisgarh. These datasets are matched using geographical coordinates and location information.[2]

Sample selection in rural areas proceeds in the following distinct steps:

1. **Sampling universe:** We restrict the universe of our study to rural villages with at least one private school in five study districts: Bemetara, Kanker, Raigarh, Raipur, and Surajpur. Whether a village has a private school or not is determined based on a match between U-DISE data and SHRUG.

2. **Classification of villages:** We label villages as "undersubscribed" if they have unfilled seats (based on the administrative data from the state RTE program). We label villages as "disadvantaged" if they are in the top tercile of either the fraction of the SC/ST population or the top tercile of the poverty rate.

3. **Sample of villages:** We divide the sample of all rural villages with at least one private school into three randomly-assigned groups that are mutually exclusive. Within each of these groups, we randomly pick 80 villages. In the first group, these are picked out of the full set of villages through simple random sampling ("representative"). In the second group, these are randomly sampled from only those villages that are classified as "disadvantaged". In the third group, these are randomly sampled only from those

[2]Since there aren't consistent identifiers across the three major data sources, and administrative data on RTE applications is not accompanied by shape files, this initial matching was approximate. It was later verified by in-person surveyor visits.

villages that are "undersubscribed". These 240 villages together represent the intended sample of villages in the study.

4. **Village-level mapping:** The surveyor goes to the villages selected and verifies whether the site is fit for survey or not.[3] From a selected reference point in the village[4], then the surveyors mark the boundaries of the area to be surveyed.[5]

5. **Household listing**: In the identified geographical areas, surveyors conducted a household listing exercise and retained households that had a child of RTE-eligible age. Within the site, surveyors started at the center chosen during the mapping survey and then moved outwards. We surveyed all households with a child of eligible age up to a limit of 30 complete surveys. Our final baseline sample includes 5881 households in rural areas.

3.1.2 Urban areas

1. **Sampling universe:** We restricted the sampling to the Raipur/Bhilai/Durg metropolitan area. Specifically, we took the train station of each of these towns and created an 8 km buffer around them. We then identified all the schools within those limits. We removed any school that is more than 1km apart from any other school (we wanted to keep densely populated urban areas). We then created a buffer of 500 mts around each of the remaining schools. We then made a grid of this area of pixels that are 600x600 mts. We removed any pixel without a private school within 1.5km (walking distance). This is the "sampling frame".

2. **Sampling of survey sites:** We found the average fee charged by private schools within 1.5km of each pixel. We then classified pixels into deciles of average fees. We sampled a total of 200 initial sites, stratified by decile groups.

[3] S/he can reject the site mainly on the following grounds: a. No private school in the locality, b Posh colony, c. Industrial/Market area, d. Not enough houses, e. The site falls in another district. After location visits, we discarded 3 villages from the "disadvantaged" sample and 3 from the "under subscription" sample.

[4] Surveyors were provided a randomly-sampled private school as a suggested reference point. They verify if the school is sufficiently central in the school to be used as a centre. s/he looks for another center point, preferably a private school. In cases where surveyors can't find any private school that can be marked as the center, they mark an Anganwadi center or a govt. school, whichever is centrally located in the community as the center point. If they cannot locate the pinned school in the village, he will look for another private school. If the village has no private school, we dropped the village from the sample.

[5] After selecting the center point, he goes around the village, selects 5 points such that the village is best captured, and records the GPS of these five points. Each of the 5 points should be approximately 1 km from the selected center point. However, in some cases, these points might not be exactly 1km meters from the center point because of some hindrances while recording one or more points (e.g., there is a pond/road/posh colony/ the village is small). In most cases, we mapped the whole village for our survey because villages were small, and we had to reach the village boundary before completing the 1km distance from the center point. We went up to 1km from the center point on each side for large enough villages.

3. **Site-level mapping:** Surveyors verify in person whether a site is fit for sample inclusion.[6] The final sample has 158 urban "sites". For each site, surveyors identified a central location based on pre-set criteria.[7] After selecting the center point, he goes around the community and selects 5 points to capture the community and record the GPS of these five points.[8] This is the enumeration area (site).

4. **Household listing**: In the identified geographical areas, surveyors conducted a household listing exercise and retained households that had a child of RTE-eligible age. Within the site, surveyors started at the center chosen during the mapping survey and then moved outwards. We surveyed all households with a child of eligible age up to a limit of 30 complete surveys. Our final baseline sample includes 4208 households in urban areas.

3.2 Random assignment of interventions

3.2.1 Information treatment (cluster-level)

We administered a baseline survey to all households identified as belonging to the sample between mid-January 2024 and mid-March 2024. This baseline collected basic household information, socio-economic status, demographic details, and child education information. In 25% of the surveyed households, we collected data on with hypothetical enrollment choices for their children in order of preference under various hypothetical scenarios. Additionally, we posed a series of questions to ascertain the household's anticipated benefits and costs of different school options.

At the end of the baseline survey, in randomly-selected (treated) sites, we asked households about their awareness and prior take-up of RTE quotas and provided them with detailed information on the RTE policy and its application process.

The random assignment for this treatment was done as follows:

1. **Rural areas**: In each of the three samples of 80 villages in rural areas — representative villages; disadvantaged villages; undersubscribed villages — we randomly assigned

[6]They can reject the site mainly on 4 grounds: a. No Anganwadi or private or govt. school b. Posh colony, c. Industrial/Market area, d. Not enough houses, e. Site falling in another district.

[7]Once the site is accepted for the survey, the surveyor looks for all the Anganwadi within the pixel. He selects the Anganwadi, which is the most centrally located in the community, and records its location. For sites that have no Anganwadi falling within the pixel, we look for an Anganwadi within 300 meters outside the box. If they still don't find Anganwadi, they look for a government. school or a private school, which can be marked as the center of the community.

[8]Each of the 5 points should be approximately 300 meters from the selected center point. However, in some cases, these points might not be exactly 300 meters from the center point because of some hindrances while recording one or more points (e.g., there is a pond/road/posh colony on one or more sides).

40 villages each to treatment status and retained the others as controls. This was done before village mapping and stratified by district.

2. **Urban areas:** In urban areas, we randomly assigned 81 sites (out of 158 sites) to treatment, stratified by district.

3.2.2 Application assistance (household-level)

The household-level application support was provided in a tighter time window between the end of the baseline in mid-March and the end of the application period on April 20, 2024. The sample targeted for this intervention was selected as follows:

1. In rural areas, we selected 160 villages (out of the full baseline sample) for inclusion in this experiment. In urban areas, all sites were included.

2. In these sites, we randomly sampled households to include in application assistance intervention. Sampling was stratified by the information treatment status and by district. In rural areas, we selected 50% of children in each stratum to be offered application support. In urban areas, we assigned 50% of children in households with a child in AWCs and 33% of children in households without AWC children to be offered application support.

3. We eventually attempted to offer support to 1578 households in urban areas and 2047 households in rural areas.

4 Outcomes

4.1 Primary outcome

Our main outcome is whether households apply for an RTE seat. This will be measured directly in follow-up household surveys and through administrative data on lottery applications and allocations.

4.2 Secondary outcomes

We intend to analyze effects on four further (secondary) outcomes:

1. Whether a child in the household is *offered* an RTE quota seat

2. Whether the household enrolls a child in an RTE quota seat

3. Whether the household enrolls a child in a private school

4. Characteristics of the school enrolled in (such as fees and medium of instruction)

These outcomes reflect successive points of the causal chain from our intervention to the eventual intended outcome of the policy, i.e., expanded school choice for disadvantaged students. While our immediate interest is in understanding and addressing constraints to applying, we intend to map this full sequence of causal effects. We expect the pass-through from one outcome to the next to vary across different samples, reflecting the differences in household characteristics, the socioeconomic composition of neighborhoods, and the availability of RTE seats in the location.

These four outcomes should be available from follow-up surveys planned between July 2024 and September 2024; the first two outcomes should also be evident from administrative data on RTE offers and admissions.

For all outcomes, we will explore the incidence of benefits across socio-economic groups.

4.3 Potential downstream outcome

Our ability to statistically distinguish any effects on learning will depend on the first stage (effect of our treatments on school choice). We will conduct power calculations after the effects on the first four outcomes are known. If we are powered to detect moderate effect sizes, we shall then conduct further survey rounds to measure learning outcomes at the end of the academic year 2024-25. We anticipate, however, that the experiment is unlikely to be sufficiently powered to detect moderate effects.

5 Statistical analysis

5.1 Empirical specifications

We will conduct standard linear regressions to study treatment effects. When studying information treatment, we will cluster standard errors at the village level (the unit of treatment). When studying application assistance, we will cluster at the household level (the unit of treatment) and recognize that the treatment effect represents a counterfactual where some households had gotten information assistance in the past. We will add strata-fixed effects (district-fixed effects in the case of the information treatment and village-fixed effects for the application assistance treatment). In addition, we will include basic demographic controls (children's age, sex, caste, and household socioeconomic status).

We will also investigate, as a secondary analysis, whether there are complementarities between information provision and application support (i.e., interaction between being

treated in both the first and second intervention). This will be done in a separate "long" regression, following the procedure outlined in Muralidharan, Romero, and Wüthrich (2023).

5.2 Sub-groups

We are particularly interested in the heterogeneity of treatment effects across two main dimensions:

1. **By experimental sample:** Specifically, we will estimate treatment effects separately across our four samples (urban and three rural samples)
2. **By whether a household has a child enrolled in an anganwadi center at baseline:** We foresee this characteristic as a proxy of whether a household is likely to enroll children to a government school afterwards.

These subgroups represent different potential criteria for targeting interventions for information or application support. Comparing treatment effects across these subsamples will be informative of potential gains from targeting (as opposed to the untargeted provision of information and/or application support).

References

Muralidharan, K., M. Romero, and K. Wüthrich (2023): "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments," *The Review of Economics and Statistics*, 1–44.