The Role of Affirmative Action in Enrollment, Test Scores, and School Quality: Evidence from India [*]

Monica Agarwal [†]
Northwestern University

June 10, 2025

Click here for the most recent version

Abstract

I provide among the first causal evidence on the impact of the world's largest affirmative action program in entry-level primary education on student achievement. India's Right to Education Act (RTE) mandates private schools to reserve 25% of seats for disadvantaged students, yet despite being in existence for over a decade, its effectiveness remains understudied. Using lottery-based allocation of seats and combining administrative and primary survey data from a large Indian state, I find that RTE beneficiaries experience significant improvements in English test scores (0.18 SD), driven by access to higher quality private schools and increased time spent in educational activities. Importantly, I document substantial heterogeneity within the private sector: students attending elite private schools show gains of 0.48-0.69 SD in English. I make three key contributions: First, I demonstrate the effectiveness of affirmative action targeting early education when most such policies focus on higher education. Second, I provide causal evidence of how quality variation within the private sector affects children's outcomes, documenting that private school premium is not homogeneous. Third, I find that higher-quality private schools adapt better to educational disruptions in remote learning contexts, while demonstrating external validity of my overall results.

---

[*] I am extremely grateful to Laura Schechter, Jeff Smith and Priya Mukherjee for their invaluable guidance on this project. I also thank Moshi Alam, Brad Barham, Leah Bevis, Paul Dower, Alejandro Ganimian, Tarun Jain, Steven Lehrer, Ian McCarthy, Nishith Prakash, Diane Whitmore Schanzenbach, conference and seminar participants at NEUDC, Indian Statistical Institute, the AAE Development Lab, and AMIE mentoring group for helpful comments and discussions. I thank the team at NEERMAN, Nandish Kenia, and the team of enumerators for conducting excellent fieldwork and phone surveys. I gratefully acknowledge funding support for this research by J-PAL's Post Primary Education Initiative and a Dissertation Research Grant from the Department of Agricultural and Applied Economics at University of Wisconsin-Madison. Any errors are my own.

[†] Email: monica.agarwal@northwestern.edu

1 Introduction

Affirmative action policies serve as key policy instruments to promote social equity across the world. While the majority of such policies target individuals at later life stages through college admissions or workplace interventions, addressing disparities early in life may be more effective (Cunha, Heckman and Schennach, 2010). As one of the world's largest affirmative action policies that targets children of school entry age, India's Right to Education Act (RTE) provides a unique opportunity to study early-life affirmative action. The scale of the policy is huge - in 2018-19 alone, the policy benefited approximately 4 million children, and has the potential to impact about 16 million children, if implemented nationally (Indus Action, 2019; Romero and Singh, 2024). Despite being in existence for more than a decade, the effectiveness of this policy remains understudied due to lack of administrative data and absence of standardized test scores. In this paper, I provide one of the first assessments of the policy's effectiveness in improving educational outcomes among disadvantaged children.

The RTE mandates all private schools in India to reserve up to 25% seats for disadvantaged children at entry-level grades, with the goal of reducing segregation within classrooms. As a direct effect, the policy improves disadvantaged children's access to private schools. Thus, a first order question is to study the effectiveness of this policy, or in other words, the impact of attending a private school as a beneficiary under the policy.

In addition, the private schooling sector in India has been steadily growing and accounted for 45% of the primary grade enrollment in 2020.[1] Given the rapid growth in the market share of fee-charging private schools, both at the upper and lower end of the quality distribution, school effectiveness is likely to vary within the private sector. This in turn, signals the importance of examining the distribution of effects *within* the private sector.

Hence, in this paper I study two main questions. First, I ask: does enrollment in private schools through RTE quotas improve disadvantaged children's educational outcomes? Second, do the effects of this policy vary by the quality of private schools that beneficiaries attend?

I study this in the context of Maharashtra, the second most populous state in India. Allocation of private school seats to applicants under the policy is based on a lottery mechanism which ensures that applicants who submit the same school preferences, and live in the same neighborhood, have an equal chance of winning a seat at any given private school that they listed in their application. Those who win entry to private schools under this policy are eligible to get tuition-free education from these schools until they finish grade 8, with the government reimbursing the schools up to a cap. I exploit the feature of lottery-induced allocation of oversubscribed private school seats to identify the causal impact of the RTE policy on children's educational outcomes in an instrumental variables framework. To overcome the absence of standardized tests for young children in India, I collect primary survey data that includes comparable assessments across students and combine them with administrative data on school applications and lottery outcomes. Additionally, to account for endogenous school preferences submitted at the time of application, I utilize the within-variation in lottery outcomes

---

[1] See this link on data from the World Bank (2020).

of applicants who had a similar simulated ex-ante probability of winning the private school lottery given the allocation mechanism following Abdulkadiroğlu, Angrist, Narita and Pathak (2017). This approach not only identifies the overall impact of attending private schools under RTE but, additionally, enables causal analysis of the impacts of quality differences within the private sector—a rare quasi-experimental variation that provides an advantage compared to previous studies on quality differences within the private schooling sector.

I implement this using administrative data of the population of children who applied for grade 1 private school lotteries under RTE's 25% quotas, in the 2020-21 school year in Maharashtra. I supplement this with primary survey data from phone surveys, which I designed and administered with a sample of applicant households, to collect detailed information on children's education, schooling, and performance on phone-based assessments in English and Math. This gives me a sample of 2329 applicant households for whom I have rich data on household characteristics, children's schooling, their performance on phone-based assessments, their time-use, parental investments, and school inputs.

My data corresponds to the period of COVID-19 induced school closures.[2] While schools were closed for in-person instruction, the majority of schools transitioned to various forms of remote instruction (both asynchronous and synchronous) during the 2020-21 school year. Thus, this context provides me with a unique setting to also study how private schools adapt to external shocks to education provision. Furthermore, even though my results are from a remote learning context, I explain the external validity of my results later in the paper.

I show that the RTE policy led to economically substantial and statistically significant improvements in educational outcomes of children. One and a half years after exposure to RTE, quota children who won the private school lottery were more likely to be enrolled in school, with enrollment increasing by 13.3 and 4.6 percentage points in the 2020-21 and 2021-22 academic years, respectively.[3] While the increase in 2020-21 largely reflects RTE's role in insuring disadvantaged children against non-enrollment during the pandemic, the continued enrollment effect in 2021-22 suggests that RTE did more than merely mitigate this risk, and that RTE's enrollment effects extended beyond the pandemic. Moreover, the benefits are not just limited to enrollment, but also include gains in test scores of children - being a quota student at a private school improved performance in English by 0.18 SD (p-value < 0.05) and I find suggestive evidence of gains in Math (by 0.14 SD, p-value=0.12).

Understanding who benefits from the RTE policy—and how they benefit—presents an interesting empirical challenge. Unlike treated compliers who are a homogeneous group attending private schools under the RTE quotas, control compliers have multiple outside options to choose from: attending private schools as a fee-paying student, attending government schools, or being out of school altogether. In order to unpack this, I follow Abadie (2003), to estimate *counterfactual destinies* for the compliers and characterize the distribution of enrollment across

---

[2] However, the application window for RTE closed before the onset of pandemic in India and hence the applications submitted for private school admissions under RTE were not impacted by the pandemic.

[3] Applications for private school admissions under the RTE 25% quotas were made for the 2020-21 academic year, and primary data collection with a sample of these applicants was conducted during the middle of the following academic year i.e., 2021-22. This allows me to study enrollment decisions in these two academic years.

school types and quality for compliers. I find that the RTE policy's most significant impact lies not in merely expanding access to private schools, but in substantially improving disadvantaged children's access to higher-quality schools. My estimates of the counterfactual destinies substantiates this insight through two striking patterns: First the private schooling sector itself exhibits substantial quality variation. Among the treated compliers, 1 in 2 students are assigned to higher quality or *elite* private schools, and the other half are assigned to lower quality or *budget* private schools. Second, and more importantly, although 67.5% of the control compliers attend private schools anyway (as fee-paying students),[4] only 1 in 3 of these private school attendees – and just 1 in 5 of all control compliers—attend high quality private schools. Thus, despite suggestive inframarginality on the extensive margin, the RTE policy *nonetheless* improves disadvantaged children's access to higher-quality schools on the intensive margin.

Exploring mechanisms using administrative and primary survey data, I find that school inputs, and children's time use in school-based educational activities are the primary channels that explain these gains. Importantly, parental responses to school quality could potentially mask program effects on educational gains if parents respond by adjusting their own investments. However, I find limited evidence of this behavior: while parental monetary and time investments increase slightly for lottery winners, these effects are small, indicating that household responses play a minor role in explaining the observed gains. The key mechanism is the school quality itself: RTE quota students attend schools that are higher quality—scoring 0.6 SD units better in the overall school quality index.[5] These schools are more likely to have English as the primary language of instruction, teach more subjects, and have a longer school week (by 3 hours/week). This quality advantage is also evident in how these schools adapted to shocks to education provision: conditional on being enrolled, quota students were more likely to receive remote instruction in both academic years—by 7 and 3 percentage points in 2020-21 and 2021-22, respectively. Moreover, they were more likely to receive synchronous online modes of instruction (by 13.6 percentage points), relative to non-quota students who predominantly received asynchronous modes of instruction.[6]

One of the main insights of my paper is the substantial quality variation within the private sector itself: RTE beneficiaries at elite private schools score 0.48-0.69 SD higher in English than beneficiaries at budget private schools.[7] This variation reveals that the quality of the private school matters as much as—or more than—simply gaining access to the private sector. I estimate this heterogeneity, by first classifying schools as either elite or budget using two alternate measures of school quality.[8] Focusing on beneficiaries who won the RTE private

---

[4] Romero and Singh (2024) find 75% of lottery losers in the state of Chhattisgarh attend private school as fee-paying students.

[5] I create a school quality index using Principal Component Analysis (PCA) that combines administrative data on school infrastructure, digital facilities, and teacher qualifications.

[6] Asychronous modes include text-based communication via WhatsApp, and pre-recorded audio and video clips.

[7] I find no statistically significant differences in Math.

[8] The first measure uses a principal component analysis (PCA) based school quality index. The second measure uses administrative data on each private school's annual fee charged to fee-paying students. The two measures show a strong positive correlation, indicating that schools classified as elite based on the PCA-index measure are

school lottery, I estimate the LATE of attending an elite private school as a quota student. This utilizes the within-variation in lottery outcomes of children who live in the same neighborhood and had a similar simulated ex-ante probability of winning the elite private school lottery but faced a randomization in winning the lottery at elite versus budget private schools, given their school preferences (Abdulkadiroğlu et al., 2017). I find substantial heterogeneity within the private schooling sector - relative to quota students at budget private schools, those in elite private schools score 0.48 SD higher in English (when eliteness is defined using school fee), and 0.69 SD (when eliteness is defined using school quality index).

What drives these substantial performance differences between elite and budget private schools? I find that while both elite and budget schools were equally likely to provide remote instruction during the pandemic, elite schools were more likely to provide *synchronous* online instruction (10-18 percentage points more likely) and offered longer hours of class instructional hours, thereby increasing children's time in school based learning activities (2.1-3.1 additional hours weekly). Elite schools also offered a more diverse curriculum—though equally likely to teach core subjects like Math and English, they were significantly more likely to include additional subjects such as general knowledge, arts and crafts, music, and dance. Further, elite schools employ teachers with higher educational qualifications. These findings extend my earlier results in an important way: not only do private schools as a whole adapt better to educational disruptions than non-private schools, but within the private sector itself, there exists a quality gradient in adaptation, with elite schools adapting better to educational shocks compared to their budget counterparts.

My results document that the RTE policy significantly improves educational outcomes for disadvantaged students by providing access to better quality schools. Considered alongside Muralidharan and Sundararaman (2015) findings that private schools achieve comparable gains at lower per-student costs, this suggests that the RTE policy may be highly cost-effective as a means of expanding educational access while improving learning outcomes.

The external validity of my findings is supported by comparisons to prior studies that examine private school effectiveness in various *in-person* contexts. My estimates are strikingly similar to prior estimates of private school effectiveness when learning happens in-person. Muralidharan and Sundararaman (2015) find gains of 0.19 SD units (ITT) in English, but none in Math, for winners of private school vouchers in India after 2 years, Singh (2015) finds effect sizes of similar magnitudes using value-added estimates in Andhra Pradesh, India, and Romero, Sandefur and Sandholtz (2020) find gains of 0.18 SD in English studying the impact of allocating private management bodies to existing government schools in Liberia. For my second set of results which examines heterogeneity within the private sector, I compare my findings to Andrabi, Bau, Das and Khwaja (2024), the only other evidence on heterogeneity within the private schooling sector, who find gains of 0.21 SD within the private sector in Pakistan and highlight that different identification strategies can lead to varying estimates of the private school premium.

---

also likely to be elite based on the fee-measure.

My paper makes three key contributions. First, I provide among the first causal evidence on the effectiveness of one of the world's largest affirmative action policies targeting primary education. While most affirmative action research focuses on college admissions,[9] the impacts of such policies at school-entry age remain understudied despite their potential to address inequities at a critical stage of cognitive development when "*mismatch hypothesis*" is less relevant. This gap in literature is due to lack of standardized tests and digitized administrative records in India. To bridge this critical gap, I collect primary survey data, where trained enumerators administered the same assessment to students, ensuring comparability of test scores across those attending different schools.[10] Next, I combine them with administrative data on school applications and lottery outcomes to provide causal evidence on RTE's effectiveness. Additionally, I also collect primary survey data on parental and school inputs, and supplement it with administrative data on school quality to provide a comprehensive analysis of the mechanisms driving the impact of the RTE policy. Concurrent work has examined other aspects of the RTE policy. On the demand side, Romero and Singh (2024) document regressive selection under the policy in Chhattisgarh and focus on barriers that prevent the poorest households in their state–the policy's intended beneficiaries–from applying to the RTE.[11] On the supply side, Sahai (2023) documents how private schools respond to the RTE through fee adjustments in the short run. They also study the impact of the policy on student achievement in school-level assessments with the caveat that the assessments are typically school-specific, making comparisons challenging. Other attempts to estimate causal impacts suffer from identification concerns, such as endogenous school preferences.[12]

My second contribution is to provide evidence on elite school premium *within* the private sector using variation from lottery-based assignment to schools. The rapid expansion of fee-charging private schools has resulted in a diverse range of school quality, which highlights the importance of understanding the distribution of effects within the private schooling sector. However, it is rare to find quasi-random variation in assignment of children to private schools of differing quality which makes it challenging to identify effects within the private sector. I contribute by leveraging lottery-based random variation in assignment to elite versus budget private schools, and simulataneously also account for endogenous school preferences, to uncover heterogeneity within the private schooling sector in India and find considerable variation within the private sector with effect sizes ranging between 0.48-0.69 SD. The only existing evidence on within-private sector variation comes from Andrabi, Bau, Das and Khwaja (2024), who estimate school value-added using an AKM framework (Abowd et al., 1999) leveraging

---

[9] See for example, Arcidiacono and Lovenheim (2016); Bagde, Epple and Taylor (2016); Bertrand, Hanna and Mullainathan (2010); Bleemer (2022); Card and Krueger (2005); Dillon and Smith (2020), and Khanna (2020).

[10] All questions were extensively piloted by the team of enumerators before the final data collection. The phone-based assessment questions, designed to capture foundational language and numeracy skills, are adapted from Angrist, Bergman and Matsheng (2020) and an earlier version of Romero and Singh (2024), who conducted similar phone-based assessments with children in comparable age groups in Botswana and India, respectively.

[11] In an earlier version of their paper, Romero and Singh (2024) estimate the impact of being allotted an RTE quota seat on test scores in foundational numeracy and literacy skills, in Chhattisgarh, India.

[12] A working paper by Damera (2018) examines RTE policy impacts in Karnataka but relies on comparing lottery winners and losers without controlling for randomization strata, likely yielding biased estimates due to unaccounted endogenous school preferences.

variation from students who switch schools in Pakistan.[13,14]

Third, I provide evidence of private school effectiveness during remote learning. I find that private schools, especially higher quality ones adjust better to remote instruction and provide higher quality educational inputs. Thus my third contribution relates to two strands of the literature: research on mitigating pandemic learning losses using remote education and technological interventions (Angrist, Bergman and Matsheng, 2020; Azevedo, Hasan, Goldemberg, Geven and Iqbal, 2021; Carlana and La Ferrara, 2021; Beam, Mukherjee, Navarro-Sola, Ferdosh and Sarwar, 2021; Singh, Romero and Muralidharan, 2022; Guariso and Björkman Nyqvist, 2023) and research on private school effectiveness in developing countries (Angrist, Bettinger, Bloom, King and Kremer, 2002; Angrist, Bettinger and Kremer, 2006; Muralidharan and Sundararaman, 2015; Singh, 2015). Combined with the earlier discussion on external validity of my estimates, my results suggest that private schools, especially those that are high in quality, are effective not just during in-person settings, but also when learning is remote. Unlike Crawfurd et al. (2023) where remote learning interventions showed null impacts due to limited student engagement, my results show policy relevant advantages for public-private partnerships to build more resilient educational systems in low- and middle-income countries through well-implemented affirmative action policies.

The rest of this paper is structured as follows: Section 2 describes the policy and context (RTE quotas in Maharashtra, and the lottery algorithm); Section 3 describes the data sources (administrative data, and primary data collection) and sampling strategy; Section 4 describes the empirical strategy; Section 5 discusses results and mechanisms; Section 6 discusses the within-private sector heterogeneity; Section 7 discusses external validity and robustness checks; and Section 8 concludes, followed by Appendix tables and figures at the end.

2 Background and Policy

The Right to Education (RTE) Act was enacted by the Indian government in 2009, and made education a fundamental right of every child aged 6-14 years. I focus on a specific Clause 12(1)(c) of this act under which all private schools in India are mandated to reserve at least 25% of the seats in entry-level grades for children belonging to low socioeconomic (SES) families.[15] Children who get admitted to private schools under this policy are eligible to get free education from the respective schools until they complete grade 8. The government reimburses private schools to cover the school's tuition fee for children admitted under the quota. Children admitted under this quota are also eligible to get free textbooks and uniforms from the respective schools but the enforceability of this varies across states and schools. These quotas

---

[13] They find an average impact of 0.21 SD within the private sector. They also validate their SVA estimates by using plausibly exogenous switches resulting from school closures.

[14] Prior evidence on heterogeneity within schooling sectors from other country contexts provides mixed evidence - Pop-Eleches and Urquiola (2013) and Jackson (2010) find positive impacts of attending a better school in Romania, and Trinidad and Tobago, respectively. In contrast, Abdulkadiroğlu, Angrist and Pathak (2014); Dobbie, Fryer et al. (2011) and Cullen, Jacob and Levitt (2006) find no additional gains on test scores as a result of attending elite and high performing schools in the US.

[15] Religious and linguistic minority schools are exempted under the RTE Act. Entry level grades comprise grade 1 and pre-primary grades (for example, nursery or kindergarten).

were motivated in part due to the rapid increase in fee-charging private schools which made them inaccessible to the disadvantaged sections. Fee-charging private schools accounted for a total of 5.8% of enrollment in rural India in 2002 (Kingdon, 2007), however, their share has shot up to about 31% (primary school enrollment) in rural areas, and 50% in urban areas (Pratham, 2019). Due to the rapid growth in demand, there were growing concerns about the rise in segregation within classrooms with the well-off moving to private schools, and the relatively worse-off being in the government schools (which are free of cost). Thus, one of the goals of these quotas is to desegregate classrooms on the basis of socioeconomic status and improving access to quality schooling for all. The policy has been adopted by several states over the years, however remains unimplemented in several other states.

2.1 RTE quotas in Maharashtra: context and lottery mechanism

2.1.1 Private school quotas in Maharashtra

I study the impact of this policy in the context of the second most populous state in India, Maharashtra. Maharashtra adopted this policy in 2010 and the eligibility criteria includes children from historically disadvantaged caste groups, low income backgrounds, and children with disabilities.[16] The government reimburses schools for each child who is enrolled under this policy by sponsoring the school fee up to a certain limit and schools are not allowed to charge any fee to the quota students.[17]

2.1.2 Online applications

Maharashtra adopted a centralized online application system under this policy, in the academic year of 2017-2018. The online application to apply to schools under this policy begins in the month of February and is open for a month, following which the allocation of students to schools begins based on a centralized lottery algorithm. The majority of schools in the state follow the June to April school year.[18] The process of online application includes filling out the child's details along with household characteristics, for example the child's name, date of birth, gender, and household characteristics like religion, caste and income (if applying under the low income quota). The most important information that is filled out is the house address details, after which the system generates a list of all private schools available under the policy, in the child's neighborhood in three distance bins - all schools available within 1 km radius of the house address, within 1-3 km of the house address, and beyond 3 km of the house ad-

---

[16] Historically disadvantaged castes include Scheduled Castes, Scheduled Tribes, and Other Backward Classes (OBC). Low income families are defined as those earning less than INR 100,000 per annum ($4746 in PPP). In my administrative data for the year 2020-21, the majority of applications were received under the low income and disadvantaged caste category. Applications received under the disability category comprised 0.6% of the total applications.

[17] The reimbursement received by schools is equal to the value determined using the smaller of the these two amounts: school fee charged to fee-paying students, or the upper cap set by the government based on per-pupil expenditure in government schools in the state. The reimbursements have to be borne by centre and state governments in a 60:40 ratio. The policy has been slightly controversial since private schools may choose not to comply with RTE quotas if their fee levels exceed the reimbursement limits. As of year 2020-21, the per child reimbursement under RTE in Maharashtra was capped at INR 17,640 per annum (approximately 213 USD).

[18] A small number of schools follow the May to March school year.

dress (within the district). This is an important detail of the application process, which I come back to in my estimation strategy. Parents are allowed to choose a maximum of ten schools combining all three distance bins, but they cannot rank schools in order of their preference. They are also required to indicate the eligibility criteria which could be any one of these: low income category, disadvantaged caste category, or child disability category. Finally, parents sign an online declaration which says that in the event of winning a seat, parents are required to show a proof of house address (which must match the address reflected in the online application) and a valid proof that establishes their eligibility criteria under this policy. According to the rules, admission at allotted schools is guaranteed conditional on the house address documentation and other eligibility proofs being valid.[19] Importantly, the declaration states that the documents must be genuine, and in the case that any documents are found to be false or counterfeit, it may lead to monetary penalties and cancellation of the admission offer.[20] Since the policy is targeted towards disadvantaged households, help centers are organized during the weeks of the online application window (oftentimes in schools, and community centers) to specifically assist interested households with filling out the online application and answer questions. Similarly, in the weeks leading up to the start of the online application, the policy is advertised through notifications and billboards outside school premises, community centers, and local newspapers.

2.1.3 Lottery algorithm

States have considerable autonomy in how they implement the RTE quotas. Thus, the lottery mechanism that determines the allocation of students to schools under this policy also varies across states. In Maharashtra, it is designed such that each school assigns the highest priority to applicants who reside and applied in the nearest distance bin of the school (within 1 km radius of school, henceforth, distance bin 1), followed by those who reside and applied in the next distance bin (within 1-3 km radius of the school, henceforth, distance bin 2), followed by those who reside and applied in the farthest distance bin (beyond 3 km radius of the school, henceforth, distance bin 3). Hence, the overarching goal is to allocate applicants to schools which are closer to their house address. Importantly, parents are not allowed to submit rank ordered lists and can choose a maximum of ten schools. The lottery mechanism is a two-part process where the first part involves determining applicants who end up winning at a school, and the second part involves determining applicants who end up being waitlisted at a school. Applicants who are neither winners, nor waitlisted by the end, are those who lost at each and every school they applied to. The end result is that each applicant has one final lottery outcome which is tied to a unique school - they are either a winner at a unique school, or, waitlisted at a unique school (with a waitlist priority), or, have lost everywhere. In other words, if an applicant is a winner then they only won at one unique school; if they are waitlisted, then they did not win anywhere, but were waitlisted at one unique school; if they are neither a winner, nor waitlisted, then they lost at each and every school they applied to. Appendix Section B.1

---

[19] This could be an income certificate, caste certificate, or disability certificate based on whether the eligibility condition chosen is low income category, disadvantaged caste category, or disability category.

[20] In the administrative data I see that 0.6% of the admission offers were cancelled ex-post, due to false or improper documentation.

provides even more detail about the mechanism.

2.1.4 RTE School lotteries in Maharashtra, 2020-21

My administrative data corresponds to the universe of applications made under the RTE Act, for private school admissions in the academic year of 2020-21. Private school lotteries in the state were extremely competitive in the 2020-21 academic year. A total of 8848 private schools across the state participated in RTE quota admissions, and received applications from 291,365 children. Of these applicants, 35% won, 39% lost and 26% were waitlisted. Most applications were made under the disadvantaged caste category (63.5%), followed by the low income category. Since the applications under the RTE school lotteries were open only till the end of February 2020, the decision to apply to these school lotteries was made before the COVID-19 pandemic hit India (early March, 2020). However, the decision to take admission (in the event of winning a seat) is likely to have been disrupted due to the nationwide lockdown which was imposed in mid-March and thus unexpectedly coincided with the time when schools were offering admissions.[21]

3 Data

My data comes from four sources. First is the administrative data, which gives me details of the universe of children who applied to private school lotteries for grade 1 under the RTE quotas, in the entire state of Maharashtra, for the academic year of 2020-2021. Second, I use the U-DISE (Unified District Information System for Education) data which contains the administrative data of school characteristics of the population of schools in India. I use data from the 2019-2020 school year as that contained the most recent information on school characteristics prior to the RTE applications. Third, I use the administrative data on the annual school fee for all private schools in the state of Maharashtra from the 2019-2020 school year. Finally, the phone survey data, which I collected during the months of Nov-Dec 2021, by contacting a sample of households who applied to these lotteries (using the phone number provided by the household in their RTE application).[22]

3.1 Administrative data of RTE quota applications

This data provides the details of the universe of applicants who applied for grade 1 private school lotteries in the state of Maharashtra for the academic year of 2020-2021. These were publicly available at the Maharashtra Education Department website. For each child who applied, there was information about the child's name, child's date of birth, parents' name, parent's contact number, house address, religion, caste, household income, list of private schools chosen by the applicant in the three distance bins (within 1km, 1-3km, beyond 3km), and the

---

[21] Because of the nationwide COVID-19 lockdown beginning 24 March 2020, RTE admissions continued to be open through the month of December, 2020. Parents were offered the flexibility to complete the admission formalities either remotely or in-person.

[22] The administrative data on the population of applicants under this policy contained phone numbers of the child's parents which allowed me to conduct phone surveys with applicant households.

distance of each school to the house address of the applicant. For each child who applied, there was detailed information about their lottery outcome and how it evolved over time. To be precise, for every child who applied, there was data on the initial status of the application - whether their application was selected, wait-listed, or not selected anywhere. Each child could only have one of these statuses to begin with.

To explain this in further detail, if a child's application status was declared as selected, then it meant that they had won a seat at one of their preferred schools (if they win, they only win at one school and are excluded from all other schools that they had indicated); if the application status was wait-listed, then it meant that their application was wait-listed at just one of their preferred schools, and they were in the consideration set for admission to this school if a previously selected candidate gave up their seat (each wait-listed child would get a wait-list priority number such that a priority number of 1 would mean that this child would be the next in line for admission, if a vacancy was created at this school. This child was also excluded from all other schools if they had applied to multiple schools); if a child's application was not selected anywhere, then it meant that they were neither selected, nor wait-listed at any school that they had indicated in their application. Over time, the status of the application of a child evolves, and for each selected application, there is data on whether the child formally secured admission to the private school that was allotted to them and the corresponding date on which admission was secured (some students forgo their admissions and this creates vacancies for wait-listed children); for each wait-listed child, whether this child was finally admitted to the school that wait-listed them and if so, when they secured admission.[23]

3.2 Administrative data on school characteristics and private school fees

School characteristics: I use publicly available data on school characteristics from U-DISE for the 2019-2020 school year. This data covers the population of all private and public schools in India and has rich information on schools.

Private school fees: I use administrative data on school fees for all the private schools in the state which participated in the RTE lotteries in 2020-21. The data comes from the official website of the State Department of Education, Maharashtra and reflects school fees for the 2019-20 year.

Constructing two measures of school quality: I utilize these two datasets to construct two measures of school quality, categorizing schools as either elite (high-quality) or budget (low-quality) based on each measure. Specifically, schools that fall in top 25th percentile of the distribution of the PCA index are classified as elite on the PCA-based measure. Similarly, those falling in the top 25th percentile of the distribution of annual fees are classified as elite on the fee-based measure. A more detailed explanation is provided in Section 6.1.

---

[23] The status would evolve over time and the website put a notice of the deadlines by which selected candidates must approach their allotted schools to secure admission after which their admission would be null and void. Similar notices were put for the waitlisted candidates along with their priority numbers, and the process would continue to extend admission to candidates with lower priorities, until all seats were filled. Candidates were also sent SMS notifications about the deadlines on their registered contact numbers.

3.3 Primary survey data collection

I conducted phone surveys with a sample of applicants during the months of Nov-Dec 2021, to collect a rich data on children's outcomes, and household characteristics. A total of 4259 applicant households were contacted during this period, and successful interviews were completed with 2329 households (response rate of 55%). For each successful interview attempt, I also conducted a short interview with the applicant child to collect data on their learning outcomes in English and Math.[24] Among the full sample, a total of 695 households provide data on children's learning outcomes.[25] Response rates among winning and non-winning applicant households were about 57.7% and 52.4%, respectively. I discuss attrition and non-response bias in Section 7.2 and find that my results are robust to differential attrition, using inverse probability reweighting.

3.3.1 Sampling strategy

To select the sample of applicant households for conducting phone-surveys, I designed a sampling strategy. It is carefully designed to select a sample of comparable winners and losers under the policy, who are otherwise ex-ante similar in their household location and the school preferences that they listed in the RTE application.

The ideal comparison would involve comparing winners and losers who had the same school preferences by each distance bin, to begin with (as indicated at the time of submitting the online application). However, full stratification of applicants based on their distance bin-specific school preferences eliminates many schools and students from consideration (Abdulkadiroğlu, Angrist, Narita and Pathak, 2017).[26] In order to remedy this, I chose my sample such that the applicants who win and lose the private school lottery are comparable to one another to the extent that they made the same school choices in the *nearest distance bin*, i.e., schools chosen within 1 km radius of the house address; or, in other words, had chosen the same *school vector* in the nearest distance bin.[27] This in turn facilitates the comparison of winners and losers under the policy, who were ex-ante similar in their school preferences in the nearest distance bin and resided in the same geographic location. An important point to note is that the sampling strategy is designed to take into account only those schools which were *oversubscribed*, i.e., schools that conducted lotteries to admit applicants, and those applicants who were subjected to lotteries. The strategy is explained in detail in Appendix Section B.2, and a schematic

---

[24] The questions to test children on phone-based assessments come from Romero and Singh (2024) and Angrist, Bergman and Matsheng (2020) and are designed to capture foundational language and numeracy skills. The exact questions administered to children are shown in Figure A3 in the Appendix.

[25] To minimize non-response bias, the following rule was followed for calling households - each household was attempted to be called up to five times before discarding that number. The protocol was to attempt to call each household once during: the morning, afternoon and evening of a weekday; once on a Saturday, and once on a Sunday.

[26] The most ideal comparison would involve comparing children who differ in their lottery outcome but indicated the same school choice in each of the three distance bins, as this takes care of their endogenous choice of schools, household location, and their ex-ante likelihood of winning entry into schools as determined by the lottery algorithm. However, implementing this is difficult in practice given the high dimensionality of possible school choices over the full population of applicants.

[27] Throughout the paper, I frequently use the term *school vector* to refer to a unique combination of schools chosen in distance bin 1.

flowchart for the same is given by Appendix Figure B2.

3.3.2 Summary statistics

Table A1 summarizes the characteristics of applicants in the phone survey and also shows the key variables associated with the applicants and their household characteristics. The average applicant is about 7.6 years old at the time of interviews, slightly more likely to be male, and applied to about 5 schools in the RTE application. Some instances of non-enrollment exist in both the academic years, however, there is improvement in enrollment rates in 2021-22, with the easing of pandemic-related restrictions. Conditional on school enrollment, there is variation in the likelihood of schools providing instruction. Several other variables are summarized, such as monetary and time investments in children, their time use, and performance in phone-based assessments; these comprise my outcome variables.

3.3.3 Balance

I test for balance across winning and non-winning applicants to examine if they are similar on baseline observed characteristics. Appendix Table A2 presents the results, conditioning on the ex-ante propensity of winning at any bin. The majority of the characteristics are balanced across the two groups, with some exceptions - for example, father's education, religion, and household SES index. The joint F-test with a p-value of 0.16 fails to reject the null hypothesis of joint balance between winning and non-winning applicants. This suggests that the winners and non-winners are balanced on baseline observed characteristics.

4 Empirical Strategy

Using the administrative data of applicants who applied for private school admissions for grade 1 under the RTE quotas in the academic year of 2020-21, my goal is to estimate the impact of enrolling in a private school as a quota student on children's educational outcomes. The treatment group comprises the beneficiaries under the policy i.e., those who are enrolled as RTE quota students in private schools and the control group comprises non-quota students who may be attending private schools (as fee-paying students), or government schools (free of cost), and those not enrolled anywhere.

There are two endogeneity concerns here, and I address both of them. First, schools selected at the time of submitting the application are endogenous, and second, conditional on winning, the decision to enrol as a quota student is also an endogenous choice. Both these choices might correlate with unobserved household characteristics which might be simultaneously correlated with children's outcomes. I address both these concerns by using a conditional instrumental variables strategy. The idea is that, given the lottery algorithm, conditional on the school choices listed in the application, winning the lottery to a private school is random.[28]

---

[28] This follows from the lottery algorithm which satisfies the Equal Treatment of Equals (ETE) property (Abdulkadiroğlu, Angrist, Narita and Pathak, 2017). ETE is satisfied when students with the same preferences and priorities have the same chance of getting allocated at any given school. If the object of interest is winning a lottery at a school chosen in distance bin 1, then ETE is satisfied each time there is a group of applicants who had listed the

While conditioning on the school choices listed in the application solves the endogeneity in unobserved preferences for schools, the second endogenity problem is solved by instrumenting quota enrollment with the indicator of winning the lottery, which in turn is random conditional on controlling for the school choices that were listed in the application. Thus, I estimate the local average treatment effect of being enrolled as a quota student on children's outcomes in an instrumental variables framework. In addition to the main results which are LATE, I also show the intent-to-treat (ITT) effects of winning the RTE lottery on the outcomes of interest in Appendix Tables A4 & A5.

As I explain in the previous section, given the high dimensionality of school preferences, my sampling strategy is designed such that I can control for the vector of schools chosen in bin 1 and compare applicants who are similar to the extent that they had the same school preferences in bin 1. Conditioning on the vector of schools chosen in bin 1 is one way of addressing the endogeneity in school preferences listed at the time of the application. However, note that given the lottery algorithm and the ETE property, the relevant instrument to be used in such a case is winning the lottery in distance bin 1, which in turn means that the causal effect is estimated for compliers, defined by those who attend private schools as quota students because of winning the lottery in bin 1, and those who don't because they lost lotteries at bin 1 schools. On the other hand, if the instrument is winning the lottery in *any* distance bin, that leads to a much more heterogeneous composition of compliers, i.e., those who are quota students because of winning the lottery in any bin, and those who are not quota students because of losing the lottery in all bins.

Such an estimation can be executed by conditioning on the simulated ex-ante propensity scores of winning the private school lottery (Abdulkadiroğlu, Angrist, Narita and Pathak, 2017). This strategy is useful because it helps reduce the dimensionality of preferences and does not require me to explicitly control for the full set of schools chosen at the time of application. The idea is the following: taking the distance-bin-specific school preferences of applicants as given, I simulate the lottery algorithm a large number of times to arrive at the simulated ex-ante likelihood of winning the private school lottery, for each applicant. Since the simulated likelihood or propensity score takes into account the school preferences that were listed by the applicant, controlling for these propensity scores essentially performs a similar function as is achieved by explicitly controlling for the full set of schools chosen at the time of application. Since the goal is to estimate the LATE of being enrolled as a quota student under the RTE policy, the identifying assumption in this estimation strategy is that winning the lottery to a private school is conditionally exogenous after controlling for the ex-ante propensity scores of winning the lottery. Below I discuss the implementation of this strategy, which is my preferred

---

exact same schools in distance bin 1. If the object of interest is winning a lottery at a school chosen in distance bin 2, then ETE is satisfied each time there is a group of applicants who had listed the exact same schools in distance bin 1, and distance bin 2. Finally, if the object of interest is winning a lottery at a school chosen in distance bin 3 or, winning a lottery at *any* school in any of the three distance bins, then ETE is satisfied each time there is a group of applicants who had listed the exact same schools in each of the three distance bins.

approach.[29,30]

Following Abdulkadiroğlu, Angrist, Narita and Pathak (2017), my preferred estimation strategy involves controlling for the vector of dummies of narrow bins of ex-ante propensity scores of winning a lottery in any distance bin.[31] This strategy relies on comparing the winners and losers of private school lotteries, who had a similar ex-ante propensity of winning the lottery to an RTE private school (in any distance bin). This exploits the within-variation that results from comparing winners and losers who had a similar ex-ante propensity of winning any private school lottery, and does not require them to have chosen the same sets of schools. I estimate this using a two-stage least squares (2SLS) procedure, where the first stage is the effect of a random assignment of a private school seat on enrollment, and the second stage estimates the impact of quota enrollment on student outcomes.

I estimate the following equations via 2SLS:

$$RTE\_Enrolled_i = \alpha_1 WinningLotteryAnyBin_i + X_i'\alpha_2 + \sum_{x=1}^{100} \gamma_x d_i(x) + \epsilon_i \tag{1}$$

$$Y_i = \beta_1 \widehat{RTE\_Enrolled}_i + X_i'\beta_2 + \sum_{x=1}^{100} \delta_x d_i(x) + e_i \tag{2}$$

where, $d_i(x)$ are dummies taking a value of 1 if child i's estimated propensity score of winning a lottery at a private school in any bin lies in the respective 0.01 wide probability bin, X_i is the vector of child and household characteristics like sex and age of child, indicator for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. These covariates are added only to increase the precision of my estimates and the results are robust to excluding them. The coefficient of interest is given by β_1, which captures the LATE of attending a private school as a quota student on child outcomes. The compliers are those who attend private schools as quota students because they won the lottery to a private school (in any bin), and those who are without a quota because they lost the lottery at all schools that they listed in their application, and may be attending private schools as fee-paying students, or government schools, or may be out-of-school.

For some of the surveyed households, responses on certain conditioning variables are sometimes missing. Instead of a listwise deletion of observations that have missing values for co-

---

[29] I discuss the calculation of these propensity scores in Appendix Section B.3. I also show the distribution of these ex-ante propensity scores (Appendix Figure A2). Appendix Table B1 shows the detailed distribution of simulated propensity scores for the full population, and the sample.

[30] This strategy is powerful to deal with issues of stratification and sampling such as the one caused by fully stratifying applicants on the basis of their distance-bin-specific school preferences. It relies on comparing winners and losers who had a similar ex-ante likelihood of winning and does not require them to have chosen the exact same set of schools, thus bypassing some of the power issues which may occur if comparisons are based on controlling for school fixed effects.

[31] In the Appendix, I present results that condition on the school vector chosen in bin 1, and compare these results to the case which conditions on the simulated ex-ante propensity of winning in bin 1. Tables A10, A11, A12 show the results for the main outcomes. The two specifications produce very similar results thus providing confidence in the fact that conditioning on simulated ex-ante propensity scores performs a similar function as is achieved by conditioning on the school vectors.

variates, I re-code missing values of covariates to their mean value in the sample and control for these re-coded covariates, and include a separate missing value indicator in all the specifications. This is because listwise deletion of observations missing any of the conditioning variables would mean non-randomly dropping a substantial fraction of the sample (King, Honaker, Joseph and Scheve, 2001; Black, Smith and Daniel, 2005).

5 Results

5.1 First stage

Table 1 shows the first stage which captures the relationship between lottery offers and enrollment as a quota student. The endogenous variable of interest, i.e., enrollment in a private school as a quota student is instrumented by the indicator of winning the lottery at a private school, under the RTE policy. The instrument is random conditional on controlling for the narrow bins of simulated ex-ante propensity scores of winning. The results show that winning the lottery is strongly and positively correlated with enrollment as a quota student. The first-stage estimates are smaller than one because of non-compliance among lottery winners - some lottery winners choose to opt out of the quota seat at allotted schools as they may prefer other schools.[32] Another reason for the reduced estimate of the first stage is that some applicants who did not win any lotteries in the beginning, received an offer through the waitlist (at a later date).[33]

Table 1: First stage of winning the RTE lottery on enrollment as a RTE quota student

	Enrolled as RTE student
	(1)
Instrument = Winning the Lottery (any bin)	0.792*** (0.013)
Outcome mean	0.44
Control mean	0.08
Observations	2,329
R^2	0.66
Pscores of winning	Yes
Controls	Yes

Notes: This table shows the first stage effects of winning the RTE private school lottery in any distance bin on enrollment as an RTE quota student in a private school. This first stage corresponds to the 2SLS regression where the outcome of interest is school enrollment. Control variables include sex and age of child, dummy of father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

5.2 Main results

There are two outcomes of interest. First, enrollment in the two academic years and second, performance on phone-based assessments. Table 2 shows statistically significant gains in en-

---

[32] Some of this non-compliance may also stem from the fact that the timing of seeking admissions at allotted schools under the RTE policy coincided with the COVID-19 lockdown. However, the extent of COVID-19 induced non-compliance among lottery winners was reduced to some extent, as a result of schools allowing admission formalities to be completed over the phone.

[33] The first stage estimates differ across outcomes due to changes in sample composition - for example, the phone-based assessments are for a sub-sample of the surveyed households and some outcomes have missing responses leading to a reduced sample.

Table 2: LATE of being a RTE quota student on enrollment

	Enrollment (2020-21)	Enrollment (2021-22)	Grade 2 and above (2021-22)
	(1)	(2)	(3)
Enrolled as RTE student	0.141*** (0.016)	0.048*** (0.009)	0.194*** (0.017)
First stage F-stat	3,911.06	3,938.19	3,934.19
Outcome mean	0.89	0.97	0.86
Control mean	0.84	0.94	0.78
Observations	2,328	2,328	2,327
R^2	0.10	0.07	0.15
Pscores of winning	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on children's enrollment. The outcomes in columns (1) and (2) measure the indicator of school enrollment in the two academic years. Column (3) measures the indicator for whether the child is in grade 2 or grade 3 in the 2021-22 academic year. Controls include sex and age of child, dummy of father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

rollment in both academic years for treated compliers. Gains in enrollment are approximately three times higher in 2020-21 relative to 2021-22 suggesting that some of the children who were out of school during the first year of the pandemic (in 2020-21), are enrolled in schools in the following year (2021-22), with the easing in restrictions surrounding the pandemic. In 2020-21, the enrollment effect largely reflects that RTE acted as a safety-net for disadvantaged children during the pandemic and prevented them against the risk of non-enrollment. However, the fact that enrollment effects continue to exist in the following year reflects that the RTE was not just a safety-net during the pandemic but also helped in maintaining enrollment even as pandemic restrictions eased. Column (3) sheds light on another important aspect which is that RTE quota students were more likely to be able to maintain the right grade-for-age trajectory following their timely enrollment. Appendix Table A4 shows the ITT impacts of winning the lottery on these outcomes and the effect sizes are both qualitatively and quantitatively similar to the LATE results.

Table 3: LATE of being a RTE quota student on phone-based assessments

	Test score (standardized)	
	English	Math
	(1)	(2)
Enrolled as RTE student	0.187** (0.090)	0.143 (0.091)
First stage F-stat	1,129.88	1,129.88
Dependent mean	0.06	0.05
Dependent mean (control group)	-0.04	-0.04
Observations	695	695
R^2	0.17	0.13
Pscores of winning (any bin)	Yes	Yes
Controls	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on children's performance on phone-based assessments. Outcomes measure children's standardized test scores on English and Math and are standardized using the control group mean. Controls include sex and age of child, dummy of father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Being a quota student not only increases the likelihood of being enrolled and being enrolled in the right grade-for-age, but also leads to gains in performance on phone-based assessments. As can be seen in Table 3, there is a 0.18 SD unit increase in English performance for the treated compliers. Although, the effect on Math is statistically indistinguishable from zero (at conventional levels), it is quite similar in magnitude to English. My intent-to-treat (ITT) estimates are similar to the LATE, with lottery winners scoring 0.15 SD higher in English and 0.12 SD higher in Math (Appendix Table A5). These ITT effects align closely with Muralidharan and Sundararaman (2015), who find ITT gains of 0.19 SD in English for private school voucher winners after 2 years in Andhra Pradesh, India. Even though my findings are from a remote learning context, they demonstrate that private schools maintain their effectiveness when learning is remote. Similar to my findings, Muralidharan and Sundararaman (2015) find null impacts on Math, with mechanisms showing that these effects are primarily driven by differences in instructional time spent across subjects.[34]

The majority of existing evidence of school effectiveness is in the context of in-person learning. More recently, since the COVID-19 pandemic, there has been a growing interest in understanding the impact of remote instruction on children's educational outcomes, but we know relatively little about how schools adapt to changes in learning environments, and how this varies by school sectors, and whether private and higher quality schools in general, are still differentially effective in the context of remote learning.[35] My results suggest that virtual learning can be effective, and that private schools do a better job at adapting to, and implementing, remote educational technologies, and in doing so, they also enhance children's learning.

5.2.1 Counterfactual Destinies for lottery compliers

Since my estimates are of a LATE parameter, it is important to understand the composition of the compliers, as the causal impact of interest is relevant to this group. In particular, while the treated compliers comprise a homogeneous group of students (enrolled in private schools as quota students because of winning the lottery), the same is not true for the control compliers. The latter group comprises fee-paying students at private schools, students who go to government schools, and students who are out of school. Additionally, there is large variation in quality within the private schooling sector, a topic that I explore in greater detail in Section 6.1. Therefore it is helpful to examine the distribution of enrollment across school types in order to interpret the results. Angrist, Hull and Walters (2023) refer to this as *counterfactual destinies* for compliers. Following Abadie (2003), I not only characterize the distribution of enrollment status for compliers, across the various school types, but I also disaggregate schools by their quality given the underlying variation in quality in private schooling sector in India.

---

[34] In an earlier version of their paper, Romero and Singh (2024) find that RTE quotas for grade 1 students lead to gains of .27 SD units on student test scores (LATE; combined effects on Hindi, English, and Math) in a neighboring state of Chhattisgarh.

[35] A related paper is by Crawfurd, Evans, Hares and Sandefur (2023), who randomize primary school students in Sierra Leone to receive phone tutoring calls from public or private school teachers during the period of COVID-19 school closures. The teachers supplemented government provided radio instruction, but the intervention did not increase children's test scores, whether provided by private or public school teachers. They attribute this non-impact to limited take-up by children.

Table 4: Distribution of enrollment by school type

Panel A: Control compliers

	Quality	
Counterfactual School Destiny	PCA	Fee
Private school (Elite)	0.218	0.187
	(0.016)	(0.015)
Private school (Budget)	0.441	0.339
	(0.019)	(0.017)
Private school (quality is missing)	0.016	0.149
	(0.005)	(0.014)
Government (Budget)	0.191	0.191
	(0.015)	(0.015)
Out-of-school	0.052	0.052
	(0.009)	(0.009)
At school (but can't match school)	0.087	0.087
	(0.011)	(0.011)

Panel B: Treated compliers

	Quality	
Assigned school type	PCA	Fee
Private school (Elite)	0.445	0.563
	(0.016)	(0.016)
Private school (Budget)	0.544	0.384
	(0.016)	(0.016)
Private school (quality is missing)	0.011	0.052
	(0.003)	(0.007)
Government (Budget)	0.000	0.000
	(.)	(.)
Out-of-school	0.000	0.000
	(.)	(.)
At school (but can't match school)	0.006	0.006
	(0.002)	(0.002)

Notes: Complier enrollment outcomes are estimated using the IV strategy described in Abadie (2003). I describe the implementation of the exercise in Appendix Section C.1. The table reports the share of control compliers and treated compliers enrolled in school, by school type and quality. Schools are classified as elite on two alternative measures (principal component analysis (PCA) and school fees) if they lie in the top 25th percentile of the distribution of that measure (details in Section 6.1). Control compliers are enrolled at fallback options and are non-quota students because they lost the RTE lottery. Treated compliers are enrolled at private schools as quota students because of winning the lottery to that school, and these schools can either be elite or budget. Hence for treated compliers, share of students in government schools and share out-of-school are zero by construction. Means are reported and these are computed from 2SLS regressions that control for the ex-ante propensity scores of winning the lottery, as described in Abadie (2003). There are some children (primarily among lottery losers) for whom the school name and the official school code could not be matched with the administrative data on the population of schools. Thus, for these children the school sector – private or government – is missing. Additionally, quality data is missing for some private schools. Robust standard errors in parentheses.

Panel A of Table 4 reports the distribution of enrollment across various destination schools for control compliers - 67.5% of the lottery losers end up enrolling in private schools as fee-paying students, 19.1% attend government schools, and about 5% are out of school. However, amongst the private school attendees within control compliers, only 1 in 3 attend elite private schools. While all of the treated compliers are enrolled as quota students in private schools, in Panel B of Table 4, I disaggregate their enrollment based on whether they got assigned at an elite or a budget private school. Comparing the shares of *assigned school type* of treated compliers (Panel B of Table 4) to the *destination schools* for control compliers (Panel A of Table 4), I find that while the share of treated compliers who attend elite private schools is 1 in 2, the share of control compliers who attend elite private schools is 1 in 5.

Two takeaways emerge from this analysis: (i) while the RTE policy improves disadvantaged children's access to private schools, there is considerable variation in quality even within the private sector and a large share of students get allocated to budget private schools, a point that I explore in much detail in Section 6.1, and (ii) more than double the share of treated compliers attend elite private schools free-of-cost as a result of being assigned through the lottery than the share of control compliers who attend elite private schools (as a fee-paying student) due to losing the lottery. It is this quality differential in schools (and within the private sector) that plays an important role in shaping children's educational outcomes.

My findings complement the findings in Romero and Singh (2024) on inframarginality in RTE in the state of Chhattisgarh, by providing two additional insights. First, inframarginality varies systematically across contexts based on private school affordability—- while 75% of lottery losers in Chhattisgarh attend private schools,[36] 67.5% do so in Maharashtra, where considerably higher school fees (Kingdon, 2020) plausibly create greater barriers to private school access for disadvantaged families. Second, even with substantial inframarginality on the extensive margin, the RTE generates meaningful impacts on student achievement by improving access to higher-quality schools within the private sector for lottery winners.

5.3 Mechanisms

Children's cognitive achievement and human development are considered to be a cumulative process that depends on the history of family and school inputs, and on children's innate ability (Becker and Tomes, 1976; Todd and Wolpin, 2003). Following that, I explore the various mechanisms that might explain these improvements in test scores. In particular I study three channels - school inputs, parental inputs, and children's own time use and educational effort.

5.3.1 School inputs

First, I discuss the channel of school inputs. Table 5 looks at school characteristics that might matter in children's educational production function. The first two columns look into the likelihood of attending a private school and whether the school's primary language of instruction is English. Being enrolled as an RTE quota student increases the likelihood of both these outcomes for the compliers. A notable observation is the magnitude of these effects - the likelihood of attending a private school increases only by 20 percentage points. However, it is not surprising to see a small effect size, given the evidence of gradual exodus of children from government schools as a result of the increased affordability and demand for private schools (Kingdon, 2020). This is also consistent with findings from Romero and Singh (2024) and Damera (2018).

Table 5 also shows that conditional on being enrolled, treated compliers were more likely to be enrolled in schools that were actively providing instruction in the two academic years (columns (3) and (4)). The magnitude of the effect size is larger in the 2020-21 academic year (7 percentage points), relative to the 2021-22 academic year (3 percentage points). This indicates that schools attended by quota students were especially more effective in providing remote instruction during the year that coincided with pandemic-induced school closures, and that this effect size was reduced by about half in the following academic year. The reduction might come from a combination of various channels - schools being attended by control compliers might have become better at providing remote instruction over time, and the extent of being out of school fell among control compliers.

In the last three columns of Table 5, I look at the modality of instruction being offered at the school, conditional on being enrolled. Treated compliers were 13.6 percentage points more

---

[36] Since in Chhattisgarh parents are allowed to rank schools, Romero and Singh (2024) further find that 50% of the lottery losers end up enrolling in their top choice private school anyway. The RTE application rules in my context do not allow parents to rank schools hence, by design, I am unable to study whether the control group enrolls at their top choice school; however, I use school quality measures to show that despite high rates of private school attendance in the control group, the rate of enrollment at high quality schools is low.

Table 5: LATE of being a RTE quota student on school inputs

	School type		School instruction		School instruction modality		
	Private (2021-22)	English medium (2021-22)	Provides instruction (2020-21)	Provides instruction (2021-22)	Synchronous (online) (2021-22)	Recordings shared (audio/video) (2021-22)	Activity plans (WhatsApp/SMS) (2021-22)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Enrolled as RTE student	0.199*** (0.016)	0.089*** (0.013)	0.072*** (0.017)	0.030*** (0.007)	0.136*** (0.021)	-0.034* (0.018)	-0.066** (0.026)
First stage F-stat	3,856.22	3,859.05	3,472.00	3,877.71	3,788.83	3,788.83	3,788.83
Outcome mean	0.88	0.94	0.89	0.98	0.77	0.14	0.55
Control mean	0.79	0.89	0.85	0.97	0.70	0.15	0.57
Observations	2,249	2,250	2,083	2,255	2,210	2,210	2,210
R^2	0.16	0.08	0.11	0.05	0.15	0.08	0.10
Pscores of winning	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on school inputs. Columns (1) and (2) show the indicator of a child being in a private school, and the school having English as the primary language of instruction. Columns (3) and (4) show outcomes for whether the school provides instruction in the two academic years. Columns (5) - (7) show the type of instruction modality offered at the child's school in the 2021-22 academic year. The question was: in the past month, what were the types of instruction offered by child's school (select all that apply) - (i) online classes with teacher, other students (ii) pre-recorded lectures were sent (audio/video) (iii) written learning activity plans were shared via Whatsapp/SMS (iv) other, specify. This question was asked only to children who were enrolled in school in 2021-22, and whose school was providing instruction. Controls include sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

likely to be enrolled in schools that offer synchronous (online) classes, whereas control compliers were more likely to be in schools that offer recordings of lectures and share text-based activity plans (via WhatsApp or SMS).

Finally, Table A8 shows the LATE of winning the lottery on the likelihood that the school the student attends teaches any given subject, after conditioning on enrollment. Relative to the non-quota students, quota students are more likely to be in schools that teach English, Hindi, Environmental studies, Computers, General knowledge, Arts, Music and Dance. However, they are no more likely to teach Math.[37] These estimates suggest that gains in English could be driven through a combination of reasons. First, schools are more likely to teach English. Second, since they teach more subjects and given that the primary language of instruction is English, it is likely to further complement students' understanding of the English language.

The above exercise is more reflective of school inputs that might matter during remote learning, however, schools that are better in providing inputs that matter more during remote instruction could also be better in providing in-person instruction. I explore this next using measures of school quality that are more reflective of business-as-usual settings.

School quality differences reflective of business-as-usual times: I measure school quality by creating indices of broad categories of school-level characteristics using principal component analysis (PCA). Appendix Table A7 reports the estimates and I find that quota students are more likely to enroll in schools that have better infrastructure facilities, digital facilities, teacher quality, and have a less diverse student composition.

---

[37] A caveat is that this information comes from parental responses, and not from the school, so there could be measurement error in the data.

Overall, these results suggest that treated compliers end up in schools that were more likely to provide instruction and also provide synchronous modes of instruction - which is arguably more effective and holds both teachers and students more accountable, by offering real-time interaction. Furthermore, these schools have better infrastructure and have teachers with more qualifications, both of which are likely to enhance children's learning during business-as-usual times. Finally, these results are reflective of my earlier discussion on counterfactual destinies where a large share of treated compliers attend elite private schools who offer better school inputs compared to the control compliers.

5.3.2 Parental inputs

Next I discuss the channel of parental inputs in children's education production function. If parental investments change as a result of quota receipt (Das et al., 2013; Pop-Eleches and Urquiola, 2013), then the LATE of attending a private school as a quota student on test scores reflects an overall effect of school inputs and home inputs on children's achievement (Becker and Tomes, 1976; Todd and Wolpin, 2003). I collect primary survey data on parental monetary and time investments to test the extent to which parents adjust their time and monetary investments in children in response to winning the quota seat both on the extensive and intensive margin. I find that parents increase their investments in winners, however the effect sizes are small, indicating that even though parental inputs are increasing, they only explain a small part of the story.

Table 6: LATE of being a RTE quota student on parental investments

	Time investments		Monetary investments	
	Receives help with homework	Hours of help (hrs/week)	Any expense (past year)	Expenditure (past year)
	(1)	(2)	(3)	(4)
Enrolled as RTE student	0.020 (0.013)	0.543* (0.308)	0.065*** (0.014)	-90.182 (158.303)
First stage F-stat	3,915.91	3,915.91	3,822.52	3,822.52
Outcome mean	0.93	9.50	0.93	3,462.86
Control mean	0.92	9.31	0.91	3,467.37
Observations	2,329	2,329	2,227	2,227
R^2	0.08	0.06	0.06	0.06
Pscores of winning	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on parental time and monetary investments in children - both on the extensive and intensive margins. Column (1) measures the extensive margin of whether the child receives any help with educational activities in the household, and column (2) measures the intensive margin of the number of hours of help. The survey questions were: "Does the child receive any help with educational activities from any members of the household?" followed by details of each person who helps and their relationship with the child. Next, it was asked: "Among all those who help, who is the person who most often helps the child with educational activities?", followed by details about number of hours per day of help on a typical day, and number of days of help per week in the past week, to calculate weekly hours of help coming from the main helper. Hence, data on hours of help are collected only for the main helper. Column (3) measures the extensive margin of any educational expenses in the child in the *past one year* (on curriculum books, notebooks, and stationary), and column (4) measures the intensive margin of the amount of expenditure incurred on child's education in the past one year. There are some missing values for the monetary investment questions due to item non-response. Controls include sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Column (1) of Table 6 shows that while there is no statistically significant impact on the extensive margin of children receiving household help with educational activities, there is evidence on the intensive margin as treated compliers are likely to receive more help per week with educational activities (approximately 30 mins more per week), relative to the control compliers. Further analyzing this in Appendix Table A9, I find that among all the household members, mothers are the ones who help children with educational activities. Turning to monetary investments, Column (3) of Table 6 shows that parents of lottery winners are 6.5 percentage points more likely to spend on the educational needs of children in the past year (on curriculum books, and stationary). However, there is no detectable impact on the intensive margin (Column (4) of Table 6).

Together, these results suggest that parents respond to the receipt of the RTE quota seat by reinforcing investments in children. However, the small magnitude of the effects suggest that this channel represents only a small part of the story.

5.3.3 Children's time use

Third, using my primary survey data, I explore the channel of children's own effort by studying the impact on their time use. LATE estimates reported in Table 7 shows that quota students spend more time in educational activities. Specifically, treated compliers spend 3 more hours per week doing school-related activities, and approximately 20 more minutes per week doing homework. I don't find any statistically significant differences in time spent on private tutoring (after school classes) and in non-educational activities although the estimate is negative which could suggest substitution away from private tutoring in response to winning the lottery.[38] These increases in children's time use is reflective of my earlier discussion on counterfactual destinies where a large share of treated compliers attend elite private schools who offer longer school week. Additionally, my findings are also consistent with Muralidharan and Sundararaman (2015), who find no impact of winning private school vouchers on home study- and play-habits except increased time spent in school for voucher winners.

6 Winning in Elite versus Budget RTE private schools

My results so far provide evidence that winning entry to RTE private schools improves children's learning outcomes. The mechanisms suggest that school's mode of instruction, and children's effort in educational activities play important roles in achieving these gains. However, even among the class of private schools attended by winners, I find substantial variation in the quality of schools attended (Table 4). Private schools that have a higher school quality index (PCA) or levy a high yearly school-fee are likely to have qualified and motivated teachers, and thus likely to offer higher quality of education, have better resources, and as a result might have a higher value-added.[39] In contrast, private schools on the lower end of the quality

---

[38] Private schooling differs from private tutoring in the Indian context. Tutoring typically happens after school hours.

[39] Previous literature has used school fees as a proxy for school quality. Rao (2019) defines elite schools as those charging a fee greater than 2000 INR per month, in New Delhi. Andrabi, Bau, Das and Khwaja (2024) find a positive correlation between school value-added (SVA) and school fees in Pakistan. Romero and Singh (2024) analyze the impact of winning a quota seat under RTE on the market price of the school being attended, in the state of Chhattisgarh, India, and find that quota students enroll in costlier schools.

Table 7: LATE of being a RTE quota student on children's time use

	Educational activities			Non-educational activities		
	School (hrs/week)	Tuition (hrs/week)	Homework (hrs/day)	Playing (hrs/day)	Television (hrs/day)	House chores (hrs/day)
	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled as RTE student	2.945***	-0.405	0.262***	-0.099	-0.073	-0.006
	(0.400)	(0.315)	(0.038)	(0.062)	(0.047)	(0.021)
First stage F-stat	3,915.91	3,915.91	3,915.91	3,909.99	3,938.97	3,915.91
Outcome mean	12.12	4.67	1.40	2.45	1.10	0.39
Control mean	10.79	4.92	1.31	2.50	1.15	0.40
Observations	2,329	2,329	2,329	2,328	2,322	2,329
R^2	0.13	0.08	0.06	0.05	0.09	0.06
Pscores of winning	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on children's time use in educational and non-educational activities. School hours are set equal to zero for those who report being not enrolled in any school. Tutoring hours (differs from formal schooling, typically happens after school) are also set equal to zero for those who report being not enrolled in any private tuition. The question for homework hours is not always zero for not enrolled children, as the question asked - "how much time does child spend doing homework, or any educational activities after school?". There are some missing values for playing and watching television due to item non-response. All time use data are winsorized at the 99th percentile. Controls include sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

distribution are likely to have fewer resources, and fewer teachers with high qualifications and thus likely to have lower value-added. Thus, even among the RTE winners who benefit from a quota seat at private schools, school quality is likely to differ, which may lead to differences in children's achievement. But do these differences in school quality matter during periods of remote learning?

There is no evidence on the how school effectiveness varies within the private sector in India. The only such evidence from a similar context is from Punjab in Pakistan, by Andrabi, Bau, Das and Khwaja (2024). Using value-added models (VAMs), they find evidence of substantial within-village variation in school quality within the private schooling sector. Contrary to the existing literature that has largely focused on a single private school premium, their findings suggest a range of causal estimates of the private school premium, resulting from a substantial *within-sector* variation in school quality. If there exists variation in quality within private schools, then it might be misleading to focus on a single estimate of the private school premium.

I examine this in my context using lottery-based variation in assignment to elite versus budget private schools. The next question that arises is how one should arrive at a reliable measure of school quality. In this section, I uncover the impacts of relative differences in private school quality on children's educational outcomes by using two alternative measures of school quality which I discuss in detail in Section 6.1. The next paragraph briefly summarizes how the literature defines school quality.

The idea of school quality is a latent concept, and the literature has looked at various ways of measuring the *true* school quality. The bulk of the literature on school quality in the US focuses on achievement-based measures of quality, and more recently on outcomes other than student achievement, for example, crime, employment, earnings and non-cognitive outcomes (Angrist et al., 2023).[40] Several papers use peer ability and socioeconomic composition of peers as proxy for school quality (Abdulkadiroğlu et al., 2014; Pop-Eleches and Urquiola, 2013; Dobbie and Fryer Jr, 2013). Greaves et al. (2023) use school inspection ratings as a source of information on school quality in the context of England. School management interventions that improve the quality of leadership practices have also been utilized to get at measures of school quality.[41] Specific dimensions of school inputs have also been used to measure school quality, such as class size (Datar and Mason, 2008; Fredriksson et al., 2016) or school resources (Houtenville and Conway, 2008; Das et al., 2013). Another important and related strand of literature is on college quality, where the goal is to study the educational and labor market effects of the quality of college that individuals attend. Black and Smith (2006) discuss the issues with using a single proxy of college quality (such as the average SAT score of the entering class) as it leads to substantial measurement error in the quality measure, and propose several solutions. One of the proposed solutions is to create a quality index that combines multiple individual quality measures (or proxies) via factor analysis. The larger the number of quality variables, the less is the measurement error in the index (Black, Smith and Daniel, 2005). I take inspiration from them to define one of my school quality measures in a similar way, as I discuss in more detail in the next section.

6.1 Two alternate measures of school quality and eliteness

In the absence of panel data on standardized test scores across schools, and school level peer achievement, I consider two ways of defining school quality - one that uses data on a rich set of school-level characteristics, and another that uses school fees (charged to fee-paying students).[42] I start with a simple case of categorizing schools as *elite* or *budget*, based on two measures of quality.[43]

The first measure of school eliteness is based on a school quality index that I construct using Principal Component Analysis (PCA). The data for this analysis comes from UDISE, which allows me to make use of a rich dataset on school characteristics – infrastructure details, digital facilities, teacher qualifications, and peer SES composition – which might matter in determin-

---

[40] Angrist et al. (2023) provide a useful review of this literature by summarizing the various econometric strategies for estimating school effectiveness - school lotteries using the instrumental variables approach, regression-discontinuity approach where students are admitted based on a cutoff score, centralized school assignment where school allotment happens via conditional randomization based on rank ordered lists submitted by parents, and finally value-added models (VAMs) which control for lagged outcomes and covariates by making use of panel data of student test scores.

[41] Anand et al. (2023) conduct a meta analysis of the impact of school management interventions on student learning using data from multiple evaluations, and provide a systematic review of this literature.

[42] The school-quality-based measure is in the spirit of the college quality literature that uses multiple college characteristics to create an index of college quality, such as Black, Smith and Daniel (2005) and Black and Smith (2006). The school fees based measure is in the spirit of prior literature that uses fees to define school eliteness and quality, such as Rao (2019); Romero and Singh (2024) and Andrabi, Bau, Das and Khwaja (2024).

[43] Note that, while both my measures of quality are continuous measures of quality, the discretization of schools into elite and budget is done following the identification strategy which relies on the within-variation in lottery outcomes of children with similar ex-ante propensities of winning the lottery *at elite* schools. This in turn requires that each school that was chosen during the time of application, be categorized as a binary of either elite or not-elite to get at the simulated ex-ante propensity score of winning at an elite school. A caveat however is that such discretization leads to a loss of information and in turn causes researcher-induced measurement error in the quality index (Black and Smith, 2006).

ing the overall school quality. The complete list of variables that are used for creating this index is shown in Figure A4 in the Appendix. I use the first component of the PCA to create the quality index. The figure also shows the factor loadings on each of the variable - it shows that all these different types of school inputs are positively associated with school quality.[44]

The second measure uses administrative data of schools' annual fee to categorize each school as elite or budget. Figure A5 shows the distribution of annual school fee for private schools in the state using the administrative data of annual school fee charged by RTE private schools in the state. As the figure shows, most of the private schools are concentrated on the lower end of the fee distribution. This is in line with Kingdon (2020), who documents that the vast bulk of private schools in India are low-fee schools, when benchmarked against the state per capita income and daily wage laborer's incomes. The author also points out that this increase in affordability has led to a rapid migration of students towards private schools, and an emptying of government schools. Taking the distribution of annual school fee for all the RTE private schools in the state, I define a school as elite if the annual school fees exceeds the 75^{th} percentile in the distribution of fees of all private schools in the state, and budget, otherwise.[45]

These two measures of school quality display a strong and positive correlation - Table A14 shows the results from OLS estimates of a simple linear model where I regress school fees on school's PCA based quality index.[46] While both measures capture school quality, I prefer the PCA measure over the fee-based measure because it relies on school characteristics that are less likely to be endogenously adjusted by schools. In contrast, school fees may be susceptible to such adjustments. [47]

6.2 Estimating the impact of attending elite private schools as a quota student

For the group of lottery winners, what is the impact of attending elite private schools as a quota student, relative to attending budget private schools? I estimate this using two-stage

---

[44] Government schools, which do not charge any school fees, are classified as budget on the fee-based measure. They also fall in the bottom 75^{th} percentile of the PCA index, and are thus classified as budget on the PCA-based measure. Most of the government schools being attended by the non-compliers among the lottery winners are zilla parishad schools (state-run schools which are established, funded and supervised by the district councils of India), which lie at the lower end of government school quality distribution.

[45] I vary the bar of eliteness by lowering and increasing the threshold to the 50^{th} and 90^{th} percentiles, respectively. The treatment contrast in school quality is low with 50^{th} percentile and using the 90^{th} percentile leads to power issues (as few schools are categorized as elite). In the sub-sample of lottery winners (which is the relevant sample for this exercise), approximately 4% of the children attend government schools and only 2% of the children attend elite schools but not as an RTE student.

[46] Tabulating schools on these two measures of eliteness shows that majority of the schools are consistent in the elite definition across the two measures (Figure A6). About 65% of the schools are elite on both measures, about 25% of the schools differ in classification of eliteness across the two measures, the rest have missing data on one of the two measures.

[47] Sahai (2023) documents that private schools in Madhya Pradesh respond to RTE quotas by raising fees in the short run (6 years after policy implementation). However, I consider the school-fee based measure reliable for my analysis, as I am estimating the effect of this policy 10 years after the RTE was first introduced in Maharashtra, implying that any general equilibrium effects on fee hikes would likely have already been realized by my analysis period. Additionally, in Maharashtra, private schools are not fully autonomous in raising fees; government rules mandate that schools can *propose* fee hikes of up to 15% every two years.

least squares framework on the sub-sample of lottery winners:

$$RTE\_Enrolled\_Elite_i = \alpha_1 WinningLotteryEliteAnyBin_i + X_i'\alpha_2 + \sum_{x=1}^{50} \gamma_x d_i(x) + \epsilon_i \tag{3}$$

$$Y_i = \beta_1 \widehat{RTE\_Enrolled}\_Elite_i + X_i'\beta_2 + \sum_{x=1}^{50} \delta_x d_i(x) + e_i \tag{4}$$

where, $RTE\_Enrolled\_Elite_i$ is the indicator that child i attends an Elite private school as a quota student, $WinningLotteryEliteAnyBin_i$ is the indicator that child i won the lottery at an Elite school in any bin, X_i is the vector of child and household characteristics, $d_i(x)$ are dummies taking a value of 1 if child i's estimated propensity score of winning a lottery at an elite private school lies in the respective 0.02 wide probability bin. As before, identification comes from within variation in lottery offers at elite or budget schools for groups of applicants who are otherwise similar in their ex-ante propensity of winning at elite schools but faced a randomization in whether they won the lottery at an elite or a budget private school.[48]

Table 8: First stage of winning the RTE lottery at elite school on enrollment at an elite school

	RTE student at Elite school	
	Elite (PCA)	Elite (Fee)
	(1)	(2)
Won RTE lottery at Elite school	0.880***	0.869***
	(0.027)	(0.028)
Outcome mean	0.39	0.51
Control mean	0.00	0.00
Observations	1,019	973
R^2	0.85	0.82
Pscores of winning at elite	Yes	Yes
Controls	Yes	Yes
Avg quality (Elite=1)	4.37	43046.62
Avg quality (Elite=0)	2.34	12320.27

Notes: This table shows the first stage effects of winning the RTE private school lottery at an elite school on enrollment at an elite school as a quota student. The sample is restircted to lottery winners. Eliteness is defined using the 75^{th} percentile cutoff. Control variables include sex and age of child, indicators for father's and mother's education being greater than the respective means, indicators of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

6.2.1 First stage

Table 8 shows the first stage - winning the lottery at an elite school, defined as schools lying above the 75^{th} percentile of the quality distribution, increases the likelihood of attending one, by 87pp - 88pp depending on the quality measure. The dependent mean shows the proportion of children enrolled at elite schools under the quota - this is 39% based on the PCA index measure, and 51% based on the fee-based measure. These differences stem from each measure identifying a different aspect of school quality and relatedly the fact that the same school might be categorized as elite based on one measure, but as budget on the other. It is also informative to learn about the average quality of schools that are categorized as elite versus those catego-

---

[48] The number of bins in the estimation is reduced to 50 to account for the lower sample size in this analysis which is restricted to lottery winners only. The detailed step by step process of calculating these ex-ante propensities of winning at elite schools is explained in Appendix Section B.3.

rized as budget. Table 8 shows this for both the quality measures - the mean school fee for elite schools is about 3.5 times higher than that for budget schools, and this ratio is about 1.8 for the PCA index.[49]

Table 9: LATE of attending elite schools on performance on tests

	English	Math	English	Math
	Elite (PCA)		Elite (Fee)	
	(1)	(2)	(3)	(4)
RTE student at Elite school	0.699*** (0.270)	0.370 (0.267)	0.485** (0.242)	0.138 (0.240)
First stage F-stat	590.47	590.47	389.43	389.43
Outcome mean	0.06	0.04	0.06	0.05
Control mean	0.03	-0.05	-0.18	-0.20
Observations	318	318	303	303
R^2	0.14	0.13	0.20	0.17
Pscores of winning at elite	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on children's performance on phone-based assessments. The sample is restricted to lottery winners. As before, the number of observations is smaller here because the phone-based assessment on English and Math is available only for a subsample of lottery winners. Control variables include sex and age of child, indicators for father's and mother's education being greater than the respective means, indicators of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

6.2.2 Primary outcomes

Table 9 shows the LATE of attending an elite school on children's performance on phone-based assessments, using both measures of school quality. Treated compliers are children who attend elite private schools under the RTE quota because they won the lottery to an elite private school, and control compliers are those who do not attend elite schools under the quota because they lost the lottery to all the elite schools. Since the analysis is restricted to that of winners, control compliers constitute those attending budget private schools as quota students. Hence, the estimates reflect the causal effects as a result of attending elite private schools relative to budget private schools. Results show that relative to budget schools, elite schools increase English test scores on both the quality measures, however, there are no statistically significant gains in Math. As before I explore several mechanisms below.

6.2.3 Mechanisms

My survey data allows me to explore several school inputs such as, school's instructional modality, subjects taught and other school characteristics, and children's time use as potential mechanisms. I discuss these in depth in the subsequent paragraphs.

6.2.3.1 School inputs

While elite schools are no more likely to provide instruction in the two academic years (as measured on the extensive margin), they are however more likely to provide better instruction

---

[49] Appendix Table A16 shows how results vary when the percentile cutoff is changed to the 50th and 90th percentile.

modalities compared to the budget private schools. Table 10 shows that treated compliers are more likely to report receiving synchronous online classes (between 10 and 18 percentage points, based on the PCA and the fee-based measure, respectively), and less likely to receive text-based instruction (by 17 pp, based on the fee-based measure) during the period of remote instruction.

Table 10: LATE of attending elite schools on school instruction

	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
RTE student at Elite school	0.102* (0.060)	0.024 (0.054)	-0.064 (0.077)	0.180*** (0.051)	-0.036 (0.048)	-0.174** (0.069)
First stage F-stat	1,151.81	1,151.81	1,151.81	1,129.44	1,129.44	1,129.44
Outcome mean	0.83	0.13	0.53	0.83	0.13	0.53
Control mean	0.77	0.15	0.59	0.69	0.17	0.60
Observations	1,005	1,005	1,005	959	959	959
R^2	0.09	0.05	0.12	0.18	0.08	0.13
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on school's instruction modality and children's time use in educational activities. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A13 shows the differences in characteristics of elite and budget schools, and helps in understanding key differences across these schools. Controlling for the village fixed effects, elite schools are consistently more likely to have internet, more digital boards per pupil, more likely to be English medium, have a higher proportion of teachers trained in computers, a higher proportion of teachers with Bachelor's in Education degrees, and a higher proportion of general caste category students. The magnitude of differences in Bachelor's in Education degree is substantive, at 20 percentage points, indicating that elite schools are much more likely to hire teachers who have specifically trained to pursue a school teaching career.[50] These patterns suggest that the relative effectiveness of elite schools is likely to be a function of teachers being more qualified and also being more adept at dealing with digital technologies (which matters for remote instruction). In addition, these results are also robust to changing the elite cutoff to the 50^{th} and the 90^{th} percentile of annual fee (Table A19 in Appendix), which provides a consistent story that elite schools were doing better in terms of providing instruction during the period of remote instruction.[51]

Another channel that might explain gains in English could be related to differences in the quality of English instruction, and differences in instructional time spent across subjects, across elite and budget schools. Appendix Table A15 shows that while elite school goers are no more

---

[50] In the Indian context, Bachelors in Education is a degree program that is specifically designed for those who aspire to become school teachers. It is typically a two-year program that one pursues after a three/four year undergraduate degree program, in order to become a school teacher.

[51] I prefer the 75^{th} percentile as it strikes a balance in the tradeoff between treatment contrast in school quality (which is reduced with 50^{th} percentile) and power (power is low at 90^{th} percentile).

likley to be taught the conventional subjects (Math, English, Marathi and Hindi), they are more likely to have other subjects in their curriculum, such as General Knowledge, Arts, Music and Dance. These other subjects are taught in English (as suggested in balance table A13 which shows elite schools being more likely to be English medium) and this in turn is likely to increase children's exposure to English thereby improving their language test scores.

6.2.3.2 Parental investments

Finally, I study impacts on parental investments in children. I find that elite school goers receive lesser help from parents by about 1.5 - 2.3 hours/week (Table 11). There are no detectable effects on parental monetary investments (Table 12). Together, these findings suggest that parents lower their own time inputs as school-quality improves but do not respond on the margin of monetary inputs.

Table 11: LATE of attending elite schools on parental time investments

	Receives help with homework	Hours of help (hrs/week)	Receives help with homework	Hours of help (hrs/week)
	Elite (PCA)		Elite (Fee)	
	(1)	(2)	(3)	(4)
Quota student at Elite school	-0.043 (0.040)	-2.265** (0.964)	0.007 (0.036)	-1.538* (0.877)
First stage F-stat	982.18	982.18	937.84	937.84
Dep mean	0.94	9.75	0.93	9.81
Dep mean (control)	0.93	9.69	0.93	9.45
Observations	1,019	1,019	973	973
R^2	0.05	0.07	0.06	0.08
Pscores of winning at elite	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on parental time investments. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table 12: LATE of attending elite schools on parental time investments

	Any expense (past year)	Expenditure (past year)	Any expense (past year)	Expenditure (past year)
	Elite (PCA)		Elite (Fee)	
	(1)	(2)	(3)	(4)
Quota student at Elite school	0.016 (0.033)	238.903 (433.087)	-0.017 (0.029)	-150.604 (393.986)
First stage F-stat	957.65	957.65	891.68	891.68
Dep mean	0.96	3,414.45	0.96	3,394.49
Dep mean (control)	0.97	3,387.54	0.96	3,189.93
Observations	974	974	931	931
R^2	0.06	0.07	0.05	0.05
Pscores of winning at elite	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on parental monetary investments. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

6.2.3.3 Children's time use

Finally, studying the impacts on children's time use (Table 13), I find that elite schools provide more hours of instruction per week (2.1 - 3.1 hours/week) relative to budget private schools.

Taken together, this evidence suggests some of the plausible mechanisms which might be at play and might matter for children's performance. While the lack of data on instructional time by subject precludes me from testing the role of that channel in explaining the results, the results on differences in school characteristics on teacher quality, digital facilities, overall time spent in school, and likelihood of teaching specific subjects provides an understanding of why elite schools were more effective in providing remote instruction and also enhancing children's learning.

Table 13: LATE of attending elite schools on children's time use

	School (hrs/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)	School (hrs/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
RTE student at Elite school	3.185***	-1.936**	0.007	2.181**	-0.314	-0.140
	(1.176)	(0.957)	(0.113)	(1.066)	(0.869)	(0.103)
First stage F-stat	982.18	982.18	982.18	937.84	937.84	937.84
Outcome mean	13.58	4.43	1.51	13.57	4.49	1.52
Control mean	12.98	4.91	1.51	12.38	5.00	1.48
Observations	1,019	1,019	1,019	973	973	973
R^2	0.10	0.09	0.04	0.11	0.09	0.03
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β_1^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on school's instruction modality and children's time use in educational activities. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

7 External validity, robustness checks and additional exercises

7.1 External Validity

Even though my estimates do not capture private school effectiveness—since the control group includes children attending private schools—the closest comparison in terms of available evidence from in-person settings comes from Muralidharan and Sundararaman (2015), who find gains of 0.19 SD in English (ITT) for private school voucher recipients after 2 years, and Singh (2015), who report effect sizes of similar magnitudes in Andhra Pradesh. Despite the differences, these studies provide the most relevant comparison to my findings, given their focus on estimating the impact of private schools in India. Moreover, my findings are consistent with studies on the impact of private management in schools in other countries, such as Romero, Sandefur and Sandholtz (2020), who observe gains of 0.18 SD in English (ITT) in Liberia. The median effect size across randomized control trials (RCTs) in developing countries is 0.1 SD Evans and Yuan (2022).

For my second set of results examining heterogeneity within the private sector, I benchmark my findings against Andrabi, Bau, Das and Khwaja (2024), who provide the only other existing evidence on within-sector heterogeneity in private schooling. Their analysis of Pakistani schools identifies gains of 0.21 SD within the private sector. While my estimates are larger (0.48-0.69 SD)—potentially attributable to quality differentials—my results align with the takeaway: treating the private school premium as a homogeneous parameter obscures significant heterogeneity, and effect sizes could substantially vary with different identification strategies. This consistency strengthens the external validity of my findings on quality variation within private schools.

7.2 Attrition

Appendix Table A3 shows whether there is selection into participation in phone surveys based on observable characteristics of households at baseline, after conditioning on the ex-ante propensity of winning. The table shows this for household's participation in phone surveys, and for household's participation in phone-based assessments with the applicant child, conditional on being part of the phone surveys. As can be seen from Panel A - winning applicants were 5.7pp more likely to agree to be interviewed relative to the non-winning applicants. However, there is no systematic attrition by winning status, on participation in child assessments, conditional on survey participation. I test for robustness of my results on the main outcome (phone-based assessments), using inverse-probability re-weighting to account for differential attrition and find that the results are robust (Appendix Table A20).

7.3 Using school level values to measure outcomes

My results use survey data from household level reports on children's outcomes. However, some of these variables could be measured with error. For example, certain variables correspond to school-level information, like - does the child's school provide instruction; what is the modality of instruction at child's school; frequency of classes at child's school etc. The ideal scenario would be to obtain administrative data from schools on these school-level variables, however, since such data is not available/collected by schools, I rely on household-level reports for these school-related variables. It is possible that the household-level responses to school-level variables might have measurement error such that there are inconsistencies in how children attending the same exact school might respond to any given question about the school, which in turn might lead to biases in the estimates.[52] I attempt to address this potential noise in the household-level responses, by creating a new variable that captures the school-level unique responses to these questions. I do this by coding the value of the new variable as the response that was most frequently chosen by students attending the same school with the goal that these new variables have less noise.[53]

---

[52] For instance, consider a scenario where a total of five children attend school A, and of these five children, four children respond by reporting yes to the question that asks whether school A was providing instruction in the previous academic year, while one child responds no to this question. This would be problematic in the regression where the outcome measures the binary indicator of school providing instruction.

[53] I clean this by making a new clean variable at the school level based on whether the proportion of children who answer yes to the question at a given school exceeds half. This is under the assumption that if the same response is

Results based on this cleaning of school level variables are shown in Appendix Table A21. I see that the results are still robust - RTE schools are still statisically significantly more likely to provide school instruction in the academic year of 2021-22, they are still more likely to provide synchronous and live classes. The standard errors on the estimates have also shrunk which is a mechanical result due to a decrease in the variation in the new outcome variable. These robustness check strengthen the validity of the main results.

7.4 Varying the ex-ante propensity scores of winning

The identification in my estimates comes from using the within-variation in children who had a similar ex-ante likelihood of winning the RTE lottery but face a randomization in their final lottery outcome. To do this I control for narrow bins of ex-ante propensity scores of winning by simulating the lottery algorithm a large number of times (10,000 simulations). I show that my results are robust to increasing the number of narrow bins of these ex-ante propensities, or in other words reducing the bin width. Reducing the bin-width would lead to stricter within comparisons, comparing children who had a very similar ex-ante likelihood of winning under the lottery. Appendix Figure A7 shows this in a coefficient plot which shows how the LATE coefficient on test scores changes as the bin-width is reduced.

7.5 External validity

My results are based on a lottery-based research design. While lottery-based estimates help in removing selection bias, there are several caveats with this design. First, these estimates are specific to oversubscribed schools, which might be different from undersubscribed schools. For example, oversubscribed schools might be overrepresentatitve of urban areas, relative to rural areas.[54] Second, it relies on applicants who faced lotteries to get admitted to schools, a group that may differ from nonapplicants (Angrist, Cohodes, Dynarski, Pathak and Walters, 2016). Third, the LATE identifies a treatment effect only for compliers which is a very specific sub-population of the treated (Black, Joo, LaLonde, Smith and Taylor, 2022). Nevertheless, Kline, Rose and Walters (2022) show that LATE is the policy-relevant parameter in case of a marginal increase in the number of available seats among lottery applicants (Angrist, Hull and Walters, 2023). Following Angrist et al. (2023), I discuss complier characteristics as they can provide a partial guide to external validity in the context of lottery-based IV estimates.

7.5.1 Characterizing Compliers

The instrumental variables strategy identifies a unique causal parameter, which is specific to the sub-population of compliers for that instrument. Different valid instruments for the same causal relation therefore estimate different things, because the compliers are essentially differ-

---

reported by at least 50% of the students then that is less likely to be incorrect.

[54] Romero and Singh (2024) compare the lottery-based estimates to a random sample of applicants who are always assigned to a private school and find that the lottery-based sample of students is moderately better off than the sample of students with a guaranteed private school allocation. They point out that this might be a function of urban areas being over-represented in their core sample, which have more oversubscribed schools. This has also been observed in charter school lotteries in the US (Cohodes, Setren and Walters, 2021).

ent based on the instrument (Angrist and Pischke, 2009). Since the IV identifies the average treatment effect for the compliers, it is a useful exercise to learn more about the characteristics of the compliers. Another important reason to study complier characteristics is that they can provide insights about external validity of a set of lottery-based IV estimates (Angrist, Hull and Walters, 2023).

I use Angrist et al. (2023)'s implementation of the methods discussed in Abadie (2003), to compute complier characteristics. Table A6 shows the differences in baseline characteristics of the compliers, always- and never-takers in Maharashtra's RTE lottery. The table shows the mean of baseline characteristics for each of these groups (see Appendix Section C.1 for details on implementation). Untreated and treated compliers are very similar across all characteristics as shown in columns (1) and (2). Columns (3) and (4) show the mean characteristics for always- and never-takers. Relative to all other groups, always-takers are slightly more likely to be low income quota applicants, Muslims, and households with mothers having finished primary education. Relative to the other two groups, the average complier is slighlty more likely to be Hindu, and less likely to be from Scheduled castes. However the magnitude of the differences are small indicating that overall group characteristics are quite similar across groups. Overall, this suggests that compliers are representative of the full sample of applicants and that the external validity of the LATE extends to always- and never-takers.

8 Conclusion

This paper provides among the first causal evidence on the impact of India's Right to Education Act (RTE), on students' access to private schools, and test score gains. Using lottery-based variation in private school assignment, I find that RTE beneficiaries experience improvements in enrollment rates and a 0.18 SD improvement in English test scores. Examining counterfactual destinies, I find that while 67.5% of control compliers attend private schools anyway as fee-paying students, only 20% of them attend higher-quality elite private schools, compared to 50% among treatment compliers. This indicates that a large share of test score gains under RTE can be attributed to school quality, rather than simply enabling private school access. Importantly, my paper is among the first to document substantial heterogeneity within the private school sector in India. Specifically, in addition to expanding overall access, especially to higher-quality elite schools, I find that disadvantaged children admitted to higher-quality elite private schools through the RTE policy experience greater gains of 0.48-0.69 SD than those attending lower-quality budget private schools. These heterogeneity findings align with recent evidence from Pakistan (Andrabi, Bau, Das and Khwaja, 2024) that challenge single-parameter approaches to private school effectiveness.

My findings contribute to the literature in three ways: providing among the first rigorous causal evidence on the world's largest affirmative action policy on school entry age children's educational outcomes; documenting significant quality variation within the private sector and estimating private school quality premium using quasi-experimental methods; and demonstrating that higher-quality schools adapt more effectively to educational disruptions.

Several policy implications emerge from my findings. My results suggest that targeting high-quality schools within affirmative action programs could be more effective than approaches that simply enable private school access without considering quality. This implies that, optimal policy design should take into account not only targeting the poorest thereby reducing inframarginality (Romero and Singh, 2024), but also incorporating school quality criteria into the eligibility requirements for participating schools. However, such design must carefully consider general equilibrium effects, including the risk of fee inflation among participating schools (Sahai, 2023) and other strategic behavior by schools to avoid quota obligations. My estimates of test score gains are similar to those observed in prior studies of private school effectiveness in India (Muralidharan and Sundararaman, 2015; Singh, 2015). This in addition to demonstrating external validity of my estimates, also highlights the effectiveness of private schools in remote-learning settings, and demonstrates potential benefits of public-private partnerships in building educational resilience.

These implications extend beyond India. Private schools are growing rapidly in developing countries worldwide. My evidence that school quality varies dramatically within the private sector—and that targeting better schools matters for student learning—is relevant for school choice policies globally. The evidence that disadvantaged students can achieve substantial learning gains when given access to high-quality schools, even during disruptions, suggests that well-designed affirmative action policies can improve educational equity in many different settings.

Future research should address several key questions. Long-term tracking of RTE beneficiaries would reveal whether early learning gains translate into improvement in secondary school completion and learning gains (Angrist et al., 2006), and subsequently labor market outcomes, addressing whether the educational intervention effects persist in the long run. Finally, analysis of spillover effects on non-quota students similar to Muralidharan and Sundararaman (2015) and Rao (2019) would provide further insights on the RTE's broader impact.

References

Abadie, A. (2002), 'Bootstrap tests for distributional treatment effects in instrumental variable models', *Journal of the American statistical Association* **97**(457), 284–292.

Abadie, A. (2003), 'Semiparametric instrumental variable estimation of treatment response models', *Journal of Econometrics* **113**(2), 231–263.

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J. and Pathak, P. A. (2011), 'Accountability and flexibility in public schools: Evidence from Boston's charters and pilots', *The Quarterly Journal of Economics* **126**(2), 699–748.

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y. and Pathak, P. A. (2017), 'Research design meets market design: Using centralized assignment for impact evaluation', *Econometrica* **85**(5), 1373–1432.

Abdulkadiroğlu, A., Angrist, J. and Pathak, P. (2014), 'The elite illusion: Achievement effects at Boston and New York exam schools', *Econometrica* **82**(1), 137–196.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J. and Walters, C. R. (2020), 'Do parents value school effectiveness?', *American Economic Review* **110**(5), 1502–1539.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333.

Alasino, E., Ramırez, M. J., Romero, M., Schady, N. and Uribe, D. (2023), 'Learning losses during the COVID-19 pandemic: Evidence from Mexico'.

Anand, G., Atluri, A., Crawfurd, L., Pugatch, T. and Sheth, K. (2023), 'Improving school management in low and middle income countries: a systematic review'.

Andrabi, T., Bau, N., Das, J., Karachiwalla, N. and Ijaz Khwaja, A. (2024), 'Crowding in private quality: The equilibrium effects of public spending in education', *The Quarterly Journal of Economics* p. qjae014.

Andrabi, T., Bau, N., Das, J. and Khwaja, A. I. (2024), 'Heterogeneity in school value-added and the private premium', *American Economic Review* .

Andrabi, T., Daniels, B. and Das, J. (2021), 'Human capital accumulation and disasters: Evidence from the Pakistan earthquake of 2005', *Journal of Human Resources* pp. 0520–10887R1.

Angrist, J., Bettinger, E., Bloom, E., King, E. and Kremer, M. (2002), 'Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment', *American Economic Review* **92**(5), 1535–1558.

Angrist, J., Bettinger, E. and Kremer, M. (2006), 'Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia', *American Economic Review* **96**(3), 847–862.

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A. and Walters, C. R. (2016), 'Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice', *Journal of Labor Economics* **34**(2), 275–318.

Angrist, J. D., Pathak, P. A. and Walters, C. R. (2013), 'Explaining charter school effectiveness', *American Economic Journal: Applied Economics* **5**(4), 1–27.

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Angrist, J., Hull, P. and Walters, C. R. (2023), 'Methods for measuring school effectiveness', *Handbook of the Economics of Education* .

Angrist, J. and Imbens, G. (1995), 'Identification and estimation of local average treatment effects'.

Angrist, N., Bergman, P. and Matsheng, M. (2020), School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic, Technical report, National Bureau of Economic Research.

Arcidiacono, P. and Lovenheim, M. (2016), 'Affirmative action and the quality–fit trade-off', *Journal of Economic Literature* **54**(1), 3–51.

Azam, M., Chin, A. and Prakash, N. (2013), The returns to English-language skills in India, Technical Report 2.

Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K. and Iqbal, S. A. (2021), 'Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates', *The World Bank Research Observer* **36**(1), 1–40.

Bacher-Hicks, A., Goodman, J. and Mulhern, C. (2021), 'Inequality in household adaptation to schooling shocks: COVID-induced online learning engagement in real time', *Journal of Public Economics* **193**, 104345.

Badaracco, N. (2020), 'Time investment responses of parents and students to school inputs'.

Bagde, S., Epple, D. and Taylor, L. (2016), 'Does affirmative action work? Caste, gender, college quality, and academic success in India', *American Economic Review* **106**(6), 1495–1521.

Bagde, S., Epple, D. and Taylor, L. (2022), 'The emergence of private high schools in India: The impact of public-private competition on public school students', *Journal of Public Economics* **215**, 104749.

Bandiera, O., Buehren, N., Goldstein, M. and Rasul, I. (2020), 'Do school closures during an epidemic have persistent effects? Evidence from Sierra Leone in the time of Ebola'.

Beam, E. A., Mukherjee, P., Navarro-Sola, L., Ferdosh, J. and Sarwar, M. A. H. (2021), 'Take-up, use, and effectiveness of remote technologies', *Endline Report. Innovations for Poverty Action project "Bangladesh COVID-19 Remote Learning Technologies," protocol* **15594**.

Becker, G. S. and Tomes, N. (1976), 'Child endowments and the quantity and quality of children', *Journal of Political Economy* **84**(4, Part 2), S143–S162.

Berry, J. and Mukherjee, P. (2016), 'Pricing of private education in urban India: Demand, use and impact', *Unpublished manuscript. Ithaca, NY: Cornell University* .

Bertrand, M., Hanna, R. and Mullainathan, S. (2010), 'Affirmative action in education: Evidence from engineering college admissions in India', *Journal of Public Economics* **94**(1-2), 16–29.

Black, D. A., Joo, J., LaLonde, R., Smith, J. A. and Taylor, E. J. (2022), 'Simple tests for selection: Learning more from instrumental variables', *Labour Economics* **79**, 102237.

Black, D. A. and Smith, J. A. (2006), 'Estimating the returns to college quality with multiple proxies for quality', *Journal of labor Economics* **24**(3), 701–728.

Black, D., Smith, J. and Daniel, K. (2005), 'College quality and wages in the United States', *German Economic Review* **6**(3), 415–443.

Bleemer, Z. (2022), 'Affirmative action, mismatch, and economic mobility after California's Proposition 209', *The Quarterly Journal of Economics* **137**(1), 115–160.

Bol, T. (2020), 'Inequality in homeschooling during the Corona crisis in the Netherlands. First results from the LISS panel'.

Bonesrønning, H. (2004), 'The determinants of parental effort in education production: do parents respond to changes in class size?', *Economics of Education Review* **23**(1), 1–9.

Bruhn, J. (2019), 'The consequences of sorting for understanding school quality', *Unpublished working paper). Retrieved from https://1b50402b-a-62cb3a1a-s-sites. googlegroups. com/site/jessebruhn3/jesse_bruhn_jmp. pdf* .

Buhl-Wiggers, J., Kerwin, J. T., de la Piedra, R. M., Smith, J. and Thornton, R. (2023), 'Reading for life: Lasting impacts of a literacy intervention in Uganda'.

Card, D. and Krueger, A. B. (2005), 'Would the elimination of affirmative action affect highly qualified minority applicants? evidence from California and Texas', *ILR Review* **58**(3), 416–434.

Carlana, M. and La Ferrara, E. (2021), 'Apart but connected: Online tutoring and student outcomes during the COVID-19 pandemic'.

Chabrier, J., Cohodes, S. and Oreopoulos, P. (2016), 'What can we learn from charter school lotteries?', *Journal of Economic Perspectives* **30**(3), 57–84.

Cohodes, S. R., Setren, E. M. and Walters, C. R. (2021), 'Can successful schools replicate? scaling up Boston's charter school sector', *American Economic Journal: Economic Policy* **13**(1), 138–67.

Crawfurd, L., Evans, D. K., Hares, S. and Sandefur, J. (2023), 'Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone', *Journal of Development Economics* **164**, 103114.

Cullen, J. B., Jacob, B. A. and Levitt, S. (2006), 'The effect of school choice on participants: Evidence from randomized lotteries', *Econometrica* **74**(5), 1191–1230.

Cunha, F., Heckman, J. J. and Schennach, S. M. (2010), 'Estimating the technology of cognitive and noncognitive skill formation', *Econometrica* **78**(3), 883–931.

Damera, V. K. (2018), Essays on school choice, PhD thesis, University of Oxford.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V. (2013), 'School inputs, household substitution, and test scores', *American Economic Journal: Applied Economics* **5**(2), 29–57.

Datar, A. and Mason, B. (2008), 'Do reductions in class size "crowd out" parental investment in education?', *Economics of Education Review* **27**(6), 712–723.

Del Boca, D., Monfardini, C. and Nicoletti, C. (2017), 'Parental and child time investments and the cognitive development of adolescents', *Journal of Labor Economics* **35**(2), 565–608.

Deming, D. J., Hastings, J. S., Kane, T. J. and Staiger, D. O. (2014), 'School choice, school quality, and postsecondary attainment', *American Economic Review* **104**(3), 991–1013.

Dillon, E. W. and Smith, J. A. (2020), 'The consequences of academic match between students and colleges', *Journal of Human Resources* **55**(3), 767–808.

Dobbie, W. and Fryer Jr, R. G. (2011), 'Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem children's zone', *American Economic Journal: Applied Economics* **3**(3), 158–187.

Dobbie, W. and Fryer Jr, R. G. (2013), 'Getting beneath the veil of effective schools: Evidence from New York City', *American Economic Journal: Applied Economics* **5**(4), 28–60.

Dobbie, W., Fryer, R. G. et al. (2011), Exam high schools and academic achievement: Evidence from New York City, Technical report, National Bureau of Economic Research.

Evans, D. K. and Yuan, F. (2022), 'How big are effect sizes in international education studies?', *Educational Evaluation and Policy Analysis* **44**(3), 532–540.

Fredriksson, P., Öckert, B. and Oosterbeek, H. (2016), 'Parental responses to public investments in children: Evidence from a maximum class size rule', *Journal of Human Resources* **51**(4), 832–868.

Gelber, A. and Isen, A. (2013), 'Children's schooling and parents' behavior: Evidence from the Head Start Impact study', *Journal of Public Economics* **101**, 25–38.

Glewwe, P. and Kremer, M. (2006), 'Schools, teachers, and education outcomes in developing countries', *Handbook of the Economics of Education* **2**, 945–1017.

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

Greaves, E., Hussain, I., Rabe, B. and Rasul, I. (2019), Parental responses to information about school quality: Evidence from linked survey and administrative data, Technical report, ISER Working Paper Series.

Greaves, E., Hussain, I., Rabe, B. and Rasul, I. (2023), 'Parental responses to information about school quality: Evidence from linked survey and administrative data', *The Economic Journal* **133**(654), 2334–2402.

Guariso, A. and Björkman Nyqvist, M. (2023), The impact of the COVID-19 pandemic on children's learning and wellbeing: Evidence from India, Technical report, Stockholm School of Economics, Mistra Center for Sustainable Markets (Misum).

Hanushek, E. A. (2003), 'The failure of input-based schooling policies', *The Economic Journal* **113**(485), F64–F98.

Hassan, H., Islam, A., Siddique, A., Wang, L. C. et al. (2021), Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh, Technical report, TUM School of Governance at the Technical University of Munich.

Houtenville, A. J. and Conway, K. S. (2008), 'Parental effort, school resources, and student achievement', *Journal of Human resources* **43**(2), 437–453.

Hsieh, C.-T. and Urquiola, M. (2006), 'The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program', *Journal of public Economics* **90**(8-9), 1477–1503.

Imbens, G. W. and Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Indus Action (2019), The bright spots report: Status of social inclusion through RTE section 12(1)(c), Technical report, Indus Action.

Jack, R., Halloran, C., Okun, J. and Oster, E. (2023), 'Pandemic schooling mode and student test scores: Evidence from US school districts', *American Economic Review: Insights* **5**(2), 173–190.

Jackson, K. (2010), 'Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago', *The Economic Journal* **120**(549), 1399–1429.

Jackson, K., Johnson, R. C. and Persico, C. (2016), 'The effects of school spending on educational and economic outcomes: Evidence from school finance reforms', *The Quarterly Journal of Economics* **131**(1), 157–218.

Khanna, G. (2020), 'Does affirmative action incentivize schooling? Evidence from India', *Review of Economics and Statistics* **102**(2), 219–233.

Khanna, G. (2023), 'Large-scale education reform in general equilibrium: Regression discontinuity evidence from India', *Journal of Political Economy* **131**(2), 549–591.

King, G., Honaker, J., Joseph, A. and Scheve, K. (2001), 'Analyzing incomplete political science data: An alternative algorithm for multiple imputation', *American Political Science Review* **95**(1), 49–69.

Kingdon, G. G. (2007), 'The progress of school education in India', *Oxford Review of Economic Policy* **23**(2), 168–195.

Kingdon, G. G. (2020), 'The private schooling phenomenon in India: A review', *The Journal of Development Studies* **56**(10), 1795–1817.

Kline, P., Rose, E. and Walters, C. (2022), Systemic discrimination among large US employers, Technical Report 4.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E. and Liu, J. (2020), 'Projecting the potential impact of COVID-19 school closures on academic achievement', *Educational Researcher* **49**(8), 549–565.

Lai, F., Sadoulet, E. and de Janvry, A. (2011), 'The contributions of school quality and teacher qualifications to student performance evidence from a natural experiment in Beijing middle schools', *Journal of Human Resources* **46**(1), 123–153.

Landerso, R., Nielsen, H. S. and Simonsen, M. (2017), 'How going to school affects the family', *Department of Economics Aarhus University* .

Moscoviz, L., Evans, D. K. et al. (2022), *Learning loss and student dropouts during the COVID-19 pandemic: A review of the evidence two years after schools shut down*, Center for Global Development.

Muralidharan, K. and Kremer, M. (2006), 'Public and private schools in rural India', *Harvard University, Department of Economics, Cambridge, MA* **9**, 10–11.

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

Patrinos, H. A., Vegas, E. and Carter-Rau, R. (2022), 'An analysis of COVID-19 student learning loss'.

Pop-Eleches, C. and Urquiola, M. (2013), 'Going to a better school: Effects and behavioral responses', *American Economic Review* **103**(4), 1289–1324.

Pratham (2019), Annual status of education report, Technical report, Pratham.

Rabe, B. (2020), 'Schooling inputs and behavioral responses by families', *Handbook of Education Economics: A Comprehensive Overview* pp. chap. 16, 217—-227, 2nd ed.

Rao, G. (2019), 'Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools', *American Economic Review* **109**(3), 774–809.

Romero, M., Sandefur, J. and Sandholtz, W. A. (2020), 'Outsourcing education: Experimental evidence from Liberia', *American Economic Review* **110**(2), 364–400.

Romero, M. and Singh, A. (2024), 'The incidence and effects of affirmative action: Evidence from quotas in private schools in India'.

Sahai, H. (2023), 'School voucher design and strategic pricing: Evidence from India'.

Singh, A. (2015), 'Private school effects in urban and rural India: Panel estimates at primary and secondary school ages', *Journal of Development Economics* **113**, 16–32.

Singh, A., Romero, M. and Muralidharan, K. (2022), COVID-19 learning loss and recovery: Panel data evidence from India, Technical report, National Bureau of Economic Research.

Todd, P. E. and Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *The Economic Journal* **113**(485), F3–F33.

Tooley, J. (2013), *The beautiful tree: A personal journey into how the world's poorest people are educating themselves*, Cato Institute.

Tooley, J. and Dixon, P. (2007), 'Private education for low-income families: Results from a global research project', *Private schooling in less economically developed countries: Asian and African perspectives* pp. 15–39.

Weiland, C., Unterman, R., Dynarski, S., Abenavoli, R., Bloom, H., Braga, B., Faria, A.-M., Greenberg, E., Jacob, B., Lincove, J. A. et al. (2023), 'Lottery-based evaluations of early education programs: Opportunities and challenges for building the next generation of evidence'.

A Appendix: Figures and Tables

Figure A1: Timeline of events

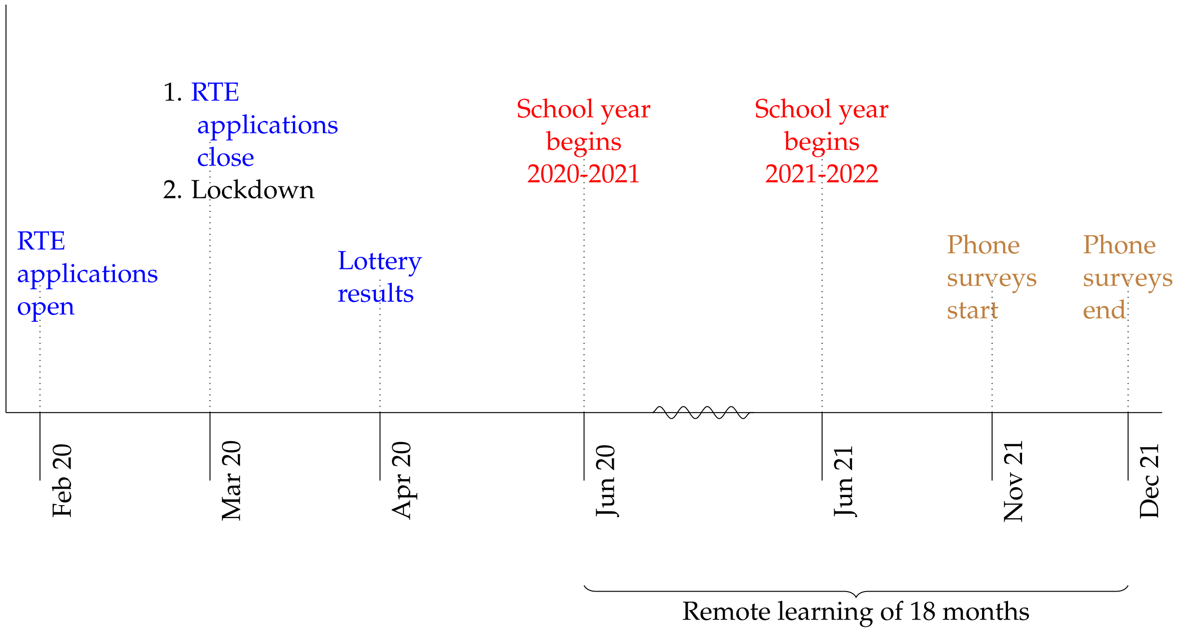
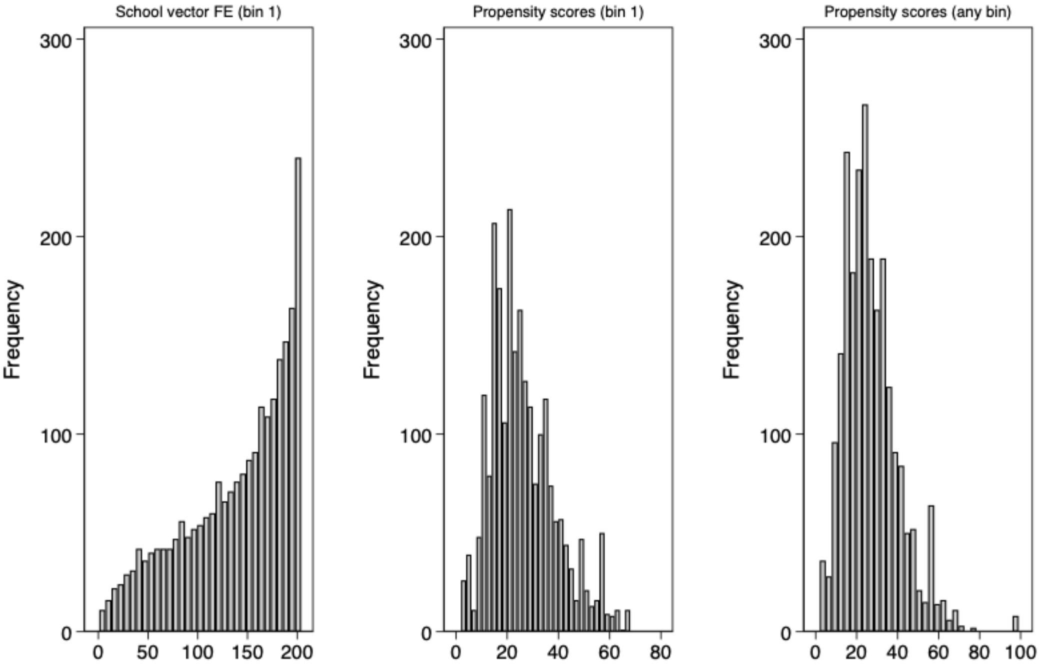
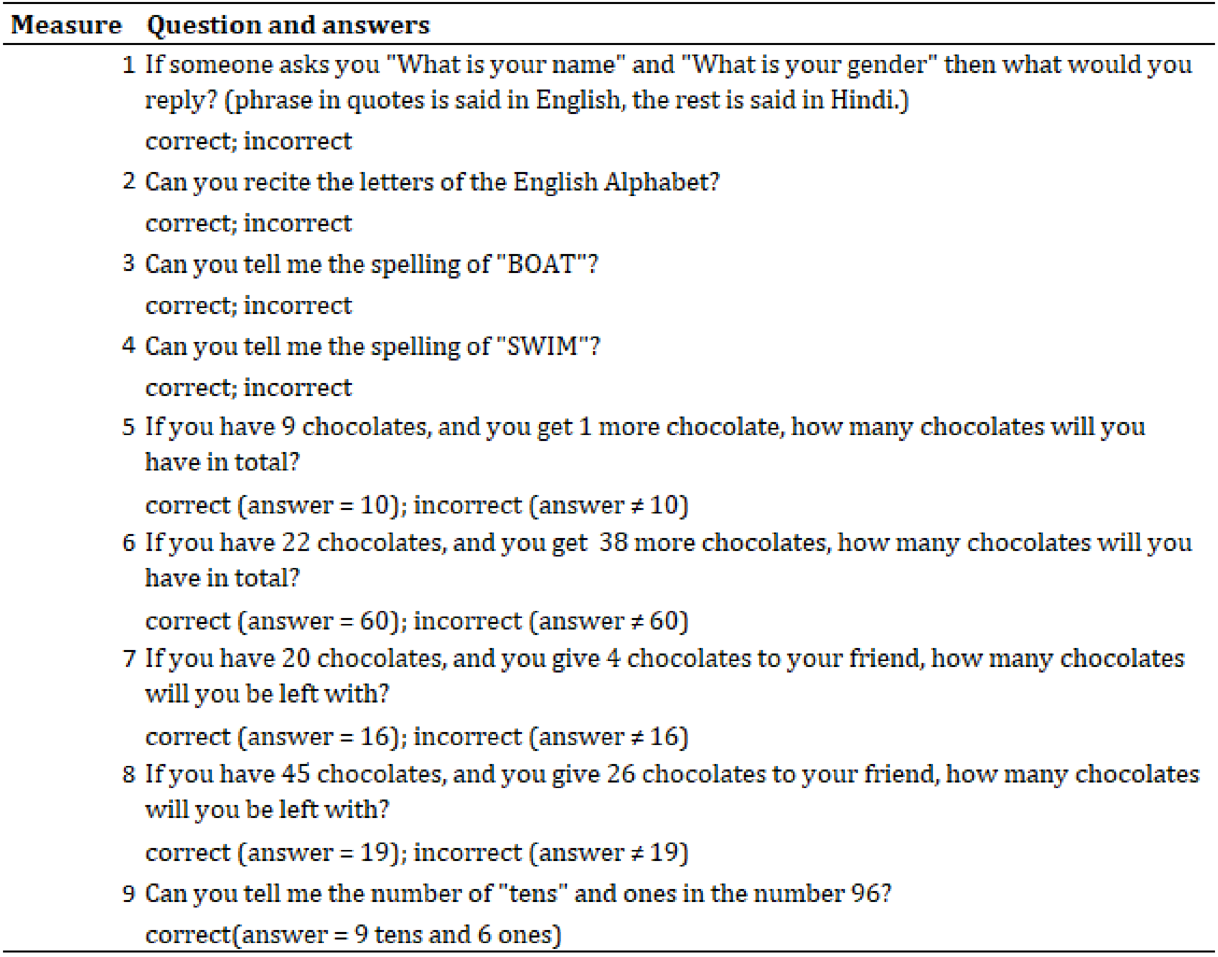


Figure A2: Distribution of school vector fixed effects and ex-ante propensity scores of winning



Notes: This is a histogram showing the distribution of school vector fixed effects (chosen in bin 1), and the simulated ex-ante propensity scores of winning the school lottery (in distance bin 1, and in any distance bin). The sample comprises surveyed applicants. In the sample of surveyed applicants, there are a total of 204 unique school vectors that are chosen in bin 1. A total of 193 school vectors out of these 204 vectors contribute to the identifying within-vector variation, i.e., they have at least one winner and at least one loser within bin 1. Distribution of simulated ex-ante propensity score bins in distance bin 1, and in any distance bin is also plotted. Here the propensity score bins are 0.01 interval wide.

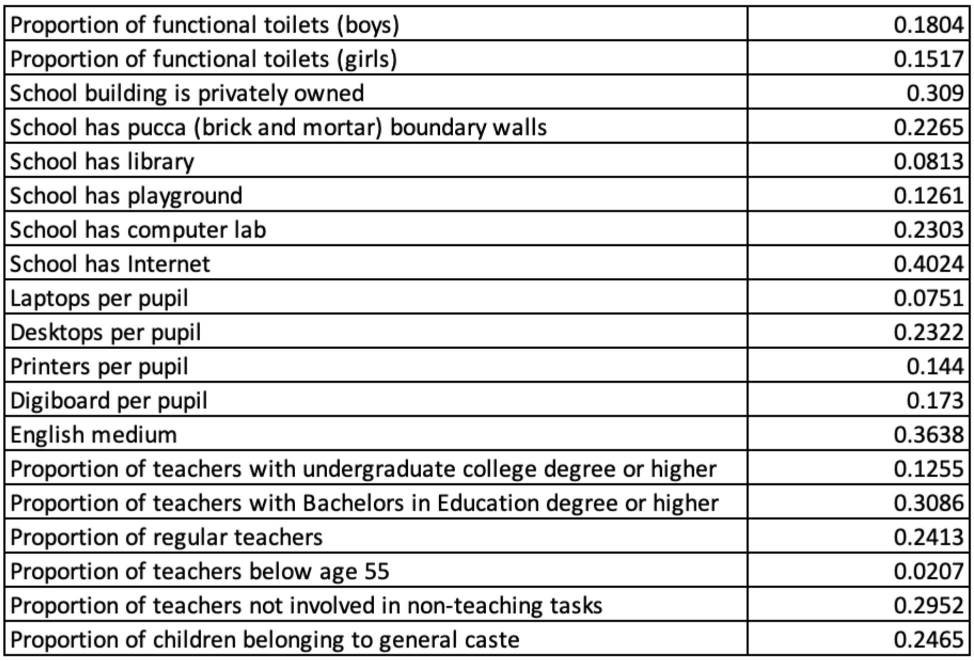
Figure A3: English and Math questions asked during phone based assessments



Measure	Question and answers
1	If someone asks you "What is your name" and "What is your gender" then what would you reply? (phrase in quotes is said in English, the rest is said in Hindi.) correct; incorrect
2	Can you recite the letters of the English Alphabet? correct; incorrect
3	Can you tell me the spelling of "BOAT"? correct; incorrect
4	Can you tell me the spelling of "SWIM"? correct; incorrect
5	If you have 9 chocolates, and you get 1 more chocolate, how many chocolates will you have in total? correct (answer = 10); incorrect (answer ≠ 10)
6	If you have 22 chocolates, and you get 38 more chocolates, how many chocolates will you have in total? correct (answer = 60); incorrect (answer ≠ 60)
7	If you have 20 chocolates, and you give 4 chocolates to your friend, how many chocolates will you be left with? correct (answer = 16); incorrect (answer ≠ 16)
8	If you have 45 chocolates, and you give 26 chocolates to your friend, how many chocolates will you be left with? correct (answer = 19); incorrect (answer ≠ 19)
9	Can you tell me the number of "tens" and ones in the number 96? correct(answer = 9 tens and 6 ones)

Notes: This table shows the list of questions asked to children during phone-based assessments. For all questions, the question was said in Hindi, but the key phrases/numbers were said in English. For example, the following things were said in English - the phrase in quotes "What is your name" and "What is your gender" (for question 1); English Alphabet (for question 2); the words "Boat" and "Swim" (for question 3, 4); numbers like 9 chocolates, 20 chocolates etc. (for questions 5-9).

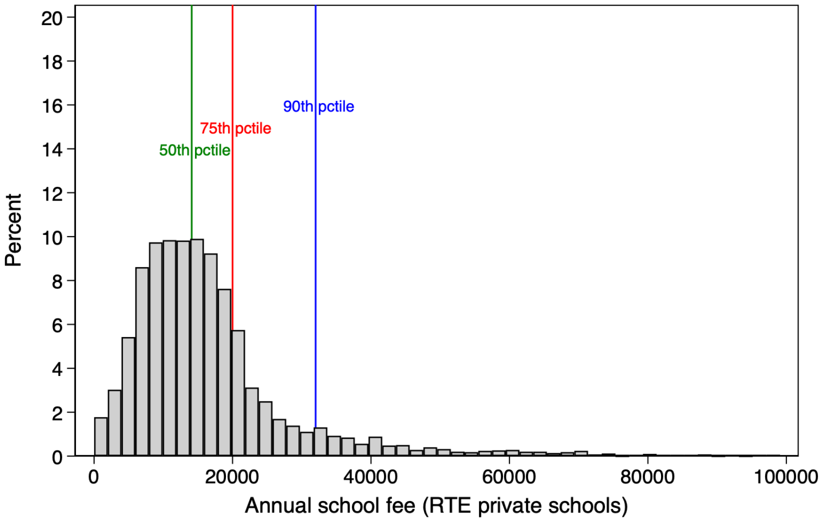
Figure A4: Factor loadings from first component of PCA



Proportion of functional toilets (boys)	0.1804
Proportion of functional toilets (girls)	0.1517
School building is privately owned	0.309
School has pucca (brick and mortar) boundary walls	0.2265
School has library	0.0813
School has playground	0.1261
School has computer lab	0.2303
School has Internet	0.4024
Laptops per pupil	0.0751
Desktops per pupil	0.2322
Printers per pupil	0.144
Digiboard per pupil	0.173
English medium	0.3638
Proportion of teachers with undergraduate college degree or higher	0.1255
Proportion of teachers with Bachelors in Education degree or higher	0.3086
Proportion of regular teachers	0.2413
Proportion of teachers below age 55	0.0207
Proportion of teachers not involved in non-teaching tasks	0.2952
Proportion of children belonging to general caste	0.2465

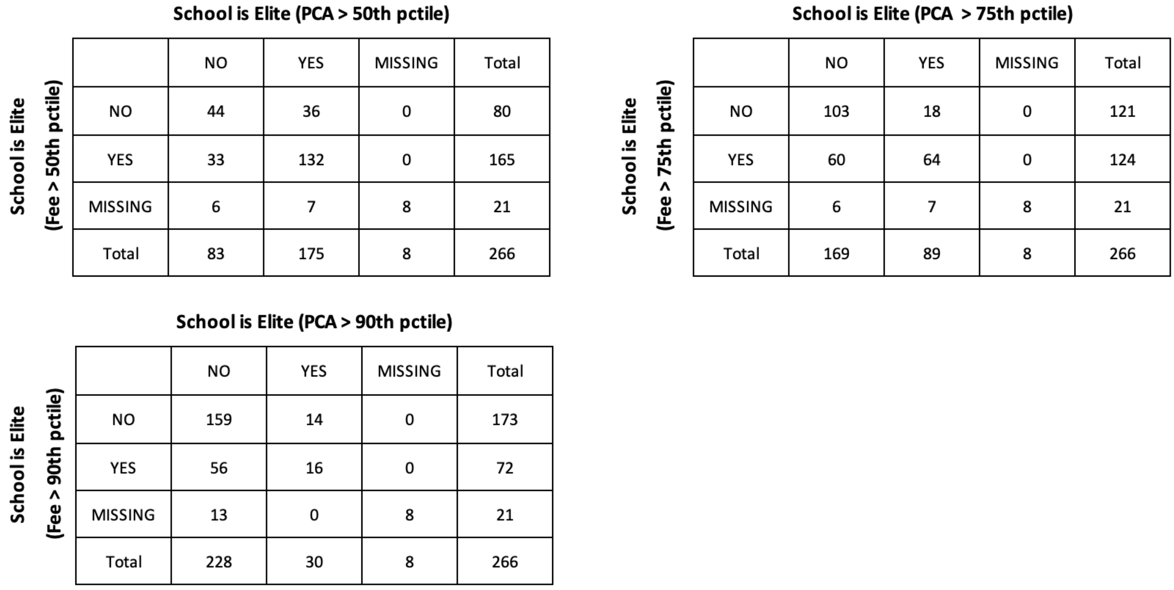
Notes: This shows the factor loadings on each of the variable that is used in the construction of the school quality index using principal component analysis. The first component explains 18% variation in the data.

Figure A5: Distribution of annual school fee for private schools



Notes: This histogram shows the distribution of annual school fee (in INR) for all the private schools in the sample. The data comes from the official website of the State Department of Education, Maharashtra.

Figure A6: Tabulating eliteness across fee and PCA index measure



School is Elite (PCA > 50th pctile)

School is Elite (Fee > 50th pctile)

	NO	YES	MISSING	Total
NO	44	36	0	80
YES	33	132	0	165
MISSING	6	7	8	21
Total	83	175	8	266

School is Elite (PCA > 75th pctile)

School is Elite (Fee > 75th pctile)

	NO	YES	MISSING	Total
NO	103	18	0	121
YES	60	64	0	124
MISSING	6	7	8	21
Total	169	89	8	266

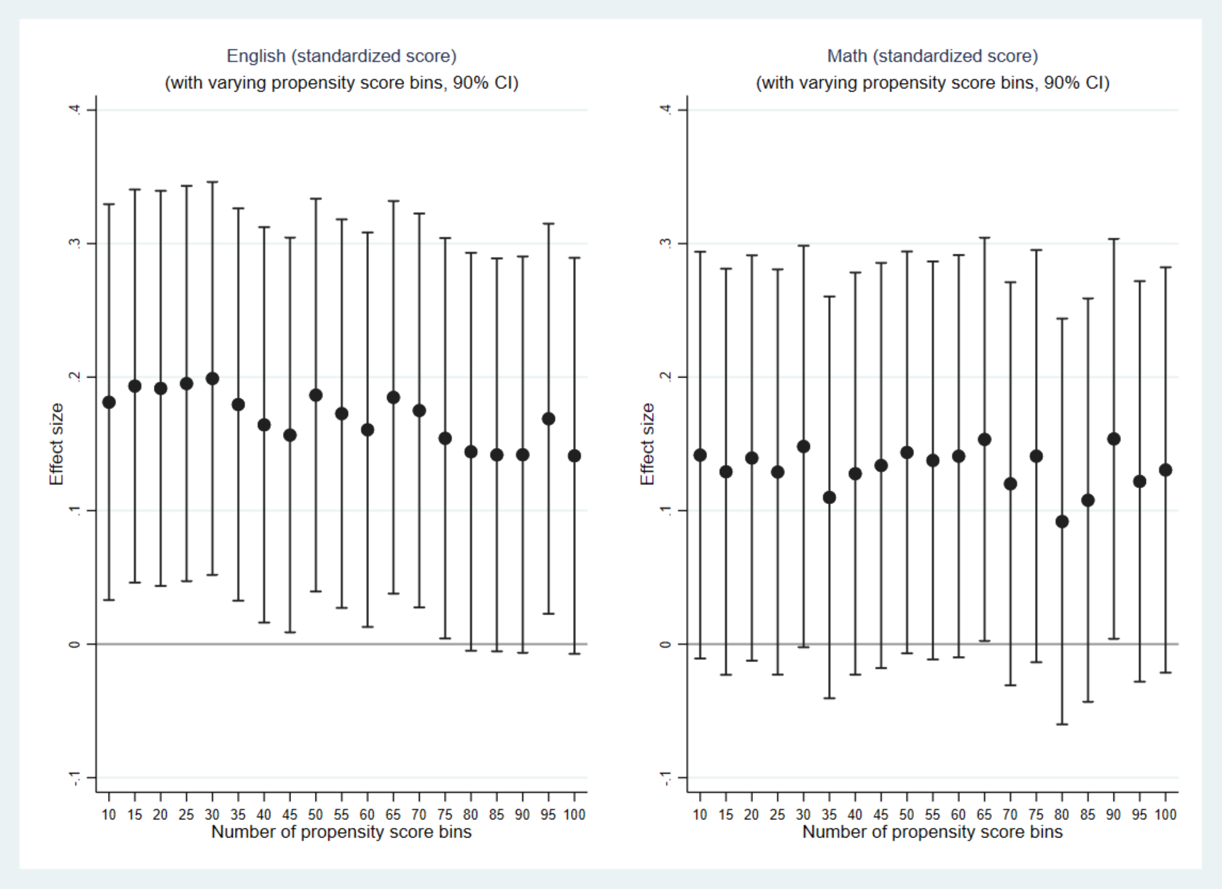
School is Elite (PCA > 90th pctile)

School is Elite (Fee > 90th pctile)

	NO	YES	MISSING	Total
NO	159	14	0	173
YES	56	16	0	72
MISSING	13	0	8	21
Total	228	30	8	266

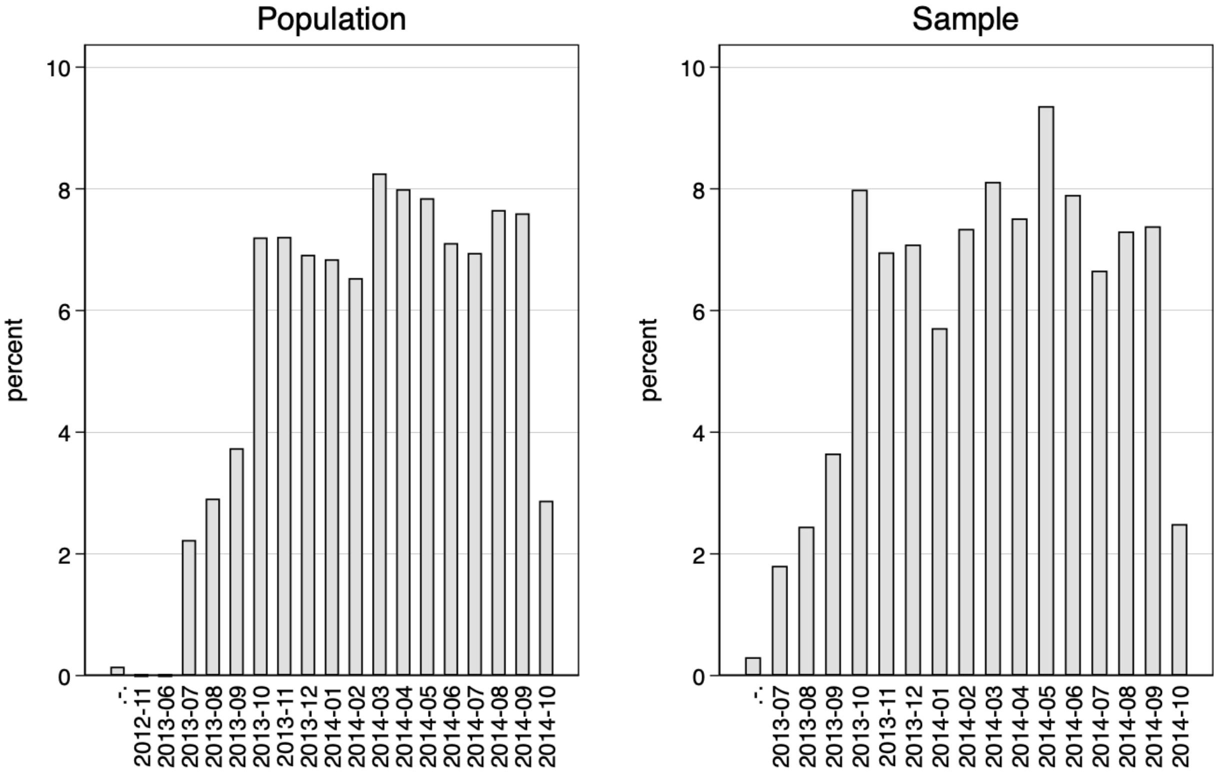
Notes: This provides a cross-tabulation of schools chosen by lottery winners, based on whether the school is categorized as elite or budget as per the PCA index measure and the school fee measure.

Figure A7: Robustness: LATE of being a RTE quota student on phone-based assessments



Notes: This figure plots the LATE of enrolling as an RTE student on children's performance in English and Math. It shows how the LATE changes as the number of bins of ex-ante propensity scores of winning are increased. The within comparisons become stricter as the number of propensity score bins are increased. The number of propensity score bins vary from 10, 15, 20, ..., 100. This utilizes the within variation resulting from comparison of treated and control students who have a similar ex-ante propensity of winning.

Figure A8: Robustness: Histogram of birth year and month



Notes: This figure shows the histogram of birth year-months for applicants to grade 1 private school lotteries under RTE policy in the 2020-21 school year. The left panel shows the distribution for the population and the right panel shows it for the sample. Some missing values exist. Birth year-months given by 2012-11 and 2013-06 are pertaining to only disability quota applicants and only appear in the population histogram. Disability quota is chosen very rarely and constitutes only 0.6% of the applications in the population. My sample does not contain any disability quota applicants. The majority of applications for grade 1 in 2020-21 school year can be seen as coming from those born in July 2013 and October 2014. Among these, applicants born between July 2014 and October 2014 are age-eligible to re-apply for grade 1 in the following year i.e., during the 2021-22 RTE lotteries. In one of my robustness checks, I remove these applicants who were still age-eligible to re-apply for the RTE lotteries in the 2021-22 school year, and find that my results are robust to removing them.

Table A1: Summary statistics

Variable	N	Mean	SD	Min	Max
Characteristics of applicants in Phone Survey					
Winner (distance bin 1)	2,329	0.44	0.50	0	1
Winner (any distance bin)	2,329	0.45	0.50	0	1
Waitlisted (any distance bin)	2,329	0.26	0.44	0	1
Loser	2,329	0.29	0.46	0	1
Age	2,329	7.62	0.33	7.05	8.34
Male	2,329	0.55	0.50	0	1
Number of schools chosen (RTE application)	2,329	4.86	2.89	1	10.00
Applied under low income quota	2,329	0.28	0.45	0	1
Schooling details for applicants					
Academic year: 2020-21					
School enrollment	2,329	0.89	0.31	0	1
School provides instruction	2,083	0.89	0.31	0	1
Academic year: 2021-22					
School enrollment	2,329	0.97	0.18	0	1
School provides instruction	2,255	0.98	0.14	0	1
School is Private	2,255	0.88	0.32	0	1
School is English medium	2,255	0.94	0.25	0	1
Instructional days at school	2,255	5.30	1.42	0	7
Number of subjects taught	2,107	5.93	1.85	1	12
Number of years in preschool	2,301	2.67	0.76	0	5
Monetary investments in applicants					
Any educational expense (in the past year)	2,227	0.93	0.26	0	1
Annual educational expenses (INR; in the past year)	2,227	3,514	3,234	0	24,000
Time investments in applicants					
Child gets help with homework in the household	2,329	0.93	0.26	0	1
Hours of household help with homework (hours per week)	2,329	9.50	5.91	0	49
Time use of applicants					
Attending school (hours per week)	2,329	12	7.98	0	36
Attending tuition (hours per week)	2,329	4.67	6.10	0	21
Doing homework (hours per day)	2,329	1.40	0.73	0	3.30
Playing (hours per day)	2,328	2.45	1.18	0	6
Watching Television (hours per day)	2,322	1.10	0.91	0	4
Helping with household chores (hours per day)	2,329	0.39	0.40	0	2
Performance on phone assessments by applicants					
English score (standardized)	695	-0.00	1.00	-1.56	1.92
Math score (standardized)	695	-0.00	1.00	-1.81	1.91
Parental education					
Mother’s education > primary	2,302	0.96	0.20	0	1
Fathers’s education > primary	2,241	0.95	0.22	0	1
Household characteristics					
Number of household members	2,329	5.14	2.10	2	20
Number of siblings of applicant child	2,329	0.88	0.57	0	5
General Caste	2,329	0.26	0.44	0	1
Scheduled Caste	2,329	0.25	0.43	0	1
Scheduled Tribe	2,329	0.04	0.19	0	1
Other Backward Class (OBC)	2,329	0.46	0.50	0	1
Hindu	2,329	0.81	0.39	0	1
Muslim	2,329	0.09	0.29	0	1
Buddhist	2,329	0.09	0.29	0	1
Other religion	2,329	0.01	0.09	0	1
Household SES index (PCA)	2,329	0.00	1.21	-2.51	6.23
Annual household earnings (INR 1000)	2,001	180	132	2.40	1,200

Notes: This table shows the summary statistics of survey participants who comprise the sample. Most of the data in this table comes from phone-survey data conducted during the months of Nov-Dec 2021 (18 months after RTE results came out). Characteristics of applicants, religion, and caste information comes from the administrative data of RTE applications. Some of the variables are conditional on other variables, such as indicator of whether school provides instruction, and other variables under schooling details, are conditioned on school enrollment. Monetary investments are asked for the past year i.e., 2020-21, and includes expenses on child’s education on stationary, books etc. (excluding school fee). Time investments by parents and household members is calculated by asking about time spent helping child with educational activities on a typical day of the week in the past week (along with number of days). Applicants’ time use is calculated by asking about time spent on each activity on a typical day in the past week, and additionally, number of days per week for variables that measure weekly hours. English and Math scores are standardized - the English assessment had four questions, the Math assessment had five questions. Household SES index is created using Principal Components Analysis using data on asset ownership of television, air conditioner, two-wheeler, and four-wheeler.

Table A2: Balance in baseline characteristics

Variable	(1) Non winners (any bin)	(2) Winners (any bin)	(3) Difference ((2)-(1))
Age of applicant (as on 1st Nov 2021)	7.622	7.611	-0.014
	(0.326)	(0.329)	(0.014)
Male	0.545	0.560	0.012
	(0.498)	(0.497)	(0.021)
Schools chosen overall (RTE application)	4.935	4.759	-0.094
	(2.905)	(2.877)	(0.113)
Applied under low income quota	0.288	0.275	-0.015
	(0.453)	(0.447)	(0.019)
Mother's education > primary	0.954	0.967	0.013
	(0.210)	(0.179)	(0.008)
Father's education > primary	0.945	0.953	0.010
	(0.228)	(0.213)	(0.010)
Number of household members	5.130	5.149	0.028
	(2.119)	(2.078)	(0.089)
Number of siblings of applicant	0.881	0.874	-0.011
	(0.586)	(0.544)	(0.024)
General Caste	0.261	0.251	-0.012
	(0.439)	(0.434)	(0.018)
Scheduled Caste	0.259	0.233	-0.023
	(0.439)	(0.423)	(0.018)
Scheduled Tribe	0.038	0.034	-0.006
	(0.191)	(0.181)	(0.008)
Other Backward Classes	0.442	0.482	0.042**
	(0.497)	(0.500)	(0.021)
Hindu	0.795	0.823	0.030*
	(0.404)	(0.382)	(0.017)
Muslim	0.097	0.086	-0.011
	(0.296)	(0.280)	(0.012)
Buddhist	0.098	0.086	-0.014
	(0.298)	(0.280)	(0.012)
Other religion	0.010	0.006	-0.004
	(0.100)	(0.076)	(0.004)
Household SES index (PCA)	0.057	-0.071	-0.117**
	(1.250)	(1.145)	(0.051)
Observations	1,291	1,038	2,329
	F-stat	**p-value**	
Joint F-test	1.15	0.16	

Notes: This table shows the balance in baseline characteristics across non-winning and winning applicants (in any bin). The differences in column (3) control for the fixed effects of ex-ante propensity of winning the lottery in any bin such that the comparisons across winners and losers are for ex-ante similar applicants. Columns (1) and (2) contain the mean and standard deviation of the variables for non-winners and winners. Column (3) contains the coefficient in front of the dummy of being a winner from the regression of the outcome variable (displayed in the rows) on the dummy of winning, after controlling for the ex-ante propensity of winning in any bin (propensity score bins are 0.01 wide). Column (3) shows standard errors in parentheses.

Table A3: Attrition: participation in the phone survey

Panel A: Sample includes everyone who was ever called for phone surveys

	(1) Participation, Survey = 0	(2) Participation, Survey = 1	(3) Difference ((2)-(1))
Winner (any distance bin)	0.394	0.446	0.057***
	(0.489)	(0.497)	(0.016)
Age of applicant (as on 1st Nov 2021)	7.639	7.617	-0.022**
	(0.335)	(0.328)	(0.011)
Male	0.523	0.547	0.027*
	(0.500)	(0.498)	(0.016)
Schools chosen overall (RTE application)	4.753	4.857	0.055
	(2.928)	(2.893)	(0.065)
Applied under low income quota	0.296	0.282	0.010
	(0.457)	(0.450)	(0.013)
General Caste	0.265	0.257	0.016
	(0.441)	(0.437)	(0.013)
Scheduled Caste	0.250	0.248	0.002
	(0.433)	(0.432)	(0.013)
Scheduled Tribe	0.033	0.036	-0.002
	(0.179)	(0.186)	(0.006)
Other Backward Classes	0.452	0.459	-0.016
	(0.498)	(0.498)	(0.015)
Hindu	0.794	0.807	0.020*
	(0.404)	(0.395)	(0.012)
Muslim	0.102	0.092	-0.008
	(0.303)	(0.289)	(0.008)
Buddhist	0.093	0.093	-0.008
	(0.290)	(0.290)	(0.009)
Other religion	0.011	0.008	-0.003
	(0.104)	(0.090)	(0.003)
Observations	1,930	2,329	4,259

Panel B: Sample includes everyone who agreed to participate in surveys

	(1) Participation, Phone Assessments = 0	(2) Participation, Phone Assessments = 1	(3) Difference ((2)-(1))
Winner (any distance bin)	0.437	0.466	0.026
	(0.496)	(0.499)	(0.024)
Age of applicant (as on 1st Nov 2021)	7.610	7.635	0.029*
	(0.332)	(0.316)	(0.016)
Male	0.575	0.479	-0.099***
	(0.494)	(0.500)	(0.024)
Schools chosen overall (RTE application)	4.765	5.072	0.147
	(2.881)	(2.913)	(0.094)
Applied under low income quota	0.285	0.275	-0.015
	(0.452)	(0.447)	(0.019)
General Caste	0.260	0.249	-0.012
	(0.439)	(0.433)	(0.019)
Scheduled Caste	0.242	0.260	0.002
	(0.429)	(0.439)	(0.020)
Scheduled Tribe	0.035	0.039	-0.001
	(0.184)	(0.193)	(0.009)
Other Backward Classes	0.463	0.452	0.011
	(0.499)	(0.498)	(0.022)
Hindu	0.810	0.800	-0.014
	(0.392)	(0.400)	(0.018)
Muslim	0.091	0.095	0.005
	(0.287)	(0.293)	(0.012)
Buddhist	0.091	0.098	0.011
	(0.287)	(0.297)	(0.013)
Other religion	0.009	0.007	-0.003
	(0.092)	(0.085)	(0.004)
Observations	1,634	695	2,329

Notes: This table shows the balance across survey respondents and non-respondents. The sample comprises all the applicants who were ever called for phone-surveys. Column (2) comprises those who agreed to be interviewed and with whom interviews were successfully conducted and column (1) comprises those who did not agree to be interviewed and with whom interviews were not conducted. Columns (1) and (2) contain the mean and standard deviation of the variables for non-participants and participants. Column (3) contains the coefficient in front of the dummy of participation from the regression of the outcome variable (displayed in the rows) on the dummy of participation. The regressions control for the fixed effects of ex-ante propensity of winning in any bin. Column (3) shows standard errors in parentheses. Winner is a dummy that takes value = 1 for those who won the lottery to a school in any bin. Winning households are slightly more like to participate in the survey relative to the non-winning households.

Table A4: Intent to treat effects of winning the lottery on enrollment

	Enrollment (2020-21)	Enrollment (2021-22)	Grade 2 and above (2021-22)
	(1)	(2)	(3)
Winning the Lottery (any bin)	0.112***	0.038***	0.154***
	(0.013)	(0.007)	(0.014)
Outcome mean	0.89	0.97	0.86
Control mean	0.84	0.94	0.78
Observations	2,328	2,328	2,327
R^2	0.10	0.06	0.13
Pscores of winning	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table reports the intent to treat effects of winning the RTE lottery on children's enrollment. The outcomes in columns (1) and (2) measure the indicator of school enrollment in the two academic years. Column (3) measures the indicator for whether the child is in grade 2 or grade 3 in the 2021-22 academic year. Controls include sex and age of child, dummy of father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A5: Intent to treat effects of winning the lottery on test scores

	Test score (standardized)	
	English	Math
	(1)	(2)
Winning the Lottery (any bin)	0.150**	0.115
	(0.075)	(0.076)
Dependent mean	0.06	0.05
Outcome mean	0.00	-0.00
Control mean	695	695
Observations	0.16	0.12
R^2	Yes	Yes
Pscores of winning (any bin)	Yes	Yes

Notes: This table reports the intent to treat effects of winning the RTE lottery on children's performance on phone-based assessments. Outcomes measure children's standardized test scores on English and Math and are standardized using the control group mean. Controls include sex and age of child, dummy of father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A6: Characteristics of lottery compliers, always- and never-takers in Maharashtra's RTE

	Compliers			
Variable	Untreated (1)	Treated (2)	Always-takers (3)	Never-takers (4)
Male	0.545	0.560	0.561	0.540
	(0.019)	(0.017)	(0.010)	(0.010)
Low income quota applicant	0.295	0.276	0.300	0.248
	(0.017)	(0.015)	(0.009)	(0.009)
Caste quota applicant	0.704	0.723	0.699	0.751
	(0.017)	(0.015)	(0.009)	(0.009)
General caste	0.274	0.258	0.263	0.201
	(0.017)	(0.015)	(0.009)	(0.008)
Scheduled Caste	0.245	0.216	0.299	0.302
	(0.017)	(0.014)	(0.009)	(0.009)
Scheduled tribe	0.043	0.035	0.019	0.016
	(0.007)	(0.006)	(0.002)	(0.002)
Other caste	0.437	0.489	0.417	0.480
	(0.019)	(0.017)	(0.010)	(0.010)
Hindu	0.796	0.833	0.796	0.773
	(0.015)	(0.013)	(0.008)	(0.009)
Muslim	0.096	0.082	0.106	0.094
	(0.011)	(0.009)	(0.006)	(0.006)
Buddhist	0.094	0.077	0.096	0.123
	(0.011)	(0.009)	(0.006)	(0.007)
Mother education > primary	0.951	0.967	0.981	0.958
	(0.008)	(0.005)	(0.002)	(0.004)
Father education > primary	0.945	0.957	0.921	0.949
	(0.009)	(0.007)	(0.005)	(0.004)
Mother works	0.233	0.212	0.230	0.260
	(0.016)	(0.014)	(0.009)	(0.009)
Father works	0.943	0.946	0.949	0.967
	(0.008)	(0.007)	(0.004)	(0.003)
Share of observations	.81		.07	.12

Notes: This table reports the estimates of average baseline characteristics of compliers, always-takers, and never-takers among lottery applicants to private schools under Maharashtra's RTE quotas. Means are computed from 2SLS and OLS regressions that control for lottery risk set indicators (or,ex-ante propensity scores of winning the lottery), as described in Abadie (2003) (see Appendix Section C.1 for details on implementation). Robust standard errors in parentheses.

Table A7: LATE of being a RTE quota student on school quality index

	Joint index	Infrastructure index	Digital index	Teacher index	Peer SES index
	(1)	(2)	(3)	(4)	(5)
Enrolled as RTE student	0.613***	0.329***	0.434***	0.411***	0.219***
	(0.050)	(0.053)	(0.052)	(0.052)	(0.048)
First stage F-stat	3,608.10	3,608.10	3,608.10	3,608.10	3,608.10
Outcome mean	-0.00	0.00	0.00	0.00	-0.00
Control mean	-0.27	-0.15	-0.15	-0.21	-0.11
Observations	2,086	2,086	2,086	2,086	2,086
R^2	0.20	0.10	0.14	0.14	0.28
Pscores of winning	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on school quality, when the instrument is winning the lottery in any bin. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A8: LATE of being a RTE quota student on subjects taught

	Math	English	Marathi	Hindi	Science	Environmental studies
	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled as RTE student	0.017	0.028**	-0.001	0.112***	0.021	0.113***
	(0.014)	(0.014)	(0.019)	(0.022)	(0.020)	(0.026)
First stage F-stat	3,877.71	3,877.71	3,877.71	3,877.71	3,877.71	3,877.71
Outcome mean	0.92	0.93	0.84	0.78	0.18	0.53
Control mean	0.91	0.92	0.84	0.73	0.17	0.48
Observations	2,255	2,255	2,255	2,255	2,255	2,255
R^2	0.07	0.06	0.05	0.09	0.05	0.10
Pscores of winning	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

	Computers	General knowledge	Art/craft	Music	Dance	Physical education
	(7)	(8)	(9)	(10)	(11)	(12)
Enrolled as RTE student	0.136***	0.096***	0.108***	0.067***	0.049***	-0.000
	(0.024)	(0.024)	(0.024)	(0.013)	(0.012)	(0.011)
First stage F-stat	3,877.71	3,877.71	3,877.71	3,877.71	3,877.71	3,877.71
Outcome mean	0.29	0.32	0.31	0.07	0.05	0.95
Control mean	0.23	0.27	0.27	0.04	0.03	0.96
Observations	2,255	2,255	2,255	2,255	2,255	2,255
R^2	0.08	0.08	0.09	0.08	0.08	0.06
Pscores of winning	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on the subjects taught at school when the instrument is winning the lottery in any bin. Outcomes measure the indicator of whether school teaches a particular subject. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A9: Time investments in children by household members: Extensive Margin

	Mother	Father	Grandparents	Siblings	Uncle/Aunt	Neighbors
	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled as RTE student	0.056***	0.026	0.010	-0.014	0.000	-0.005
	(0.020)	(0.023)	(0.006)	(0.016)	(0.010)	(0.004)
First stage F-stat	3,916.37	3,916.37	3,916.37	3,916.37	3,916.37	3,916.37
Outcome mean	0.81	0.28	0.01	0.11	0.04	0.01
Control mean	0.79	0.26	0.01	0.11	0.04	0.01
Observations	2,329	2,329	2,329	2,329	2,329	2,329
R^2	0.12	0.10	0.04	0.05	0.07	0.02
Pscores of winning (any bin)	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending RTE private schools as a quota student on the indicator of whether a specific household member helps the child with educational activities. The outcome variables capture the extensive margin of whether child gets any help from mom, dad, and grandparents. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A10: First stage of winning the RTE lottery (in bin 1) on enrollment as a RTE quota student

	Enrolled as RTE student	
	(1)	(2)
Instrument = Winning lottery in Bin 1	0.790***	0.787***
	(0.013)	(0.013)
Outcome mean	0.44	0.44
Control mean	0.09	0.09
Observations	2,329	2,329
R^2	0.66	0.64
School vector FE (bin 1)	Yes	No
Pscores of winning (bin 1)	No	Yes
Controls	Yes	Yes

Notes: This table shows the first stage effects of winning the RTE private school lottery in distance bin 1, on enrollment as an RTE quota student in a private school. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Column (1) controls for the fixed effects of school vector chosen in bin 1, and Column (2) controls for the ex-ante propensity of winning the lottery in bin 1. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A11: LATE of being a RTE quota student (in bin 1) on enrollment

	Enrollment (2021-22)		Enrollment (2021-22)		Grade 2 and above (2021-22)	
	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled as RTE student	0.130*** (0.015)	0.140*** (0.016)	0.045*** (0.009)	0.047*** (0.009)	0.186*** (0.017)	0.194*** (0.017)
First stage F-stat	3,589.78	3,751.65	3,617.60	3,776.54	3,611.65	3,772.53
Outcome mean	0.89	0.89	0.97	0.97	0.86	0.86
Control mean	0.84	0.84	0.94	0.94	0.78	0.78
Observations	2,328	2,328	2,328	2,328	2,327	2,327
R^2	0.20	0.11	0.13	0.08	0.20	0.15
School vector FE (bin 1)	Yes	No	Yes	No	Yes	No
Pscores of winning (bin 1)	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on children's enrollment when the instrument is winning the lottery in bin 1. The outcomes measure the indicator of school enrollment in the two academic years. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Odd numbered columns control for the fixed effects of school vector chosen in bin 1, and even numbered columns control for the ex-ante propensity of winning the lottery in bin 1. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A12: LATE of being a RTE quota student (in bin 1) on test scores

	Test score (standardized)			
	English		Math	
	(1)	(2)	(3)	(4)
Enrolled as RTE student	0.169** (0.085)	0.178* (0.091)	0.093 (0.090)	0.135 (0.093)
First stage F-stat	947.07	1,194.62	947.07	1,194.62
Outcome mean	-0.00	-0.00	-0.00	-0.00
Control mean	-0.10	-0.10	-0.09	-0.09
Observations	695	695	695	695
R^2	0.41	0.16	0.33	0.11
School vector FE (bin 1)	Yes	No	Yes	No
Pscores of winning (bin 1)	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on children's performance on phone-based assessments when the instrument is winning the lottery in bin 1. Outcomes measure children's standardized test scores on English and Math. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Odd numbered columns control for the fixed effects of school vector chosen in bin 1, and even numbered columns control for the ex-ante propensity of winning the lottery in bin 1. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A13: School characteristics in Elite and Budget schools

	Fee			PCA		
Variable	Budget (1)	Elite (2)	Difference (3)	Budget (4)	Elite (5)	Difference (6)
Proportion functional toilets (boys)	1.000	0.988	0.000	0.994	0.996	-0.000
	(0.000)	(0.101)	(0.000)	(0.079)	(0.030)	(0.000)
Proportion functional toilets (girls)	0.995	0.988	0.012	0.990	0.996	0.009
	(0.059)	(0.101)	(0.015)	(0.095)	(0.030)	(0.015)
School building is privately owned	0.437	0.602	0.172	0.356	0.870	0.464***
	(0.498)	(0.492)	(0.108)	(0.480)	(0.339)	(0.100)
School building has pucca boundary	0.778	0.971	0.009	0.812	0.986	0.142**
	(0.417)	(0.169)	(0.068)	(0.392)	(0.120)	(0.067)
School has library	0.937	0.981	0.016	0.938	1.000	0.100*
	(0.245)	(0.139)	(0.058)	(0.243)	(0.000)	(0.058)
School has playground	0.905	0.981	0.051	0.925	0.971	0.062
	(0.295)	(0.139)	(0.062)	(0.264)	(0.169)	(0.063)
School has computer lab	0.143	0.350	0.125	0.156	0.420	0.265***
	(0.351)	(0.479)	(0.078)	(0.364)	(0.497)	(0.076)
School has internet	0.849	1.000	0.139**	0.881	1.000	0.161**
	(0.359)	(0.000)	(0.065)	(0.325)	(0.000)	(0.065)
Laptops per pupil	0.003	0.005	0.004	0.003	0.005	0.005*
	(0.005)	(0.013)	(0.003)	(0.007)	(0.014)	(0.003)
Desktops per pupil	0.025	0.037	0.007	0.023	0.046	0.025***
	(0.028)	(0.034)	(0.008)	(0.026)	(0.037)	(0.008)
Printers per pupil	0.004	0.005	0.001	0.004	0.005	0.002*
	(0.003)	(0.004)	(0.001)	(0.003)	(0.005)	(0.001)
Digiboards per pupil	0.002	0.008	0.007***	0.002	0.011	0.008***
	(0.005)	(0.011)	(0.002)	(0.005)	(0.012)	(0.002)
School is English medium	0.849	1.000	0.169**	0.881	1.000	0.166**
	(0.359)	(0.000)	(0.071)	(0.325)	(0.000)	(0.072)
Prop. of teachers trained in computer	0.449	0.628	0.255***	0.511	0.571	-0.013
	(0.386)	(0.367)	(0.088)	(0.386)	(0.388)	(0.092)
Prop. of teachers who are graduates	0.736	0.859	0.074	0.759	0.867	0.079
	(0.292)	(0.197)	(0.055)	(0.281)	(0.186)	(0.055)
Prop. of teachers with Bachelors in Education	0.421	0.682	0.200***	0.462	0.716	0.213***
	(0.282)	(0.192)	(0.057)	(0.281)	(0.167)	(0.058)
Prop. of full time teachers	0.673	0.759	0.005	0.725	0.680	-0.157*
	(0.417)	(0.362)	(0.095)	(0.394)	(0.397)	(0.094)
Prop. of contract teachers	0.316	0.232	0.011	0.268	0.303	0.149
	(0.415)	(0.364)	(0.093)	(0.392)	(0.403)	(0.093)
Prop. of part-time teachers	0.011	0.009	-0.016	0.007	0.017	0.008
	(0.065)	(0.027)	(0.015)	(0.047)	(0.060)	(0.015)
Prop. of teachers < 55 years	0.953	0.965	0.009	0.951	0.975	0.049**
	(0.107)	(0.068)	(0.021)	(0.106)	(0.038)	(0.020)
Prop. of teachers not involved in non-teaching tasks	0.859	0.919	-0.015	0.855	0.959	0.068
	(0.295)	(0.234)	(0.059)	(0.302)	(0.159)	(0.060)
Teachers per pupil	0.037	0.040	0.010	0.037	0.042	0.010
	(0.029)	(0.019)	(0.006)	(0.027)	(0.021)	(0.006)
Prop. of general caste category students	0.273	0.527	0.097***	0.334	0.512	0.075***
	(0.283)	(0.271)	(0.025)	(0.296)	(0.290)	(0.026)
Observations	126	103	229	160	69	229

Notes: This table shows the balance in school characteristics for elite and budget schools, where eliteness is defined using the two measures: school fee and PCA index. Schools lying above the 75^{th} percentile value in the distribution of fee and PCA index of all the private schools in the state are classified as elite schools, and classified as budget, otherwise. The sample comprises schools being attended by lottery winners. Columns (1), (2), (3), and (4) show the mean and standard-deviations of the characteristics for budget and elite schools based on the two quality measures. Columns (3) and (4) contain the coefficient on the indicator of "school is elite" from the regression of the outcome variable (dispalyed in rows) on the indicator of school being elite, after controlling for the geography fixed effects at the village level (standard errors in parentheses).

Table A14: Correlation between fee-based school eliteness and PCA-based school eliteness

	School Fee (log)
	(1)
School quality index (standardized)	0.272***
	(0.015)
Outcome mean	9.51
Observations	4,019
R^2	0.51
Village FE	Yes

Notes: This table shows the regression of the log of school fee on the school quality index on the population of RTE schools for whom there is non-missing data on school fee.

Table A15: LATE of attending an elite schools as a RTE quota student on subjects taught

	Math	English	Marathi	Hindi	Science	Envt studies	Comp- uters	General knowledge	Art/ craft	Music	Dance	Phys ed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Elite based on Fee												
RTE student	0.014	0.029	-0.043	-0.020	-0.111**	0.033	0.054	0.241***	0.145**	0.138***	0.085**	-0.013
Elite school	(0.036)	(0.033)	(0.051)	(0.054)	(0.055)	(0.070)	(0.067)	(0.067)	(0.069)	(0.041)	(0.036)	(0.032)
F-stat	1,090	1,090	1,090	1,090	1,090	1,090	1,090	1,090	1,090	1,090	1,090	1,090
Outcome mean	0.93	0.94	0.85	0.83	0.18	0.59	0.35	0.36	0.36	0.10	0.08	0.95
Control mean	0.92	0.93	0.84	0.79	0.18	0.52	0.26	0.28	0.32	0.03	0.02	0.95
Observations	965	965	965	965	965	965	965	965	965	965	965	965
R^2	0.08	0.08	0.08	0.08	0.07	0.08	0.12	0.12	0.06	0.17	0.16	0.10
Pscores	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel B: Elite based on PCA												
RTE student	0.039	0.025	0.072	0.015	-0.057	0.153*	0.117	0.182**	0.041	0.121**	0.126***	0.004
Elite school	(0.042)	(0.038)	(0.058)	(0.061)	(0.063)	(0.078)	(0.076)	(0.076)	(0.077)	(0.047)	(0.042)	(0.035)
F-stat	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099
Outcome mean	0.93	0.94	0.84	0.82	0.19	0.58	0.35	0.36	0.36	0.10	0.07	0.95
Control mean	0.92	0.93	0.85	0.80	0.20	0.53	0.31	0.33	0.36	0.07	0.06	0.96
Observations	1,011	1,011	1,011	1,011	1,011	1,011	1,011	1,011	1,011	1,011	1,011	1,011
R^2	0.07	0.08	0.07	0.07	0.05	0.08	0.08	0.09	0.07	0.09	0.05	0.10
Pscores	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on the subjects taught at child's school. The sample is restricted to lottery winners. Envt studies refers to Environment studies, and Phy Ed refers to Physical Education. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A16: First stage (varying the elite cutoff)

	RTE student at Elite school	
	Elite (PCA)	Elite (Fee)
	(1)	(2)
Panel A: Eliteness defined at 50^{th} pctile		
Won RTE lottery at Elite school	0.878***	0.881***
	(0.032)	(0.031)
Outcome mean	0.67	0.65
Control mean	0.00	0.00
Observations	1,019	973
R^2	0.70	0.72
Pscores of winning at elite	Yes	Yes
Controls	Yes	Yes
Avg quality (Elite=1)	3.79	37143.46
Avg quality (Elite=0)	1.56	10292.29

	RTE student at Elite school	
	Elite (PCA)	Elite (Fee)
	(1)	(2)
Panel B: Eliteness defined at 90^{th} pctile		
Won RTE lottery at Elite school	0.865***	0.850***
	(0.022)	(0.026)
Outcome mean	0.10	0.31
Control mean	0.00	0.00
Observations	1,019	973
R^2	0.89	0.86
Pscores of winning at elite	Yes	Yes
Controls	Yes	Yes
Avg quality (Elite=1)	4.96	54733.34
Avg quality (Elite=0)	2.71	16090.5

Notes: This table shows the first stage effects of winning the RTE private school lottery at an elite school on enrollment at an elite school as a quota student. Here, I present the results with two different percentile cutoffs of eliteness - at 50^{th} and 90^{th} percentile in panel A and B, respectively. The sample is restricted to lottery winners. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A17: LATE of attending elite schools on performance on tests (varying the elite cutoff)

	English	Math	English	Math
	Elite (PCA)		Elite (Fee)	
	(1)	(2)	(3)	(4)
Panel A: Eliteness defined at 50^{th} pctile				
RTE student at Elite school	0.083	0.256	0.189	0.241
	(0.252)	(0.250)	(0.217)	(0.220)
First stage F-stat	229.44	229.44	333.36	333.36
Outcome mean	0.04	0.06	0.05	0.06
Control mean	-0.12	-0.08	-0.18	-0.19
Observations	318	318	303	303
R^2	0.12	0.15	0.16	0.18
Pscores of winning at elite	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

	English	Math	English	Math
	Elite (PCA)		Elite (Fee)	
	(1)	(2)	(3)	(4)
Panel B: Eliteness defined at 90^{th} pctile				
RTE student at Elite school	0.219	0.510	0.024	-0.094
	(0.421)	(0.422)	(0.308)	(0.309)
First stage F-stat	1,821.94	1,821.94	567.59	567.59
Outcome mean	0.04	0.06	0.05	0.06
Control mean	0.02	0.02	-0.07	-0.02
Observations	318	318	303	303
R^2	0.11	0.12	0.14	0.17
Pscores of winning at elite	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on children's performance on phone based assessments. The results correspond to the 50^{th} and 90^{th} percentile cutoffs of eliteness in panel A and B, respectively. The sample is restricted to lottery winners. As before, the number of observations is smaller here because the phone-based assessment on English and Math is available only for a subsample of lottery winners. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A18: LATE of attending elite schools on school instruction (varying the elite cutoff)

	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Eliteness defined at 50^{th} pctile						
Quota student at Elite school	0.195***	-0.021	-0.162**	0.244***	-0.110**	-0.194***
	(0.050)	(0.046)	(0.068)	(0.046)	(0.045)	(0.066)
First stage F-stat	788.37	788.37	788.37	864.10	864.10	864.10
Outcome mean	0.83	0.13	0.53	0.83	0.13	0.53
Control mean	0.73	0.13	0.55	0.65	0.18	0.60
Observations	1,005	1,005	1,005	959	959	959
R^2	0.16	0.07	0.09	0.23	0.07	0.11
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)	Synchronous classes (online)	Recordings shared (audio/video)	Text-based activity plans (WhatsApp/SMS)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Eliteness defined at 90^{th} pctile						
Quota student at Elite school	0.108	-0.005	0.257*	0.196***	-0.018	-0.128
	(0.103)	(0.092)	(0.139)	(0.063)	(0.058)	(0.083)
First stage F-stat	1,453.58	1,453.58	1,453.58	1,326.33	1,326.33	1,326.33
Outcome mean	0.83	0.13	0.53	0.83	0.13	0.53
Control mean	0.81	0.13	0.52	0.77	0.15	0.58
Observations	1,005	1,005	1,005	959	959	959
R^2	0.08	0.04	0.04	0.11	0.06	0.11
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on school's instruction modality. The results correspond to the 50^{th} and 90^{th} percentile cutoffs of eliteness in panel A and B, respectively. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A19: LATE of attending elite schools on children's time use (varying the elite cutoff)

	School (hrs/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)	School (hrs/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Eliteness defined at 50^{th} pctile						
Quota student at Elite school	2.874***	-0.405	0.091	3.272***	-0.825	-0.006
	(0.985)	(0.815)	(0.095)	(0.988)	(0.810)	(0.095)
First stage F-stat	725.95	725.95	725.95	750.79	750.79	750.79
Outcome mean	13.58	4.43	1.51	13.57	4.49	1.52
Control mean	12.60	4.30	1.45	12.18	4.49	1.45
Observations	1,019	1,019	1,019	973	973	973
R^2	0.13	0.09	0.06	0.13	0.10	0.06
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

	School (hrs/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)	School (hours/week)	Tuition (after school) (hrs/week)	Homework (hrs/day)
	Elite (PCA)			Elite (Fee)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Eliteness defined at 90^{th} pctile						
Quota student at Elite school	2.885	-2.674	0.312	0.528	0.040	-0.290**
	(2.053)	(1.650)	(0.194)	(1.304)	(1.047)	(0.124)
First stage F-stat	1,331.97	1,331.97	1,331.97	985.43	985.43	985.43
Outcome mean	13.58	4.43	1.51	13.57	4.49	1.52
Control mean	13.38	4.58	1.50	12.86	4.58	1.51
Observations	1,019	1,019	1,019	973	973	973
R^2	0.07	0.08	0.04	0.08	0.09	0.03
Pscores of winning at elite	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending elite RTE private schools as a quota student, on children's time use. The results correspond to the 50^{th} and 90^{th} percentile cutoffs of eliteness in panel A and B, respectively. Control variables include - sex and age of child, indicators for father's and mother's education being greater than the respective means, indicator of low income quota applicant, household's SES index, indicator of caste categories, and religion. Simulated ex-ante propensity scores of winning the lottery in any distance bin are controlled. Results are robust to increasing the number of propensity score bins. Robust standard errors in parentheses.

Table A20: Robustness to Attrition using Inverse-Probability Reweighting

	English	Math
	(1)	(2)
Enrolled as RTE student	0.332***	0.293**
	(0.126)	(0.119)
Total observations	695	695
Treatment observations	324	324
Control observations	371	371

Notes: This table shows the results for the LATE of attending private schools as a quota student on children's test scores, by using inverse probability weighting to account for the differential probability of attrition or non-response based on baseline observables.

Table A21: Robustness: LATE of being a quota student on school instruction

	School provides instruction (2021-22)	Synchronous (online) (2021-22)	Recordings shared (audio/video) (2021-22)	Text based activity plans (WhatsApp/SMS) (2021-22)
	(1)	(2)	(3)	(4)
Enrolled as RTE student	0.020***	0.152***	-0.069***	-0.085***
	(0.005)	(0.019)	(0.015)	(0.025)
First stage F-stat	3,562.84	3,557.02	3,557.02	3,557.02
Outcome mean	0.99	0.80	0.10	0.57
Control mean	0.98	0.75	0.13	0.59
Observations	2,255	2,238	2,238	2,238
R^2	0.05	0.18	0.10	0.11
Pscores of winning	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated coefficient β^{LATE} from the 2SLS regression that estimates the local average treatment effect of attending a private school as a quota student on school instruction and school modality. This table uses a new variable which is generated such that it is unique at the school level, and captures a unique response to each school being attended in the sample. To do this, I recode the new variable equal to the value that is reported by the majority of the applicants (at least 50%) attending that school. For example, if more than 50% children attending school A say that school was providing instruction, then I code the variable to reflect that school A was providing instruction (for each child who is enrolled at that school, regardless of their original response). Column (1) looks at the dummy of whether school provides any instruction in the 2020-21 academic year, and columns (2), (3), and (4) look at the instructional modality offered by school. Thus, the outcome here is recoded such that there is a unique value associated with each school. Controls include - sex and age of child, indicators for father's and mother's education being greater than the mean, dummy of low income quota applicant, SES index, dummies of caste categories, and religion. Robust standard errors in parentheses.

B Appendix: Lottery Algorithm, Sampling, and Simulation of Algorithm

B.1 Lottery algorithm

Here I explain the lottery algorithm that is implemented for RTE 25% quotas in the state of Maharashtra.

Part 1: Direct offer of admission to winners

1. Schools are arranged in the descending order of total applications received under the policy in the *previous* year. Based on this rank ordering, each school gets their turn to do the allotment of students in the *current* year.

2. There are three rounds in which the allotment happens. Each round corresponds to one of the three distance bins in which schools receive applications.

   Round 1- schools that receive applications in bin 1.

3. The first round comprises each school that received non-zero applications from students who applied to the school in distance bin 1 and allotment is done only for students who applied to these schools in bin 1.

4. The top school (as determined by the rank ordering of schools) allocates seats by lottery if the count of applications received in bin 1 $>$ seats at the school. The school allocates seats to all bin 1 applicants without any lottery if the count of applications received in bin 1 $\leq$ seats. Within the bin, all applicants are treated equally and thus have the same ex-ante probability of being selected in the lottery.[55] All the applicants who are matched to this school are removed from the consideration set and only unmatched applicants are considered for further matching. The school is removed from further matching if it has exhausted all its vacancies.[56]

5. Revised bin-level demand is calculated for all the remaining schools. The previous step is repeated for the next school based on the rank ordering list of schools. The school conducts a lottery based admission if the revised demand by bin 1 applicants exceeds the number of vacancies at school. This process is iterated over all the schools, while maintaining the same initial rank ordering.

6. After the end of round 1, all applicants have been considered at all their bin 1 school choices and all schools have tried to allot any available seats by offering them to their respective bin 1 applicants.

   Round 2- schools (with vacancies) that have applications from unalloted applicants in bin 2

---

[55] This mechanism satisfies the Equal Treatment of Equals (ETE) property following Abdulkadiroğlu, Angrist, Narita and Pathak (2017). ETE is said to satisfy when students with the same preferences and priorities have the same chance of getting allocated at any given school.

[56] If a school conducts a lottery to admit children in round 1 (i.e., for those who applied in the nearest distance bin), then this means that the school will not admit students who applied in the other two distance bins.

7. Next is the second round. The second round comprises schools which have non-zero vacancies and have non-zero applications from those who applied here in bin 2, based on revised bin-level demand at the end of round 1. In this round, allotment is only done for applicants who (i) failed to get a seat in round 1 and had applied somewhere in bin 2, and (ii) applicants who only applied to bin 2 schools.
8. The allotment process is same as before. The top school (based on the same initial rank ordering of schools) allots seats by lottery if the count of revised applications in bin 2 $>$ seats. School allots seats to everyone who applied here in bin 2 without a lottery if the count of revised applications in bin 2 $\leq$ seats.
9. Revised bin level demand is calculated for all the remaining schools, and the previous step is iterated over all the remaining schools, following the same initial rank ordering of schools.
10. At the end of round 2, all applicants who were remaining to be matched after round 1 and were bin 2 applicants somewhere, plus applicants who only applied to schools in bin 2 have been considered at all their bin 2 school choices, conditional on the fact that these school still had seats to offer.

    Round 3- schools (with vacancies) that have applications from unalloted applicants in bin 3

11. Next is the third round. The idea is same as before. Schools which feature here are those that still have vacancies after rounds 1 and 2. Hence round 3 considers applicants who are (i) remaining to be matched after the end of round 2 and had applied somewhere in bin 3, and (ii) applicants who only applied to schools in bin 3.
12. The allotment process is same as before. The top school (based on the same initial rank ordering of schools) allots seats via lottery if the count of revised applications in bin 3 $>$ vacant seats. School allots seats to everyone who applied here in bin 3 without a lottery if the count of revised applications in bin 3 $\leq$ vacant seats.
13. This marks the end of direct offer of admissions to winners.

Part 2: Waitlist determination

Even after the previous steps described in Part 1, there are many applicants who are yet to be matched. These applicants are either waitlisted at a unique school or are rejected from all the schools. There are 3 rounds in which the waitlist determination happens. The process is exactly similar to Part 1 and is explained as follows:

1. Schools are arranged in the same initial rank ordering as before and take turns to do the allotment based on this rank ordering. The rule is that the maximum number of waitlisted students at a school is equal to the number of winners at the school (where the number of winners per school is established in Part 1).

    Round 1- schools that have applications from unalloted applicants in bin 1

2. Round 1 comprises schools which have unmatched applications from those residing in

bin 1 (these are applicants who did not get matched in Part 1).

3. The top school provides offers of waitlist by lottery if the count of unmatched applications in bin 1 $>$ seats available under waitlist. Within the bin, all applicants are treated equally in the event of a lottery. Each matched applicant is assigned a waitlist priority at the school which determines the ordering in which they will be called for admission in the event that any winner at this school forgoes their seat.[57] All matched applicants are removed from the consideration set and the school is removed from any further matching if it has exhausted all its vacancies. Revised bin-level demand is calculated for all remaining schools. This process is iterated for all the remaining schools, following the same initial ranking.

   Round 2- schools (with vacancies) that have applications from unalloted applicants in bin 2

4. Round 2 comprises schools which have unmatched applications from those residing in bin 2 (these are applicants who did not get matched either in Part 1 or round 1 of waitlist). Similar as before, step 3 is iterated at each eligible school, taking into account unmatched applications received in bin 2.

   Round 3- schools (with vacancies) that have applications from unalloted applicants in bin 3

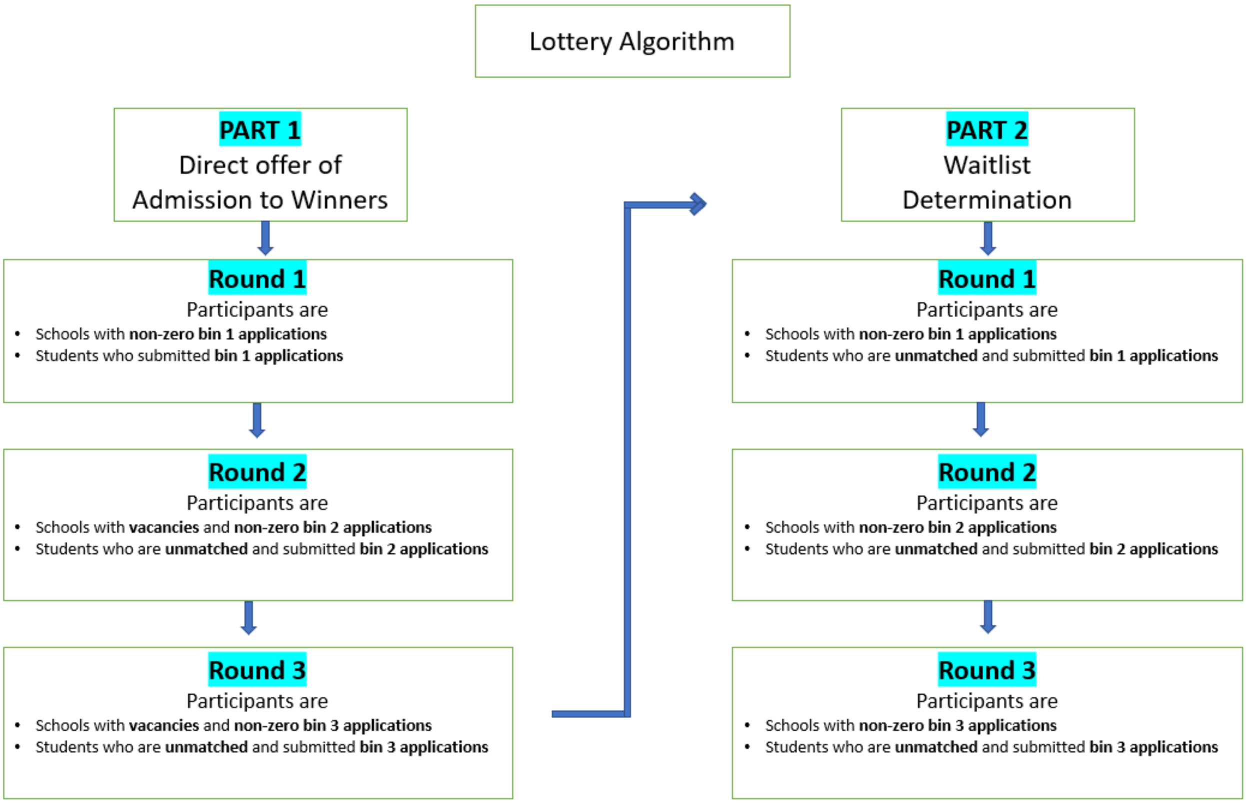
5. Round 3 marks the final round. This comprises schools which have unmatched applications from bin 3 students (these are applicants who did not get matched either in Part 1 or in round 1, and 2 of the waitlist determination). Step 3 is iterated at each eligible school, taking into account unmatched applications received in bin 3.

6. At the end of Round 3, there are still some applicants who are remaining to be matched anywhere. These are the applicants who are not selected anywhere and I refer to them as overall lottery losers.

---

[57] The waitlist priority assigned to applicants at each school is randomly generated.

Figure B1: Schematic flowchart explaining the lottery algorithm



Notes: This flowchart explains the lottery algorithm which the state of Maharashtra uses to allocate schools to applicants under the RTE 25% reservation policy at private schools. The allocation mechanism is a two part process, starting with determining the winners (Part 1, as shown in the left panel), followed by determining the waitlisted candidates (Part 2, as shown in the right panel).

B.2 Sampling strategy

1. I focus on the districts for whom I have the most complete administrative data (Mumbai, Nagpur, Pune, and Thane), and focusing on these districts I make a list of all unique combinations of schools chosen by applicants in the nearest distance bin (henceforth, bin 1). This gives me all the unique *school vectors* that were chosen in bin 1. By virtue of this, some school vectors consist of a single school, and some consist of multiple schools.
2. For each school vector chosen in bin 1, I compute the count of winners who win at any school in the vector (given by the sum of winners at each school in the vector) and count of non-winners who did not win at any school they listed in bin 1.
3. Thus, non-winners for a given school vector comprise applicants who might be: (a) winners at a school that was chosen in distance bin 2 or 3, (b) waitlisted at a school that was chosen in distance bin 1, 2, or 3, and (c) overall losers who lost their chance at each and every school that they listed in each distance bin.[58]
4. Next, I focus on those school vectors which meet the following criteria:
   (i). Count of winners in the vector is at least 4.
   (ii). Count of overall losers in the vector is at least 4.
   (iii). Share of overall losers (among non-winners) in the vector is at least 0.75.

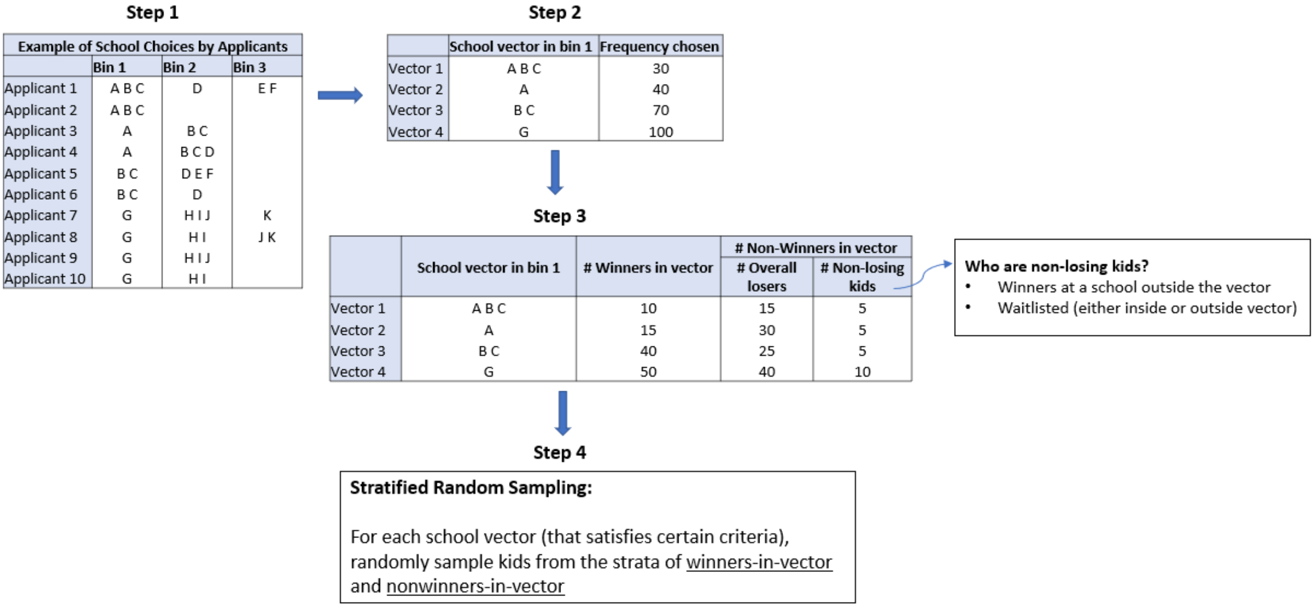
   As an aside - The rule about count of winners and overall losers being at least 4, was imposed taking into account the possibility of low response rates at the time of phone surveys.
5. Finally, from each school vector which satisfies the above three criteria, I perform a stratified random sampling where the two strata are winners and non-winners corresponding to a given school vector chosen in bin 1. Furthermore, the sampling is done such that the count of applicants sampled per school vector = min(winners, non-winners, 25)* 2.[59]

---

[58] This stratification based on the school vector chosen in distance bin 1, satisfies the Equal Treatment of Equals (ETE) property following Abdulkadiroğlu, Angrist, Narita and Pathak (2017). The ETE property is satisfied as all applicants who chose the exact same combination of schools in bin 1 are treated equally at the time of each school's randomization. They are subjected to the same randomization at each school which is listed in the vector, until they get matched at a school. Thus, on average the winners and non-winners who chose the same school vector in bin 1, are comparable to each other.

[59] I restrict the maximum count of applicants per vector in order to maximize the count of unique school vectors in my sample. Based on all these criteria, the minimum number of applicants selected per school vector is equal to 8. Importantly, when the school vector consists of multiple schools chosen in bin 1, I make sure to sample a non-zero count of winning applicants (among winners) from each school in the vector.

Figure B2: Schematic flowchart explaining sampling strategy



Step 1

Example of School Choices by Applicants			
	Bin 1	Bin 2	Bin 3
Applicant 1	A B C	D	E F
Applicant 2	A B C		
Applicant 3	A	B C	
Applicant 4	A	B C D	
Applicant 5	B C	D E F	
Applicant 6	B C	D	
Applicant 7	G	H I J	K
Applicant 8	G	H I	J K
Applicant 9	G	H I J	
Applicant 10	G	H I	

Step 2

	School vector in bin 1	Frequency chosen
Vector 1	A B C	30
Vector 2	A	40
Vector 3	B C	70
Vector 4	G	100

Step 3

	School vector in bin 1	# Winners in vector	# Non-Winners in vector	
			# Overall losers	# Non-losing kids
Vector 1	A B C	10	15	5
Vector 2	A	15	30	5
Vector 3	B C	40	25	5
Vector 4	G	50	40	10

Who are non-losing kids?

- Winners at a school outside the vector
- Waitlisted (either inside or outside vector)

Step 4

Stratified Random Sampling:

For each school vector (that satisfies certain criteria), randomly sample kids from the strata of winners-in-vector and nonwinners-in-vector

B.3 Calculation of ex-ante propensity scores of winning under the lottery mechanism

Below, I explain the step-by-step process for calculating the simulated ex-ante propensity scores of winning under Maharashtra's lottery mechanism for RTE.

1. I conduct a large number of simulations of the lottery mechanism as explained in Section 2.1 (N $\sim$ 10,000).
2. For each simulation, I record the school allotted to each child.

Then for each child, I compute:

1. Simulated ex-ante probability of winning at *each* school that the child listed in application. I do this by averaging across simulations, the probability of winning at that school.
2. Simulated ex-ante probability of winning *in bin 1*. This is given by the sum of simulated ex-ante probability of winning at each school that the child listed in bin 1. The individual simulated probabilities for each chosen school are computed in the previous step.
3. Simulated ex-ante probability of winning *in any bin*. This is given by the sum of simulated ex-ante probability winning at each school that the child listed (combining bin 1, bin 2, bin 3).
4. Simulated ex-ante probability of winning at *elite schools*. I have two measures of elitness - PCA based index and school fee-based measure.
   i. For each child in the sample, and correspondingly for each RTE school that they listed in their application, I make an indicator of whether the school is elite or budget, based on the percentile cutoff. I code the indicator variable = 1 if the school lies above the respective percentile cutoff value, and I code it = 0 if the schools lies below the respective percentile cutoff value. The indicator variable is assiged a missing value in the case where there is missing data on PCA index or fee for the school.
   ii. Next, I compute the simulated ex-ante propensity of winning at elite schools. To do this I simply take the sum of the simulated ex-ante propensities for each school that is coded to be elite based on the respective percentile cutoff.
5. Note that this is always satisfied: simulated probability $\in$ [0,1]
6. Next, I divide these into 100 bins of width = .01 each (for some estimations I reduce the number of bins to 50, in which case the width becomes .02, respectively).
7. Finally, I create dummies of narrow bins of simulated ex-ante propensity scores. In the case where I have 100 bins of propensity scores, this creates 100 dummies of narrow bins: [0,0.01] , (0.01,0.02], ..., (0.99, 1], such that only one of these 100 dummies gets activated for each applicant child.
8. In the estimations I control for dummies of narrow bins of ex-ante propensities of win-

ning, as this facilitates the within-comparison between ex-ante similar applicants who vary in their lottery outcome.

Table B1: Distribution of simulated ex-ante propensity scores of winning

Panel A: Population

Variable	N	10th pctile	25th pctile	50th pctile	75th pctile	90th pctile	95th pctile	99th pctile
NAGPUR								
For winners	6,330	0.20	0.31	0.51	0.80	1	1	1
For waitlisted	5,913	0.03	0.16	0.30	0.45	0.58	0.67	0.81
For losers	9,974	0	0	0.06	0.18	0.27	0.33	0.45
PUNE								
For winners	15,198	0.25	0.42	0.65	0.94	1	1	1
For waitlisted	13,606	0	0.12	0.30	0.47	0.62	0.71	0.86
For losers	13,385	0	0	0.01	0.15	0.26	0.32	0.43
THANE								
For winners	8,041	0.42	0.75	1.00	1	1	1	1
For waitlisted	3,756	0	0.09	0.28	0.47	0.64	0.77	0.93
For losers	1,392	0	0	0.00	0.17	0.27	0.32	0.42
MUMBAI								
For winners	4,721	0.33	0.55	0.94	1	1	1	1
For waitlisted	2,776	0	0.16	0.33	0.48	0.65	0.73	0.89
For losers	1,727	0	0	0.08	0.19	0.27	0.36	0.45

Panel B: Sample

Variable	N	10th pctile	25th pctile	50th pctile	75th pctile	90th pctile	95th pctile	99th pctile
NAGPUR								
For winners	584	0.14	0.20	0.27	0.35	0.48	0.55	0.62
For waitlisted	318	0.17	0.25	0.33	0.41	0.49	0.56	0.62
For losers	396	0.12	0.16	0.22	0.29	0.34	0.38	0.53
PUNE								
For winners	275	0.10	0.15	0.20	0.33	0.52	0.57	0.68
For waitlisted	154	0.12	0.16	0.22	0.38	0.56	0.58	0.68
For losers	228	0.08	0.11	0.16	0.21	0.26	0.30	0.38
THANE								
For winners	134	0.22	0.26	0.30	0.39	0.46	0.71	1
For waitlisted	108	0.22	0.25	0.34	0.40	0.44	0.52	0.64
For losers	43	0.21	0.22	0.28	0.31	0.39	0.44	0.45
MUMBAI								
For winners	45	0.21	0.26	0.36	0.42	0.56	0.64	1
For waitlisted	28	0.25	0.26	0.39	0.53	0.62	0.63	1
For losers	16	0.19	0.20	0.25	0.26	0.27	0.39	0.39

Notes: This table shows the distribution of simulated ex-ante propensity scores of winning under the lottery mechanism.

C Appendix: Estimating Complier Characteristics and Counterfactual Destinies

C.1 Estimation

I follow the Angrist et al. (2023)'s implementation of methods used in Abadie (2002) to compute complier characteristics and counterfactual destinies for untreated compliers. Below I discuss the steps as mentioned in Angrist et al. (2023).

The notation is as follows: $Z_i \in \{0,1\}$ is the instrument which denotes whether i wins the RTE private school lottery. $D_i(1)$ and $D_i(0)$ refer to potential treatments, indicating i's RTE enrollment status as a quota student, when $Z_i = 1$ and $Z_i = 0$, respectively. $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes for individual i as a function of RTE enrollment.

The following assumptions are made:

Assumption 1. Independence/exclusion: $(Y_i(0), Y_i(1), D_i(0), D_i(1)) \perp Z_i$.

Assumption 2. First stage: $\mathbb{E}[D_i|Z_i = 1] > \mathbb{E}[D_i|Z_i = 0]$.

Assumption 3. Monotonicity: $D_i(1) \geq D_i(0)$ ∀ i.

Angrist et al. (2023) explain the process of backing out complier characteristics, which I discuss next. While individual compliers are not coded in any data, complier characteristics can be described using methods of Abadie (2002). The monotonicity assumption implies that the population contributing to the IV analysis only consists of always-takers, never-takers, and compliers. Some of the always and never takers can be identified by the following cells of the data: $D_i = 0$ and $Z_i = 1$ are always-takers while, $D_i = 1$ and $Z_i = 0$ are never-takers. The other cells of the data contain mixtures of compliers with the other two groups: $D_i = 0$ and $Z_i = 0$ contain compliers and never-takers, while $D_i = 1$ and $Z_i = 1$ contain compliers and always-takers. The size of the compliers is given by the first stage. The data also helps in infering the share of never-takers and always-takers as these correspond to the proportion of those who reject the offer of enrollment as a quota student, and the proportion of those who choose to enroll as a quota student when not offered.

Like them, I estimate the following system of equations via 2SLS

$$g(X_i, Y_i) \times 1\{D_i = d\} = \pi_d + \gamma_d 1\{D_i = d\} + v_{id} \tag{5}$$

$$1\{D_i = d\} = \phi_d + \beta_d Z_i + e_{id}, d \in \{0,1\} \tag{6}$$

, where $g(X_i, Y_i)$ is a function of student baseline characteristics (X_i) or post-lottery outcomes (Y_i). Complier characteristics for the treated are obtained by setting $d = 1$ which amounts to using Z_i as the instrument for D_i where the outcome in the second stage is given by $g(X_i, Y_i)$ multiplied by D_i. Similarly setting $d = 0$, estimates the complier characteristics for the untreated which means using Z_i as an instrument for (1-D_i) where the outcome in the second stage is $g(X_i, Y_i)$ multiplied by (1-D_i).

Estimating complier characteristics: Setting $g(X_i, Y_i) = X_i$ yeilds the average complier characteristics for baseline covariates. Estimating equations (5) and (6) as explained in the previous paragraph (along with ex-ante propensities of winning) produces the columns (1) and (2) for Table A6. Column (3) shows always-taker means which are computed by regressing $X_i D_i (1 - Z_i)$ on $D_i(1 - Z_i)$ (with ex-ante propensities), column (4) shows never-taker means which are computed by regressing $X_i D_i Z_i$ on $(1 - D_i)Z_i$ (with ex-ante propensities).

Estimating counterfactual destinies: Table 4 shows the distribution of enrollment across sectors for lottery losers and lottery winners. For example, lottery losers could be enrolled at private schools as fee-paying students, government schools, or remain out-of-school. I first create dummies of enrollment at a particular school sector. Next, I estimate (5) and (6) by setting d=0, for a total of four times (since there are four outside options), each time setting $g(X_i, Y_i)$ as the dummy for enrollment at that specific outisde option.