Essays on School Choice:

Empirical Evidence from Implementation of India's National School Choice Policy

Vijay Kumar Damera

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Public Policy

Hilary 2018

Abstract

This thesis, written in a three-article/chapter format, explores several questions that are at the centre of the theoretical and empirical debates around school choice in developing countries. The implementation of India's national school choice policy provides the context for this inquiry. The policy (hereafter referred to as the 25 percent mandate) sets aside 25 percent of places in private schools for children from disadvantaged backgrounds with government paying the tuition fee to private schools. The empirical analysis in based on three primary datasets and several secondary data sources relating to the applicants to the 25 percent mandate (children aged 7-8 years) in the south Indian state of Karnataka.

In Chapter 1, I evaluate the impact of the 25 percent mandate. Exploiting the lottery based allocation of free places, I investigate whether the learning and psychosocial outcomes of lottery-winning children (policy beneficiaries) improved relative to lottery-losing ones. The study sample of 1,616 children is drawn from a cohort of 7-year-old children who started Grade I in 2015. After 1.5 years of schooling, I find no significant difference in test scores of lottery-winning and losing children on a range of subjects. Policy beneficiary children, however, perform significantly better than non-beneficiaries on one of the four psychosocial measures: self-efficacy. I conduct a detailed mechanism analysis to investigate the non-impact, which reveals that majority of the applicants were default private school goers and would have

attended similar schools irrespective of the policy. This explains the absence of treatment effects on student outcomes. The evidence establishes that the policy has been mistargeted, allowing ineligible households to apply and deterring low-income households from participating. I replicate the mistargeting result on another large representative sample of about 16,000 children drawn from across Karnataka state.

In Chapter 2, I estimate the value added at private schools vis-à-vis public ones and at elite/high-fee private schools vis-à-vis low-fee ones. The experimental setup, sample and data for this analysis is same as that used in Chapter 1. Using an instrumental variables strategy, I find that enrollment at a private school has no statistically significant effect on test scores. However, these estimates are in the range of 0.2-0.6 standard deviations and are not precise due to the relatively small number of government school attending children in the sample. Hence I am not able to comment on the valued added at private schools in learning. Moving beyond test scores, I find large, robust and statistically significant private school effect of 1.90 standard deviation on self-efficacy, one of the four measured psychosocial outcomes. On the impact of elite private schools, I find robust effects only on English test scores (0.58 standard deviation). This is unsurprising given the labor market premiums to English proficiency in the Indian context. However, the lack of test score gains on other subjects is puzzling and suggests the importance of non-academic factors in parental school choices.

In Chapter 3, I present evidence on the demand-side of the RTE 25 percent mandate by evaluating the determinants of parental private school choices and exploring if parental preferences vary by household and child characteristics. The study sample comprises a random set of 636 households with 6-year old children that applied to the RTE mandate in Bangalore, India's third largest city. I estimate a Rank-Ordered Logit (ROL) model that uses data on school attributes, household/ child attributes, and rank-ordered school preferences. I find that home-school distance, school fees, size and infrastructure are the primary determinants of choice. The positive average preference for school fee suggests that fee is a marker for quality in urban India; the heterogeneity across SES on this dimension indicates that the tuition-fee waiver under the RTE mandate is not sufficient to induce poorer households to choose higher-fee schools. The effects on school size, infrastructure and home-school distance are broadly in accordance with existing evidence from other contexts like the US, England and Chile. My results diverge from the extant literature on the heterogeneous preferences for school proximity across household types and on the importance of school composition. I do not find any significant differences in preferences for school proximity by household SES type, and school composition is not a determinant of choice in my setting. This divergence in results can be accounted for by the contextual specificity of the RTE 25 percent mandate and urban India.

The evidence in this thesis contributes primarily to the literature on school choice in developing countries, while also being related to the literature on targeting of social programs and price-quality correlations. From a policy perspective, the results demonstrate that the 25 percent mandate (if correctly targeted) in particular, and school choice in general, has the potential to provide some learning gains to disadvantaged students. It cannot, however, promote large improvements in learning.

Notes for the Reader

This thesis consists of a brief introduction, a review of the literature, three substantive chapters and a conclusion. Chapter 1 has a detailed discussion on India's national school choice policy that provides the context for all the three chapters. Hence a sequential reading of the chapters would be more appropriate though each chapter addresses a separate set of questions. The numbering of sections, figures and tables is specific to each chapter. Chapter-related appendices and references are at the end of each chapter. Notation may vary across chapters.

All the three chapters are solely my work.

An earlier version of Chapter 1 is published as a working paper as

'Damera, V. K. (2017). Choice for the poor or poor choice? Experimental evidence from implementation of India's school choice policy. Working paper, BSG-WP-2017/022, Blavatnik School of Government'

Word count: This thesis has approximately 52,921 words excluding references. This is calculated using the number of words on page 2 (283) multiplied by the number of pages (187, excluding pages with references).

Acknowledgements

To my parents, Laxmi Damera and Veeraiah Damera: I owe you everything. Thank you!

I am grateful to Stefan Dercon, my supervisor, for all the support and for the several critical ideas that shaped this thesis. I am also grateful to Andrew Eggers, my second supervisor, for key inputs at every stage of this work. I will miss our engaging troubleshooting sessions at Nuffield.

Advice from Julien Labonne, Patrick Kuhn, Abhijeet Singh, Ingo Outes-Leon, Natalie Quinn, Ellen Greaves and Maya Tudor was key in developing this work. Special thanks are due to Clare Leaver and Simon Quinn for their detailed feedback during Confirmation of Status.

I am grateful to the many organizations/departments that funded my education at Oxford and the fieldwork for this research: Department of Personnel and Training, Government of India; Blavatnik School of Government (BSG), Center for Study of African Economies, and George Webb Medly/OEP Fund at Oxford; and IDFC Foundation, India.

A very special gratitude goes out to three friends in India. First, Dr. PC Jaffer, IAS, from the education department of Karnataka for his unstinted support at every stage of my fieldwork. Second, Tarun Cherukuri for helping me conceptualize this research. And Gireesh, my fieldwork manager, for his dedication towards this project. Thanks are also in order to my entire field team with a special mention to Phalakshaiah, Nityananda Aardhya, Shiv Kumar and Mary Punnoose. Finally, I should thank Dr. Venita Kaul and her team at the Center for Early Childhood Education and Development (CECED), Dr. B R Ambedhar University, New Delhi for generously allowing me to use and adapt their child testing tools/ instruments.

I wish to acknowledge the encouragement and support of the State Government of Meghalaya, my employer in India. I would like to extend my gratitude to several people at the BSG for providing an intellectually rich and socially warm work environment.

And finally, last but by no means the least, thanks to all my DPhil mates at the BSG and friends at Oxford and around the world who read drafts, gave feedback, challenged my ideas and nurtured this thesis in myriad ways.

Contents

List of Tables

List of Figures

Introduction

Improving student learning and providing cost-effective education are the biggest challenges facing the school education sector in developing countries. Over the last two decades, Governments, aid-agencies and NGOs have intensified efforts at tackling these challenges by deploying myriad reforms/ interventions. Understanding these interventions and gathering evidence on their efficacy has been a central focus for economists and other researchers working on education. Glewwe and Muralidharan (2016) categorize these interventions into four broad groups. First, demand side interventions like information campaigns and conditional cash transfers that focus on households. Second, supply side interventions like improving teacher quality and quantity that seek to provide better inputs at schools. Third, interventions related to pedagogy like remedial teaching that target the “technology of instruction.” And fourth, governance reforms like monitoring, teacher performance pay and school choice that broadly relate to the management of the entire education system.

While most of these reforms concentrate on improving public/government schooling, school choice policies are singular in their focus on the overall education system, encompassing both public and private schools. This reform has been widely tried in the United States (US) and other high-income countries for the last five decades. It involves introducing market mechanisms in schooling and has foundations in economic theory. It has also been a very contentious and polarizing reform, especially in the US, with both enthusiastic supporters and ardent skeptics. The former extol its potential to improve student learning and the overall efficiency of the education system, while the latter view it as an assault on public schooling. The most commonly used choice programs are targeted school vouchers that enable low-income parents to enroll their children at a school of their choice (public or private).

Calls for introduction of choice policies have been rising in the low- and middle-income countries (LMICs), given the increasing size of the private sector and the growing societal perception of the failure of public schools. The rise of the private sector has been particularly strong in India, where about 38 percent of the 200 million elementary school age children (grades 1-8) attend private schools. An important aspect of the private sector is that it serves not just the wealthy but also the middle class and a section of the poor. Private sector enthusiasts posit that private schools are better at promoting student learning and have long argued for introduction of school choice policies like vouchers that enable even the poor to access high quality private sector education. Against this background the Indian government introduced an ambitious and unique school choice mandate as part of the Right To Education (RTE) act of 2009. The mandate sets aside 25 percent of places in all private schools of the country for children from disadvantaged backgrounds; government pays the tuition fee for these places. When fully implemented the mandate would impact 16 million children making it the world's largest program for public funding and private provision in education (Sarin et al., 2016).

This thesis investigates the various facets of the RTE 25 percent mandate and school choice in the context of the south Indian state of Karnataka. Three specific questions are addressed, each in one of the three substantive chapters.

1. What is the impact of the 25 percent mandate on childrens' learning and psychosocial outcomes? Has school choice led to improved student outcomes as predicted by theory?

2. What is the value added at private and elite (private) schools?[1] Are private schools in general, and elite private schools in particular, more effective at improving student outcomes as contended by choice enthusiasts?

3. What are the determinants of parental school choices? Do preferences for schools vary by household and child characteristics?

To answer these questions, I use multiple primary and secondary data sources. I assembled the following three primary datasets on which most the analysis relies.

- A dataset of learning and psychosocial outcomes and household attributes of 1,616 children from four districts of Karnataka (face-to-face survey of parent and children).

- A dataset on school enrollment of 16,000 children from across Karnataka (telephone survey of parent).

- A dataset of family attributes of 636 households from Bangalore, capital of Karanataka (face-to-face survey of parent).

[1] There is no agreed upon definition of elite schools in the literature. I define schools with fee higher than the 75^{th} percentile (p75) of the fee distribution as high-fee/elite.

Additionally, I used secondary data on 25 percent admissions for three academic years, from 2015-18, and data on school characteristics.

In Chapter 1, I present experimental evidence on the impact of the 25 percent mandate on childrens' learning and psychosocial outcomes. I exploit the lottery-based allocation of RTE free places and use a pairwise matching design (Bruhn and McKenzie, 2009) to estimate the causal impact of the policy. The study sample comprises 808 matched treatment-control pairs (1,616 children) drawn from the population of RTE applicants from four districts of Karnataka. The study cohort enrolled in Grade I at the start of the program and were aged 7.3 years, on average, at end line data collection. The results show that after 1.5 years of schooling, there is no difference in test scores of lottery winning and losing children on four subjects- General Cognitive Ability (GCA), Math, English, and Kannada (the local language)- and on the total test score measure. On one of the four psychosocial measures, self- efficacy, lottery winners perform better than lottery losers (effect size: 0.12 standard deviation (σ)).[2]

To understand the non-impact of the program, I conduct a detailed mechanism analysis. I find no qualitative difference in the schools attended by the treatment and the control children, implying that the policy did not shift treated children to better quality schools as envisaged. I demonstrate that the average policy applicant is a default private school goer with ability to pay the median school fee, as opposed to the targeted applicant who was to be a default government school goer. All this evidence leads to the central finding of the chapter that the 25 percent mandate is mistargeted. I replicate this result on a much larger sample of children drawn from across the state

[2] I test children on four psychosocial measures: self-efficacy, peer support, teacher support and school support.

and argue that design flaws in the 25 percent mandate have caused the mistargeting.

This chapter makes three distinct contributions to the literature. First, it contributes to the small literature evaluating the impacts of school choice in developing countries (Angrist et al., 2006; Hsieh and Urquiola, 2006; Muralidharan and Sundararaman, 2015; Glewwe and Muralidharan, 2016). This is the only evaluation of a government implemented school choice policy in a non-Latin American LMIC context. My policy impact estimates and mechanism analysis contribute to this literature by illuminating how the design of the choice program interacts with the context and determines impact or the lack of it.

Second, this chapter also contributes to the growing literature on school choice in India (Muralidharan and Sundararaman, 2015; Singh, 2015). Most of the existing evidence is from rural India. I contribute to the literature by presenting the first experimental evidence from urban India where school choice is more relevant (private school enrollment in urban areas is 49 percent compared to 21 percent in rural areas (Kingdon, 2017)).

Third, my evidence is related to the literature on targeting of social programs in developing countries. Studies in this literature have evaluated the targeting efficiency of poverty alleviation programs and the relative efficacy of different targeting methods (Coady et al., 2004; Alatas et al., 2012; Banerjee et al., 2009; Jha et al., 2013; Radhakrishna and Sharma, 2015; Svedberg, 2012). I contribute to this literature by presenting experimental evidence on the targeting failure of the 25 percent mandate. By showing the failure of income-based targeting deployed by the mandate, I strengthen the claim that targeting of social programs based on simple means-testing is infeasible in developing countries.

In Chapter 2, I use the same experimental set up and data from Chapter 1 to estimate the value added at private schools vis-à-vis public ones and at elite/high-fee private schools vis-à-vis low-fee ones. The 'private school advantage' and the 'elite school advantage' are widely held beliefs in India, despite the lack of unambiguous evidence. I exploit the RTE lottery and use Two Stage Least Squares (2SLS)/ Instrumental Variables (IV) approach to estimate private and elite school effects on childrens' learning and psychosocial outcomes. I find that enrollment at a private school has no statistically significant effect on any of the five test score measures. However, the coefficients are in the range of 0.2-0.6 σ and are not estimated precisely- my sample is only powered to detect very large effects (about 1.3 σ and above). Hence I am not able to comment on the valued added at private schools in learning. On psychosocial outcomes, I find that the effect on self-efficacy, one of the four measured psychosocial outcomes, is large, statistically significant and robust (effect size: 1.90 σ). On the impact of elite schools, I find significant, robust and economically meaningful effects only on English test scores (effect size: 0.58 σ). The English effects are unsurprising, given the labor market premiums to English proficiency. However, the absence of overall learning gains is puzzling and suggests that non-academic aspects of schooling play a critical role in parental school choices.

This chapter makes three specific contributions to the literature. First, it contributes to the large literature from developing countries evaluating the relative effectiveness of private schools vis-à-vis public ones (Ashley et al., 2014; Singh, 2015; Muralidharan and Sundararaman, 2015; Bold et al., 2011).

Despite not being able to comment on the value-added in learning at private schools owing to the small government school sample, this is only the second study in the literature to focus on psychosocial outcomes in addition to test scores. Further, my psychosocial measures are richer than those used by Singh (2015). The large self-efficacy results that I report, read together with the psychology literature that establishes the role of efficacy in promoting learning, highlight the need for incorporating non-test score measures in future evaluations.

A second contribution is the evidence on elite private school effects in LMICs. To my knowledge, this is the first systematic analysis of the elite private school effects on student outcomes. My study design, where free places of different value are randomly assigned to students enables the identification of these effects. This evidence is closely related to the small literature investigating the impact of elite public schools (Abdulkadiroğlu et al., 2014; Lucas and Mbiti, 2014; Zhang, 2012). The non-significant overall test score effects I report are consistent with the pattern of null results in this literature. Finally, the elite school results also contribute to the empirical literature on price-quality relationships. The extant evidence is primarily from the marketing literature that finds weak price-quality correlations in consumer goods (Steenkamp, 1988; Judd, 2000). My results, demonstrating the lack of a relationship between school fee and overall student outcomes, add to this evidence.

Chapter 3 explores the demand-side of the RTE 25 percent mandate. Here I investigate the determinants of choice and analyze if parental preferences for school attributes vary by household type and gender of the child. I use rank-ordered school choices data of 636 parents from Bangalore, India's third largest city, and estimate a Rank-Ordered Logit (ROL) model. I follow the classical discrete choice modelling approach of McFadden (1973; 1978) and extend it to the rank-ordered data following Chapman and Staelin (1982). I find that parents prefer schools with higher fees, size and infrastructure and those that that are closer to home. These results broadly align with the existing evidence on determinants of choice from high-income countries. On heterogeneity, however, my results diverge. I find no difference in parental preferences, by household socioeconomic status (SES), for home-school distance and composition of the student body. This is contrary to evidence from OECD countries, particularly the US, where lower-SES parents seem to have a stronger preference for school proximity and composition relative to higher-SES parents. Finally, I find that parental preferences for school attributes are same for both boys and girls.

This chapter is primarily related to the literature on the demand-side of school choice despite being based on the unique institutional context of the RTE 25 percent mandate. It makes two contributions. First, the evidence adds to the small global literature estimating the determinants of school choice (Hastings et al., 2009; Gallego and Hernando, 2009; Neilson, 2013; Burgess et al., 2015; Carneiro et al., 2016). Further, it is only the second study after Carneiro et al. (2016) from a non-OCED country context. The heterogeneity results are particularly significant as they highlight the centrality of context and policy design in predicting the impacts of school choice.

A second contribution is to the literature on gender bias in India. A small body of descriptive literature has documented gender bias against girls in intra-household resource allocation (White et al., 2016). However, more nuanced investigations have demonstrated no gender bias for younger children (5-9 years) conditional on school enrollment (Kingdon, 2005; Azam and Kingdon, 2013). My results confirm and add to this evidence.

To summarize, this thesis uses carefully collected primary data and rigorous empirical strategies to answer three fundamental questions around the RTE 25 percent mandate and school choice in the urban Indian context. It provides credible evidence that India's national school choice policy, the RTE 25 percent mandate, has not improved childrens' outcomes owing to policy mistargeting. It demonstrates that enrollment at private schools hugely improves childrens' sense of self-efficacy. It establishes that elite private schools do better than non-elite ones only in boosting English test scores and do not significantly impact overall learning. Finally, it shows that parents value proximity, fees, size and infrastructure while choosing private schools in urban India. These results contribute primarily to the literature on school choice in developing countries, while also being related to the literature on targeting of social programs, price-quality correlations and gender bias.

The policy implications of my results and the directions for future research emanating from this work are summarized in the conclusion section. In addition to this introduction, the three substantive chapters and the conclusion, I also present a review of the literature.

References

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.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *The American Economic Review*, 102(4):1206–1240.

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

Ashley, L. D., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., and Rose, P. (2014). The role and impact of private schools in developing countries. Report, 2206, EPPI-Centre.

Azam, M. and Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, 42:143–164.

Banerjee, A., Duflo, E., Chattopadhyay, R., and Shapiro, J. (2009). Targeting efficiency: How well can we identify the poorest of the poor? *Institute for Financial Management and Research Centre for Micro Finance Working Paper*, 21.

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2011). The high return to private schooling in a low-income country. *Center for Global Development Working Paper*, (279).

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289.

Carneiro, P. M., Das, J., and Reis, H. (2016). The value of private schools: evidence from Pakistan. IZA Discussion Paper No. 9960.

Chapman, R. G. and Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, pages 288–301.

Coady, D., Grosh, M. E., and Hoddinott, J. (2004). *Targeting of transfers in developing countries: Review of lessons and experience*, volume 1. World Bank Publications.

Gallego, F. and Hernando, A. (2009). School choice in Chile: Looking at the demand side. Working Paper No. 356, 2010, Pontificia Universidad Catolica de Chile.

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

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46.

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):1477–1503.

Jha, R., Gaiha, R., Pandey, M. K., and Kaicker, N. (2013). Food subsidy, income transfer and the poor: A comparative analysis of the public distribution system in India's states. *Journal of Policy Modeling*, 35(6):887–908.

Judd, V. C. (2000). The price-quality relationship: an empirical study of food products. *Journal of Food Products Marketing*, 6(1):11–24.

Kingdon, G. G. (2005). Where has all the bias gone? Detecting gender bias in the intrahousehold allocation of educational expenditure. *Economic Development and Cultural Change*, 53(2):409–451.

Kingdon, G. G. (2017). The Private Schooling Phenomenon in India: A Review. IZA Discussion Paper No. 10612.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics*, 6(3):234–263.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers of Econometrics*.

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.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Mimeo, Yale University.

Radhakrishna, R. and Sharma, A. N. (2015). Poverty and social protection in urban india. Working Papers, id:7563, esocialsciences.

Sarin, A., Kuhn, S., Singh, B. D., Khanghta, P., Dongre, A., Joshi, E., Sengupta, A., Das, R., and Rahman, F. (2016). State of the Nation: Rte Section 12 (1) (c) 2015. Technical report, Indian Institute of Management Ahmedabad and others.

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.

Steenkamp, J.-B. E. (1988). The relationship between price and quality in the marketplace. *De Economist*, 136(4):491–507.

Svedberg, P. (2012). Reforming or replacing the public distribution system with cash transfers? *Economic and Political Weekly*, pages 53–62.

White, G., Ruther, M., and Kahn, J. (2016). *Educational Inequality in India: An Analysis of Gender Differences in Reading and Mathematics*. IHDS Working Paper 2016-2.

Zhang, H. (2012). The mirage of elite schools: evidence from lottery-based school admissions in China. Working Paper, Chinese University of Hong Kong.

Literature Review

School choice policies entail introducing market mechanisms in schooling, which will lead to expansion of the choice set of low-income parents who face binding income constraints. Widely used choice mechanisms are universal or targeted vouchers, open enrollment into magnet or public schools and charter schools. The choice debate, theory and literature is largely focused on the United States (US) and to some extent on other OECD countries. Over the last few decades school choice has been a controversial area of education policy: either condemned as an attack on public schooling or extolled as an effective strategy to boost productivity in the education system (Wolfe, 2009).

In this literature review, I start with the theory and debate around school choice. I then describe the different types of school choice policies, after which I compare and contrast the existing empirical literature on the impact and demand-side of school choice in both developed and developing countries. Finally, I summarize and highlight the gaps in existing knowledge and give an overview of the specific research questions this thesis will address.

School choice- Theory and debate

Milton Friedman (1962) propounded the 'market model' for the universalization of school choice through a voucher plan designed to supplant the public education system. In this model, greater competition will eliminate failing public schools (whose budgets are tied to student enrollment), improve the productivity of some public schools through the 'competition effect' and lead to an increased supply of productive private schools due to availability of public dollars for private institutions. His ultimate vision for schooling in America was to replace the existing public system with a market place of schools that were publicly financed and privately run (Hoxby, 2003; Viteritti, 2003). A basic assumption of the model is that public schools are inefficient owing to their monopolistic nature and that private schools are more effective than public schools (Hsieh and Urquiola, 2006). The theory, therefore, posits that introduction of choice will lead to improved public school productivity as the accountability and incentive mechanisms inherent in the private system become applicable to public schools.

The primary mechanism for the overall improvement in productivity, of both the public and the private schools, is parental ability as consumers to 'vote with their feet' in accordance with the Tiebout (1956) choice model. Another channel for improvement in outcomes from enhanced choice, that is not as widely discussed, is better matching: choice empowers parents to match the ability of children with the type of schools (Cullen and Jacob, 2007; Rouse and Barrow, 2009).

In the US, the largest and oldest laboratory for theoretical and empirical examinations of school choice, the first generation debates focused on the efficiency gains from the 'market model,' while the second generation arguments favoured an 'opportunity model' with equity as the central concern. Advocates of this model view choice as a mechanism to improve educational outcomes of under-served and low-income racial minorities whose children overwhelmingly attend failing local schools due to income and mobility constraints. While the market model advocated universal vouchers, this model argues for targeted vouchers to disadvantaged children. Over the years the targeting approach has become more widely accepted both from a normative and efficiency perspective. Normatively, it is argued, that an education system where high-income households have choices[3] while low-income ones do not is inherently unfair (Viteritti, 2003). Economic logic also contends that school choice programs should only benefit children from urban poor families because the constraints (income and/or residence) they face are binding (Hoxby, 2007; Viteritti, 2003).

Opponents of choice have argued that it leads to greater segregation in schools as private schools will tend to 'cream-skim' better students than improving their productivity to survive in the market place. The cream-skimming argument goes that faced with competition, a private school can either improve productivity through the longer and costly route of investing in their teachers and so on, or through the shorter and less-expensive route of selecting better students (Hsieh and Urquiola, 2006). Another concern is the demand-side aspect of parental ability to identify and choose effective schools. Choice will lead to improved outcomes only if parents are willing and have the ability to prioritize academic quality over all other aspects of schooling. However, research shows that factors like peer groups, school environment, facilities and safety weigh heavily into the decision making process of the parents owing to preferences and their inability to observe

[3] High-income households can access better schools by changing residence and/or by sending children to fee-charging private schools. They do not need government support to do this. Low-income households , on the other hand, can neither change residence nor shift their children to private schools because the income constraints they face are binding.

school effectiveness. This in turn could create more incentives for cream-skimming leading to increased stratification of schools along household socioeconomic status (SES) (Hanushek, 1981; Hsieh and Urquiola, 2006; Rothstein, 2006; Viteritti, 2003; Musset, 2012). Finally, it is argued that over reliance on market mechanisms cannot be the default solution for improving school education, given its public goods nature (Drèze and Sen, 2013, p.137).

Hence, despite the theoretical prediction that school choice could impact student outcomes and school efficiency, the direction of the impact is not unambiguous. It seems to be dependent on several assumptions (on productivity of public schools, level of competition in education markets and availability of information) and on the specific features of the school choice policy.

Finally, it is critical to note that the dichotomy in views between the proponents and opponents of choice is a reflection of the larger ideological divides between the "libertarian" and "social contract" positions. As elucidated by Belfield and Levin (2005, p.548):

> Libertarians believe that freedom of choice should be the highest priority of voucher programs and assume that a private market place with increased options will promote greater efficiency and (possibly) equity. Advocates of social contract maintain that education generates important positive externalities that are best promoted through a free, publicly funded and operated and democratically determined education system.

Types of school choice

Several types of school choice policies have been implemented over the last few decades. Giving school vouchers, which typically cover the costs of tuition, to households is the most widespread choice policy. Friedman advocated a universal voucher program, which would get rid of the state's education monopoly, promote efficiency and further equity (Viteritti, 2003). Chile is probably the only country that has implemented a nationwide, universal and comprehensive school voucher program (since 1981) by providing vouchers to all students desirous of attending private schools and linking the budgets of public schools to enrollments. The universal model is not, however, very common in other OECD countries (Musset, 2012). The more common programs are those that target vouchers to lower-SES households who might otherwise not be able to exercise choice due to binding constraints. Targeted voucher programs are believed to be equity enhancing and redistributive and have gained greater political support in the last few decades. Notable examples are the US, where targeted voucher programs have been implemented both by local governments and private charities, several other OECD countries, Chile whose universal program became more targeted since 2008[4] and Columbia that has one of the largest targeted voucher programs in the world.

Other important school choice policies, especially in the US, include open enrollment policies that allow children to enroll into public and magnet schools outside their neighborhood, and charter schools, which are independent, open-enrollment public schools set up under a charter with greater flexibility in management and accountability than traditional public schools.

[4] Chile introduced weighted vouchers since 2008 to specifically benefit children from lower-SES backgrounds.

Empirical literature

The contentious nature of the choice debate over the last five decades, and the theoretical ambiguity on the impact of choice has fuelled a large body of empirical research (Viteritti, 2003). The literature can be broadly categorized into studies investigating the impact of choice policies and those that investigate the demand-side of school choice i.e., the question of how parents make school choices. I will first present the large body of evidence from high-income country contexts and then move on to the relatively smaller group of studies from low and middle-income countries.[5]

Impacts of choice: high-income countries

The bulk of the empirical literature on school choice is from the US and other OECD/ high-income countries. The fundamental question addressed by this literature is that of impact of various choice policies. Voucher programs, the most common school choice interventions, are unsurprisingly the best studied. In a recent review Rouse and Barrow (2009) conclude that the most credible evidence does not find any statistically significant impacts of voucher programs in the US. Musset's (2012) review on voucher programs in other OECD countries also arrives at similar conclusions.

The evidence on other forms of school choice is more mixed. For open enrollment policies, where students are guaranteed admission in their

[5] I use the World Bank categorization of high, middle,and low-income countries, rather than the ambiguous developed and developing categorization. This categorization is particularly relevant here because Chile, where a lot of school choice literature comes from, has been categorized as both developing and developed in different reviews. Based on current World Bank definitions, Chile is a high-income country, in addition to being in the OECD.

neighborhood (home) school but are allowed to choose other schools in the district, the two most rigorous studies find opposite results. Deming et al. (2009) exploit school lotteries in Charlotte- Mecklenburg school district and find that students with the lowest performing home schools benefited greatly from choice while those with even average quality of home schools did not gain. In contrast, Cullen et al. (2006) study Chicago Public Schools and find little evidence of any positive impact of winning a lottery on traditional outcome measures like standardized test scores and graduation rates; contrary to theoretical predictions, sub-groups like African-American children and below-average students who would be expected to gain the most from school choice are negatively impacted.

On charter schools, where majority of the recent literature from the US is concentrated, the evidence in favour of charter effectiveness is stronger. Research that finds positive impacts of attending a charter school include the following: Hoxby and Rockoff (2004) study seven charter schools in Chicago and report 6 percentage point statistically significant charter school effects in both reading and math. Abdulkadiroğlu et al. (2011) study charters in Boston and find 'impressive score gains' for middle and high school students, in the range of 0.25 standard deviations for English language and Arts and 0.4 standard deviations in math. Hoxby and Murarka (2009) investigate New York charter school effects and find that attending charter school is associated with 0.09 standard deviations per year increase in math and 0.04 standard deviations per year increase in reading scores. Finally, Angrist et al. (2013) study Massachusetts charter schools and conclude that urban charter schools are effective in improving student learning and that non-urban ones are not. However, one large evaluation of 36 charter middle schools across 15 states does not find any significant learning gains (Gleason et al., 2010).

Research discussed above only investigates the partial equilibrium effects of choice i.e., how the marginal students' outcomes are impacted by access to a choice school holding the existence of the choice program constant. However, a large-scale choice program would have general equilibrium effects that might alter the student composition and academic effectiveness across both public and private schools (Cullen et al., 2006). The evaluation of Chile's voucher program is probably the only rigorous research on the long-term general equilibrium effects of a choice program. Using panel data for about 150 municipalities, Hsieh and Urquiola (2006) find no evidence that choice improves education outcomes like test scores, school repetition rates and years of schooling. Summarizing the US experience, Rouse and Barrow (2009) conclude that a comprehensive voucher system might not lead to large improvements in academic outcomes. Musset's (2012, p. 30) review of the aggregate impacts of school choice in OECD countries summarizes:

> Majority of the evidence suggests that different schemes of school choice do not, through the competition they create for local schools, induce them to improve, nor does it improve the student achievement of those who take advantage of more school choice and opt out of their local school.

Impacts of choice: low and middle-income countries

The operation of school choice is different in low and middle-income countries relative to OECD countries because of the differences in functioning and funding of the public education system and the extent and nature of private schooling. First, public schools are funded by national and state government budgets and not by decentralized taxes like in the US. This precludes the possibility of operation of Tiebout choice mechanism to discipline public schools. Second, most public school teachers are tenured civil servants with pay and tenure delinked from performance measures like student test scores or enrollment. Hence, it is unlikely that choice policies could trigger competition between public and private schools.

Finally, private primary school enrollment in these contexts is much higher than that in rich countries.[6] And there is wide variation in the type and quality of private schools. Unlike private schools in high-income countries that are typically expensive and are attended only by the elite, private schools in developing countries are varied in their fee structure and cater to a broader section of the society (Patrinos et al., 2009). Choice proponents in these contexts assume that private schools are better than public schools, and that high-fee private schools are better than low-fee ones at producing education outcomes. They contend that expanded choice will improve learning outcomes of disadvantaged children, not so much because of the choice-induced competition improving overall productivity of the education system, but because of the pre-existing higher productivity of private schools in general and of elite/ high-fee private schools in particular.

The two rigorous evaluations of school choice policies in these contexts arrive at opposite conclusions. Evaluations of Columbia's large secondary school voucher program (Angrist et al., 2002, 2006) find that voucher winners have increased educational attainment, academic achievement and better non-educational outcomes. On the contrary, Muralidharan and Sundararaman (2015) find modest English language gains (0.12 standard deviation) and no gains in math, local language and aggregate test scores for children aged 10 years in southern India. The question of impact of choice policies is, therefore,

[6] For instance, private primary school enrollment in India, Pakistan and Bangladesh is higher than 30 percent, compared to less than 10 percent in the US and the UK.

not as widely studied in these contexts, given the absence of large choice programs.

Majority of the choice literature in these countries focuses on the question of relative effectiveness of private schools and public schools. Ashley et al. (2014) review the results of 21 studies from India, Pakistan, Kenya, Ghana, Nigeria and Nepal. They conclude that the evidence in favor of private schools producing better learning outcomes than government ones is moderately strong, though the size of the private school effect is ambiguous. However, barring a few studies (discussed below), most of the literature in this review is chiefly based on cross sectional estimates that are confounded by selection and omitted variables and cannot be treated as causal.

Rigorous empirical studies have used multiple approaches to overcome the endogenity problem: value added models (Singh, 2015), fixed effects estimation (French and Kingdon, 2010), aggregation (Bold et al., 2011) and randomized control trials (Muralidharan and Sundararaman, 2015). The bulk of this rigorous literature is located in India, where the private sector is one of the largest, and the belief in existence of a positive private school effect is widespread. This belief is probably driven by cross-sectional comparisons that establish the superior performance of children in private schools (Muralidharan and Kremer, 2006). However, the two most recent and rigorous studies in this literature do not fully support this general belief. Singh (2015) estimates value-added-models on a panel dataset of 9-year-old children from rural areas of the south Indian state of Andhra Pradesh. He reports large English learning gains (0.65 standard deviation) and modest receptive vocabulary gains (0.13 standard deviation). The study also has a relatively small urban sample on which he finds no private school learning gains and concludes that even if such premium exists in urban areas, it is unlikely to be substantial. Muralidharan and Sundararaman (2015) study impact of vouchers and the private school premium using a large RCT, also conducted in rural Andhra Pradesh. They report zero average impact across three subjects on which children were assessed and modest English language gains of 0.12 standard deviation at the end of four years of the study.

So the two most rigorous studies in this literature are based in the South Indian state of Andhra Pradesh and arrive at slightly different conclusions. While Singh (2015) finds a large and significant private school premium in English, Muralidharan and Sundararaman (2015) report very modest English effects and zero overall effects. Further, both these studies are based in rural areas, while the private schooling phenomenon is primarily urban (Kingdon, 2017). Finally, while summarizing the literature on school choice in developing countries, Glewwe and Muralidharan (2016) opine that the existing evidence points to private schools being more productive than government schools (produce the same results at much lower cots), but not more effective in raising test scores. They conclude that further evidence is required to understand if the higher productivity of private schools can be translated to better learning outcomes.

Demand-side literature

As explained in the theory section, the claim that expanded choice improves student learning is premised on parents prioritizing academic achievement over all other school attributes and on their ability to differentiate effective schools from less effective ones. However, several studies have established that factors such as peer groups, school environment, facilities, safety and proximity (home-school distance) weigh heavily into the decision making process of parents. This could be due to difficulties associated with obtaining information on school effectiveness or because parents rationally prioritize non-academic factors over academic ones or some combination of both. Further, evidence also suggests that the relative weighting of academic and non-academic factors in making choices seems to vary by household SES.

If non-academic factors are key determinants of school choice, and if there is significant heterogeneity across SES groups in making choices, the theoretical prediction of expanded choice leading to improved outcomes will not hold. Worse still, expanded choice could lead to increased stratification of the schooling system and widen the learning achievement gaps between different SES groups. Hence understanding the determinants of school choice and investigating heterogeneity in preferences across household types are critical empirical questions in school choice theory. Though not as widely studied as the question of 'impact of choice policies,' understanding the determinants of choice enables a thorough appreciation of the potential for school choice policies to be effective in a given context.

In his analysis of the American public school system, Hanushek (1981) argues that while choosing schools parents are not necessarily buying the best education services, but are selecting social and physical environments that they and their children prefer. Rothstein (2006) investigates the demand side of school choice by studying the distribution of student outcomes across schools within metropolitan housing markets in the US and concludes that school effectiveness is not a primary determinant of parental choices. Hsieh and Urquiola (2006) study the impact of Chile's nationwide school choice program and find that parents seem to prefer schools with good peers rather than those that are most productive. On a similar note, Abdulkadiroğlu et al. (2014) find that the highly competitive and coveted exam school education in Boston and New York has very little effect even for the highest marginal ability applicants compared to regular public education. They conclude that parents seem to care about non-learning related attributes while choosing exam schools. They, hence, posit that school choice might not have strong demand-side effects on the production of human capital in schools.

Hastings and Weinstein (2008) evaluate the importance of information in parental school choices and find that information increases the proportion of parents choosing higher- performing schools, conditional on availability of such schools within a reasonable distance. Gallego and Hernando (2009) and Neilson (2013) study the determinants of parental choice in Chile and find that parents care both about academic and non-academic attributes of schooling, implying that they face trade-offs between their preference for test scores and for other school attributes. Musset (2012) summarizes several other studies from OECD countries and argues that the determinants of parental school choices are complex and go beyond simple academically rational reasons.

On heterogeneous preferences by parental SES, Hastings et al. (2009) study a school choice policy in North Carolina and find that higher-SES parents prioritize schools' academic effectiveness much more than lower-SES ones. They posit that lower-SES families face tougher trade-offs between their preferences for test scores and that for proximity and school composition. Similarly, Burgess et al. (2015), in their study of parental choices in England, find significant heterogeneity across SES groups in their demand for academic effectiveness. Their demand elasticities show that the highest SES group responds much more significantly to improvements in academic standards than the other SES groups. Musset (2012, p. 33) argues that lower SES parents are at a disadvantage in acquiring information on school effectiveness and hence underweight it. She explains:

> information acquisition has very high costs, especially for parents who lack the needed social capital, the resources, the time, the connections, or the cultural resources to effectively choose.

The entire evidence on the demand-side of school choice is from OECD countries, barring one study from Pakistan (Carneiro et al., 2016).[7]

Gaps in the literature

The extant literature on school choice can be summarized as follows:

- The first broad area of inquiry concerns the impact of school choice on student outcomes. The bulk of the existing evidence here is from high-income/OECD countries. The best evidence on school vouchers finds relatively small impacts, while the evidence on other forms of school choice (like open enrollment and charter schools) is more mixed. The evidence from lower and middle-income countries is scant with only two rigorous studies.
- The second key area is that of the demand-side of school choice and understanding how parents choose schools. Here again, the entire evidence (but for one study from Pakistan) is from OECD countries.

[7] They find that the central determinants of school choice in rural Pakistan are the home-school distance, school fee and the characteristics of peers.

The evidence suggests that though parents prefer academic outcomes, there are trade-offs between preferences for academic outcomes and non-academic aspects of education like home-school distance and composition of student body. Further, a number of studies establish heterogeneity in parental preferences by family background indicating that choice could lead to greater stratification of the schooling system.

- In the context of low and middle-income countries, the bulk of the literature is focused on evaluating the effectiveness of private schools relative to public ones. This is critical in these contexts given the increasing role of the private sector. The best studies are from India, but the evidence is not conclusive. Further there is no evidence on the widely held claim that high-fee is correlated with higher educational quality in the private sector.

There are, therefore, three key questions concerning school choice in low and middle-income countries : First, what is the impact of school choice policies? Second, is there a private school and an elite school premium in learning? And third, what are the determinants of parental school choices, and do they vary by household type? Despite the centrality of these questions and the growing demands for introduction of school choice policies, evidence on all these questions is either sparse or completely absent.

Research questions for this thesis

In this thesis, I seek to bridge the knowledge gap in the school choice literature on developing countries. I study India's national school choice policy and answer the following three specific questions:

1. What is the impact of the 25 percent mandate on childrens' learning and psychosocial outcomes? Has school choice led to improved student outcomes as predicted by theory?

2. What is the value added at private and elite (private) schools?[8] Are private schools in general, and elite private schools in particular, more effective at improving student outcomes as contended by choice enthusiasts?

3. What are the determinants of parental school choices? Do preferences for schools vary by household and child characteristics?

[8] There is no agreed upon definition of elite schools in the literature. I define schools with fee higher than the 75th percentile (p75) of the fee distribution as high-fee/elite.

References

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., 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.

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

Ashley, L. D., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., and Rose, P. (2014). The role and impact of private schools in developing countries. Report, 2206, EPPI-Centre.

Belfield, C. and Levin, H. M. (2005). Vouchers and public policy: When ideology trumps evidence. *American Journal of Education*, 111(4):548–567.

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2011). The high return to private schooling in a low-income country. *Center for Global Development Working Paper*, (279).

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289.

Carneiro, P. M., Das, J., and Reis, H. (2016). The value of private schools: evidence from Pakistan. IZA Discussion Paper No. 9960.

Cullen, J. B. and Jacob, B. A. (2007). *Is gaining access to selective elementary schools gaining ground? Evidence from randomized lotteries*, pages 43–84. The Problems of Disadvantaged Youth: An Economic Perspective. University of Chicago Press.

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.

Deming, D., Hastings, J., Kane, T., and Staiger, D. (2009). School choice and college attendance: Evidence from randomized lotteries. *NBER Working Paper*.

Drèze, J. and Sen, A. (2013). *An uncertain glory : India and its contradictions*. London : Allen Lane.

French, R. and Kingdon, G. (2010). The relative effectiveness of private and government schools in Rural India: Evidence from ASER data. *London: Institute of Education*.

Friedman, M. (1962). *Capitalism and freedom*. University of Chicago Press.

Gallego, F. and Hernando, A. (2009). School choice in Chile: Looking at the demand side. Working Paper No. 356, 2010, Pontificia Universidad Catolica de Chile.

Gleason, P., Clark, M., Tuttle, C. C., and Dwoyer, E. (2010). The evaluation of charter school impacts: Final report. ncee 2010-4029. *National Center for Education Evaluation and Regional Assistance*.

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

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management*, 1(1):19–41.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, 123(4):1373–1414.

Hoxby, C. M. (2003). School choice and school productivity. Could school choice be a tide that lifts all boats? In *The economics of school choice*, pages 287–342. University of Chicago Press.

Hoxby, C. M. (2007). *The economics of school choice*. University of Chicago Press.

Hoxby, C. M. and Murarka, S. (2009). Charter schools in new york city: Who enrolls and how they affect their students' achievement. Working paper 14852, National Bureau of Economic Research.

Hoxby, C. M. and Rockoff, J. E. (2004). *The impact of charter schools on student achievement*. Department of Economics, Harvard University Cambridge, MA.

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):1477–1503.

Kingdon, G. G. (2017). The Private Schooling Phenomenon in India: A Review. IZA Discussion Paper No. 10612.

Muralidharan, K. and Kremer, M. (2006). Public and private schools in rural india. *Harvard University, Department of Economics, Cambridge, MA*.

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.

Musset, P. (2012). School choice and equity: current policies in OECD countries and a literature review. Technical report, OECD.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Mimeo, Yale University.

Patrinos, H. A., Osorio, F. B., and Guáqueta, J. (2009). *The role and impact of public-private partnerships in education*. World Bank Publications.

Rothstein, J. M. (2006). Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350.

Rouse, C. E. and Barrow, L. (2009). School vouchers and student achievement: Recent evidence and remaining questions. *Annu. Rev. Econ.*, 1(1):17–42.

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.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, pages 416–424.

Viteritti, J. (2003). Defining equity: Politics, markets, and public policy. *School choice: The moral debate*, pages 13–30.

Wolfe, A. (2009). *School choice: The moral debate*. Princeton University Press.

CHAPTER 1

Choice for the poor or poor choice? Experimental evidence on the impact of India's school choice policy

Abstract

I evaluate the impact of a national school choice policy in the south Indian state of Karnataka. The policy sets aside 25 percent places in private schools for children from disadvantaged backgrounds, with the government paying the tuition fee to private schools. Exploiting the lottery based allocation of free places, I investigate whether the learning and psychosocial outcomes of lottery-winning children (policy beneficiaries) improved relative to lottery-losing ones. The study cohort comprises of 7-year-old children who started Grade I in 2015. After 1.5 years of schooling, I find no difference in test scores of lottery winning and losing children on a range of subjects. Policy beneficiaries, however, perform significantly better than non-beneficiaries on one of the four psychosocial measures: self-efficacy. I conduct a detailed mechanism analysis to investigate the non-impact, which reveals that majority of the applicants were default private school goers and would have attended similar schools irrespective of the policy. This explains the absence of treatment effects on student outcomes. The evidence establishes that the policy has

been mistargeted, allowing ineligible households to apply and deterring low-income households from participating. I replicate the mistargeting result on another large representative sample drawn from across Karnataka state.

1.1 Introduction

School choice programs, like voucher systems, have been important and contentious education reforms in the last few decades. They entail introducing market mechanisms in schooling by making public funds available to any type of school (public, private, or faith-based) chosen by parents. Parental ability to choose schools (as consumers) is expected to trigger competition in schooling markets, leading to overall improvements in school performance. Furthermore, choice has an equity aspect as it potentially benefits low-income households the most: these households face binding income constraints that oblige them to enroll their children at failing neighborhood public schools. Theory, hence, predicts that school choice would promote student outcomes, school-level efficiency and social equity. However, these expectations are strongly contingent on the underlying assumptions of parental behavior and nature of schooling markets. Opponents of choice have long argued that it will not improve outcomes, and worse still, cause detriment to the schooling system by increasing stratification along race and socioeconomic status (SES) divides. Given this theoretical ambiguity, understanding the impact of school choice policies becomes an empirical question.

The empirical evidence is also, unfortunately, inconclusive: the size and direction of the impacts of choice seem to vary by context and design/type of the choice program. Rouse and Barrow (2009) and Musset (2012) summarize the evidence. Further, most of the existing empirical literature is focused on the US and other OECD countries.[1] The literature from developing country contexts is at a nascent stage (Glewwe and Muralidharan, 2016; Morgan et al., 2013) as large-scale choice programs have not been widely implemented in these settings. Calls for the introduction of school choice policies have, of late, been increasing in developing countries like India, given the large and emergent private sector and the growing perceptions of failure of government schools. India's Right to Education (RTE) Act of 2009 introduced a unique choice policy, which when fully implemented could impact 16 million children making it the world's largest program for public funding and private provision in education (Sarin et al., 2016).

In this chapter, I present experimental evidence on the impact of this policy. The policy (hereafter referred to as the RTE 25 percent mandate/ the mandate) sets aside 25 percent places (hereafter referred to as RTE free places/ free place) in entry grades of all private schools for children from disadvantaged households. Government pays the associated tuition costs for these places to private schools. Families from certain unprivileged social groups (castes) and those below defined income cutoffs are classified as disadvantaged and are eligible for free places. The stated objectives of the policy are to provide better quality education to disadvantaged children and to desegregate the Indian schooling system.

[1] The Organization for Economic Cooperation and Development is a club of 35 rich and emerging economies.Chile, from where a lot of evidence on choice comes is part of the club. Large emerging economies like China, India and Brazil are not part of it.

Despite these lofty goals and the legal mandate, several Indian states are skeptical about the program and have not implemented it. They argue that the mandate will reduce public school budgets by diverting scarce public funds to private schools and worry about the potential for incorrect identification of income-disadvantaged households, given the lack of information on household incomes. On the other hand, States implementing the mandate view it as a public-private partnership through which the government can procure higher quality education from the private sector. They believe that the mandate provides a ladder of opportunity for disadvantaged children enabling them to shift from their default low-quality learning environments (government or low-fee private schools) to high-quality ones (high-fee private schools). Against this background, I investigate the impact of the 25 percent mandate on childrens' learning and psychosocial outcomes.

My study sample of 1,616 children comes from two regions of the south Indian state of Karnataka. I exploit the lottery-based allocation of RTE free place and use a pairwise matching design (Bruhn and McKenzie, 2009) to estimate the causal impact of the policy. My matching design accounts for the admission process and the lottery algorithm and generates treatment-control pairs with each pair of children having the same ex-ante probability of admission. The study cohort enrolled in Grade I in mid-2015 and were aged about 7.3 years, on average, at end line data collection in November 2016.

The results show that after 1.5 years of schooling, there is no difference in test scores of lottery winning and lottery losing children in four subjects- General Cognitive Ability (GCA), Math, English and Kannada (the local language)- and on the total test scores measure. On the four psychosocial measures that I test children on, lottery winners perform better than lottery losers only on the self-efficacy index (effect size: 0.12 Standard Deviation (σ)).[2] The self-efficacy result is robust to alternate specifications and multiple hypothesis correction. Further, there is no heterogeneity in treatment effects for different sub-groups in the sample.

I explore the mechanisms driving the non-impact and determine that the policy did not shift children from their default low-quality schools to high-quality schools as expected. First, 99 percent of the children in treatment group (lottery winners) and 93 percent of those in the control group (lottery losers) are enrolled in private schools, establishing that most of the policy applicants are default private school goers. The poorest and most deserving households (default government school goers) have not participated in the policy. Second, both the treatment and control children, on average, attend private schools that charge around the median school fee. Hence there is no qualitative difference between the type of schools attended by treatment and control children, and the average policy applicant is a default private school goer who can afford the median school fee. All this leads to the central policy result of this chapter: the 25 percent mandate is mistargeted. I replicate this result on a much larger sample of children drawn from across the state.

I demonstrate that design flaws in the 25 percent mandate have led to the mistargeting. Making eligibility contingent on income, whose determination is difficult in the Indian context, has enabled many ineligible households to participate in the program. Further, the policy only covers the tuition costs of education and not the non-tuition costs, which I estimate to be around 1.3

[2] In addition to self-efficacy, I also test children on the level of peer support, school support, and teacher support.

times the tuition costs. This partial nature of the RTE subsidy has probably deterred the poorest parents from applying to the mandate.

My research contributes to and enriches several strands of empirical literature in economics and public policy. First, I add to the large global literature on the impact of school choice programs on childrens' outcomes. This literature comprises investigations of the impact of various types of choice programs: universal and targeted vouchers, charter schools, open enrollment and magnet schools. Rouse and Barrow (2009) and Musset (2012) summarize the evidence. My policy impact estimates and mechanism analysis contribute to this literature by illuminating how the design of the choice program interacts with the context and determines impact or the lack of it. Within this broader literature, my contribution is particularly relevant to the smaller body of empirical evidence on impact of choice policies in developing countries. Glewwe and Muralidharan (2016) summarize the few rigorous studies from Chile, Colombia and India and opine that more evidence is needed to better understand school choice in these contexts. Focusing further, my findings contribute significantly to the small literature on the impacts of school choice in India. Additionally, this is the first evaluation of the impact of a government-implemented large-scale school choice policy in the Indian context and has much higher generalizability than previous evaluations that are based on NGO-run programs (Muralidharan and Sundararaman, 2015).

Second, my headline result of mistargeting of the policy and the detailed mechanism analysis speak to the literature on targeting the poor for social programs in developing countries. A large part of this literature is centered on poverty alleviation programs like food subsidies, cash transfers and public works. Coady et al. (2004) summarize the evidence on 122 programs from 48 countries across the developing world. There is also a significant body of research demonstrating the mistargeting of welfare programs in the Indian context (Jha et al., 2013; Radhakrishna and Sharma, 2015; Svedberg, 2012; Drèze and Khera, 2010). I contribute to this literature by demonstrating the failure of income-based targeting, a relatively new and less used targeting approach in Indian. Further, my evidence holds relevance across the developing world, where ascertaining household incomes is challenging.

The rest of the paper is structured as follows: Section 1.2 presents the school choice theory and debate. Section 1.3 presents the institutional setting and describes the school choice policy being evaluated. Section 1.4 explains the research design and sampling strategy. Section 1.5 presents the data, while Section 1.6 presents the results. Section 1.7 is a detailed analysis of the mechanism driving the results. Section 1.8 is a detailed discussion along with policy implications. Section 1.9 explores the external validity of the results by presenting evidence from another sample. Section 1.10 concludes.

1.2 Theory and literature

1.2.1 School choice: theory and debate

School choice programs allow relatively poor parents to choose the type of schools their children attend. In the context of the US where the choice debate, theory, and literature is largely focused, the default school option for children, especially from underprivileged socioeconomic backgrounds, is the neighborhood public school. Their parents do not have the income and other means to afford fee-charging private schools or to move to better school districts. These twin constraints of income and residence are binding for families from particular socioeconomic groups, for instance low-income racial minorities. School choice policies seek to relax the binding constraints of income and place of residence and move lower-SES children to better schools.

Proponents of choice contend that making public funds available for private institutions, linking public school budgets to student enrollment and giving parents the choice to decide where they want to educate their child would introduce competition amongst education providers (competition effect) leading to enhanced productivity of the entire education system and improved childrens' learning outcomes. The assumptions of this 'market model' of school choice are that public schools are inefficient because of their monopolistic nature and that private schools are more effective than public schools (Hsieh and Urquiola, 2006). The underlying mechanism driving the overall improvement in productivity in both public and private schools is parental ability as consumers to 'vote with their feet' in accordance with the Tiebout (1956) choice model. Finally, choice would also promote equity if its benefits were targeted to poor and disadvantaged households (Hoxby, 2003; Wolfe, 2009).

Choice skeptics maintain that it leads to greater segregation in schools, as private schools would 'cream-skim' better students rather than improving productivity to survive in the market place.[3] Further, 'choice improves education outcomes' argument is premised on parental ability to choose

[3] The cream-skimming argument goes that faced with competition, a private school can either improve productivity through the longer and costly route of investing in their teachers and so on, or through the shorter and less expensive route of selecting better students (Hsieh and Urquiola, 2006).

schools that produce the best outcomes. However, school effectiveness is unobservable, and parents might end up choosing schools based on criteria other than quality of education. This in turn could create more incentives for cream-skimming and precipitate increased stratification of schools along household SES (Hanushek, 1981; Hsieh and Urquiola, 2006; Rothstein, 2006; Wolfe, 2009; Musset, 2012). Scholars have also argued that over reliance on market mechanisms cannot be the default solution for improving school education given its public goods nature (Drèze and Sen, 2013).

Theoretical uncertainty on the impacts of choice and the contentious nature of the choice debate has opened doors for empirical research into the questions of impact of choice on school productivity, equity and student outcomes. Unfortunately, the empirical evidence is also ambiguous on the impact questions; the results seem to be specific to the context and policy design. Rouse and Barrow (2009) and Musset (2012) summarize the evidence from the US and other OECD countries. The literature from developing country contexts is at a nascent stage (Morgan et al., 2013; Glewwe and Muralidharan, 2016).

1.2.2 School choice in developing countries

The operation of school choice is different in low and middle-income countries relative to OECD countries because of the differences in functioning and funding of the public education system and the extent and nature of private schooling. First, public schools are funded by national and state government budgets and not by decentralized taxes like in the US. This precludes the possibility of operation of Tiebout choice mechanism to discipline public schools. Second, most public school teachers are tenured civil servants with pay and tenure delinked from performance measures like student test scores or enrollment. Hence it is unlikely that choice policies could trigger competition between public and private schools.

Second, private primary school enrollment in these contexts is much higher than that in rich countries.[4] And there is wide variation in the type and quality of private schools. Unlike private schools in high-income countries that are typically expensive and are attended only by the elite, private schools in developing countries are varied in their fee structure and cater to a broader section of the society (Patrinos et al., 2009). Choice proponents in these contexts assume that private schools are better than public schools, and that high-fee private schools are better than low-fee private schools at producing education outcomes. They contend that expanded choice will improve the learning outcomes of disadvantaged children, not so much because of the choice-induced-competition improving overall productivity of the education system, but because of the pre-existing higher productivity of private schools in general, and of high-fee private schools in particular.

1.3 Institutional setting

1.3.1 Elementary education system in India

About 200 million children are part of India's elementary education system (Grades 1-8). The following contextual details enable better appreciation of the policy/ program being investigated. First, private school enrollment is one of the highest in the developing world at 38 percent (Mehta, 2015). The

[4] For instance, private primary school enrollment in India, Pakistan and Bangladesh is higher than 30 percent, compared to less than 10 percent in the US and the UK.

national average masks variation in levels of private school enrollment across states and between urban and rural areas. For instance, 49 percent of children in urban India are enrolled at private schools (Kingdon, 2017) and five Indian states have private school enrollment rates of greater than 50 percent (Pratham Foundation, 2014). There has been a steady migration of children from government to private schools over the last decade, a phenomenon described as emptying of government schools (Kingdon, 2017). This can be attributed to the general perception that private schools are of superior quality than government schools (Juneja, 2014). There is, however, no rigorous evidence supporting this perception. Survey data suggests that parents prefer private schools for the following reasons: better learning environments in private schools, unsatisfactory quality of education in government schools and absence of English medium instruction in government schools (Saha, 2016).

Second, private schools serve not just the wealthy and the middle class, but also a section of the poor. The private school fee distribution in India is log-normal with a pronounced rightward skew: there are large number of low-fee schools and a smaller number of ultra high-fee schools. The highest annual fee is around Indian Rupees (INR) 200,000 (USD 3,100), 40 times the median fee of INR 5,000 (USD 77).[5] Further, the mean private school fee is only about 9.2 percent of the per capita GDP, and about 26 percent of rural private schools have a monthly fee that is below the daily minimum wage (Kingdon, 2017).

[5] 1 USD= 65 INR. I express school fees and other monetary variables in INR in the rest of the chapter.

1.3.2 The policy: RTE 25 percent mandate

India's Right to Education (RTE) Act, 2009 is a landmark legislation that guarantees universal access to elementary education and ushers in several far-reaching changes in the elementary education system of the country. The act has drawn most attention and controversy for the section 12 (1) (c), which mandates that all private schools (except minority institutions) reserve 25 percent of places in their entry level grades for children from disadvantaged households: both social and economic. Social disadvantage is based on caste: Schedules Castes (SC), Scheduled Tribes (ST) and some socioeconomically backward castes are eligible. Economic disadvantage is means-tested and state governments are to set income eligibility cutoffs. Families with income below the defined cutoff become eligible for the program. Government reimburses the tuition fee to private schools, thus subsidizing the education of program beneficiary children.[6]

While school choice, in a market sense, was available to the poor in India for the last decade or so, thanks to the growth of a vibrant private sector providing low-fee schooling, poverty constrained the exercise of that choice (Härmä, 2011). The mandate, hence, makes school choice a reality for the poor by relaxing the binding income constraint.

Choice introduced through the mandate is a unique policy intervention and should be differentiated from traditional choice models like voucher

[6] The fee reimbursement to private schools is at a rate that is the lower of the actual amount charged from non-poor children (75 percent) by the school or the per child expenditure incurred in government schools (Sarin et al., 2016). The typical reimbursement rate is around the median private school fee. So private schools that charge above the median fee incur a revenue loss through participation in the program. Further, government regulation makes it difficult for schools to increase or decrease the number of places. Hence high-fee schools have opposed the mandate.

programs. A typical voucher program provides vouchers of a specified value and thus limits the choice set of income constrained families (who cannot top up the voucher) to private schools whose tuition fee is equivalent to the value of the voucher. For instance, in the Chilean voucher program, 6-7 percent of elite schools did not participate as their fee was much higher than the value of the voucher (Hsieh and Urquiola, 2006). In contrast, the RTE free places are available in every private school and hence the school choice set, subject to the neighborhood criterion, is much larger.[7] However, the limitation of the RTE model is that it is not universally available even within the eligible groups, given the cap of 25 percent on places and that the targeted population of socially and economically disadvantaged groups is larger than available places. Kingdon (2017, p. 31) estimates that the number of free places is less than 10 percentage of the population of eligible children.

The goals of the 25 percent mandate are two fold. First, to desegregate Indian schooling system and to create inclusive learning environments for children from different backgrounds to share interests and knowledge on a common platform (Sarin et al., 2016, p. 15). Second, to provide better quality education for children from disadvantaged backgrounds. Educationalists and public intellectuals assert the primacy of the desegregation goals, while politicians, policy makers, economists and choice activists assert the centrality of the improved learning goal. The primary focus of this inquiry is also the latter, though I will shed light on the impact of the policy on both student outcomes and on achieving desegregated learning environments.

[7] Families are allowed to apply to more than one school within their neighborhoods.

1.3.3 Implementation of the mandate

Implementation of the RTE 25 percent mandate has become a contentious issue amongst education policy makers and private schools.[8] Front-runner states in implementing the mandate like Karnataka, Rajasthan, and Madhya Pradesh argue that the mandate provides a unique opportunity to deliver better quality education to children from disadvantaged backgrounds. They view it as a new model of service provision in elementary education that could alleviate the capacity constraints faced by government schools, particularly in urban areas.

On the other hand, states like Andhra Pradesh and Telengana are not implementing the mandate despite the legal requirement. Policy makers from non-implementing states opine that the mandate will have a two pronged negative impact. First, it would divert limited government resources to the private sector and weaken the government schooling system, and second, it could lead to some kind of 'elite capture' and policy mistargeting whereby only the better-off among the disadvantaged benefit from the program. These arguments are exactly mirrored in theory and in global policy and scholarly debates on school choice.[9]

[8] High-fee private schools vigorously opposed the mandate viewing it as the state's encroachment on their autonomy and started implementing it only after the courts ruled against them.

[9] Further, state governments that invested aggressively in the government school system over the last decade through upgrading infrastructure, teacher appointment and training are concerned that it could lead to an unsustainable financial burden: paying high salaries to government teachers who cannot be retrenched and paying private schools for 25 percent places.

1.3.4 Theory of change

Though not clearly articulated, supporters of the mandate have the following theory of change in mind.

- There is a near universal preference for private schooling (Härmä, 2011). So the poor attend public schools not by choice, but because of their inability to pay for private education.
- The 25 percent mandate, by targeting disadvantaged groups, would open the doors of high-quality private schools to children from poorer backgrounds, provide access to better quality education and consequently, improve learning outcomes for these children.
- Enrollment of 25 percent of poorer children in private schools will also desegregate the schooling system and lead to social benefits.

It is therefore expected that the 25 percent mandate would move disadvantaged/income-constrained children from low-quality (government/low-fee private) to high-quality (high-fee/elite private) educational environments, and thus improve their learning outcomes. The two underlying assumptions of this theory of change are that of a private school learning advantage (private schools produce better learning than government schools), and the elite school learning advantage (high fee is correlated with high quality learning). A detailed pictorial description of the theory of change underlying the mandate is at Appendix A.1.1.

I follow Srivastava (2008) and define low-fee private schools as those with monthly fees less than the daily minimum wage. There is, however, no agreed upon definition in the literature on what constitutes a high-fee school.

For the analysis in this thesis, I define all schools charging more than the 75^{th} percentile of the fee distribution as high-fee/elite private schools.

1.3.5 Existing evidence on the mandate

Since states started implementing the mandate in 2012, it has become the focus of inquiry by the media and to a limited extent by practitioners and academics. The focus of the media has been on stories highlighting the travails of parents and children going through the admissions process and the experiences of children enrolled at free places. Academics and practitioners have written on the administrative mechanisms in the implementation of the mandate (Sarin et al., 2016), the institutional and ideological debate surrounding it (Srivastava and Noronha, 2014; Juneja, 2014) and on its early implementation (Noronha and Srivastava, 2013).

There have been calls to evaluate the impact of the policy on childrens' outcomes. The State of the Nation Report on RTE (Sarin et al., 2016, p. 13) notes that "opinion is divided on the impact of the mandate on children themselves," with claims of positive and negative impacts made by enthusiasts and skeptics of the mandate respectively. The report adds that despite the empirical nature of these claims, there is no large-scale empirical work informing them. Muralidharan and Sundararaman (2015) also comment that the implementation of the RTE act provides a fertile ground for research to understand education markets in low-income settings. Against this background, I seek to investigate if expanded choice under the RTE 25 percent mandate has impacted student outcomes. Though learning outcomes measured by test scores are the primary outcome variables of the analysis, I also focus on psychosocial outcomes as they are welfare indicators in their own right and play an instrumental role in learning (Singh, 2015).

1.4 Research design

1.4.1 Experiment

Demand for RTE free places is much higher than supply necessitating random allocation of places through school lotteries. The RTE Act itself provides for lottery based allocation, deeming it a fair way of filling places in cases of oversubscription. These lotteries divide the policy participants into two groups: a treated group that has been given the policy benefit and a control group that has not. The control group of lottery losers creates the perfect counterfactual and simulates the behavior of policy participants in the absence of the policy. Comparing the outcomes of treatment and control children should, therefore, provide unbiased evidence of the impact of expanded school choice on childrens' outcomes.

1.4.2 Setting

My research is located in Karnataka, a large south Indian state.[10] The state government is enthusiastic about the potential of the mandate to improve learning outcomes and is therefore investing significant administrative effort into its implementation. The state started fulfilling the 25 percent mandate from the academic year 2012-13.[11] The State's education department is the nodal agency implementing the mandate. A unique feature of the

[10] Karnataka has a population of 60 million, with 8 million children in the elementary education age.

[11] The school year in Karnataka is from June- March.

implementation of the mandate in Karnataka is that the 25 percent quota is further subdivided amongst the three major eligible disadvantaged groups: 7.5 percent to Scheduled Castes (SCs), 1.5 percent to Scheduled Tribes (STs), and 16 percent to Economically Weaker Sections (EWS). Thus 64 percent of the free places are blocked for income-disadvantaged, while the rest are for the socially-disadvantaged. Eligibility determination is based on certificates issued by grassroots functionaries: a caste certificate for the socially-disadvantaged groups and an income certificate for the the economically weaker sections. Procedures for ascertaining caste are well established in this context, while those for income determination are vague. The income cutoff for being eligible is INR 100,000 (USD 1,546).[12]

For the first three years of implementation, 2012-15, RTE free places were filled through a decentralized application process with the schools playing a key role. Parents were required to physically submit paper applications to schools. Schools, in turn, would scrutinize the eligibility of applicants and conduct admission lotteries in cases of oversubscription. In cases of under-subscription or exact-subscription (number of applications is equal to the number of free places), free places would be offered to all applicants. Though the education department supervised the entire process, there were complaints of non-transparent lotteries and schools subverting the system by rejecting eligible applicants.

The education department introduced a centralized online admissions system from the year 2015-16 to address complaints and to improve the efficiency of the entire admission process. Parents were allowed to choose up to five schools in their neighborhoods. In this new system, parents access the online

[12] The state average per capita income is INR 126,976 (USD 1,940).

application, upload their eligibility documents, choose their neighborhood and select up to five schools giving a preference order. Hence all parents in a neighborhood have the same set of schools to choose from.

Post the applications deadline, the education department does an intensive data cleaning exercise to weed out ineligible applications and duplicates. Following this, the lottery is held on a pre-fixed date in the presence of the media and the political and bureaucratic heads of the education department. The admissions algorithm (described below) generates admission offers and information is sent to parents through text messaging on the same day. Households/ children who have been offered free places are given a time window to claim their spot. As the school year in Karnataka begins in June, the admissions process starts in February/ March and is completed before May.

1.4.3 Lottery

Karnataka's RTE admissions lottery algorithm is based on the random serial dictatorship mechanism (Abdulkadiroğlu and Sönmez, 1998, 2003). A single central lottery is conducted on all eligible applications and admission offers are made as follows:

- Each applicant is randomly assigned a unique 15 character alphanumeric code.
- These codes are arranged in ascending order thus giving a numerical rank to each applicant.
- Every applicant is considered rank-wise and each of her preference is sequentially matched against the availability of places in the particular

school and eligibility group (SC, ST, EWS) combination. A match leads to an admission offer and the applicant becomes a lottery winner; failure to find a match means the applicant becomes a lottery loser.

The lottery winning probability depends on the eligibility category (SC, ST or EWS) and the choice profile (the schools that the parent put down in the application and the preference/ rank order). Combinations of eligibility category and choice profile constitute a randomization stratum. Within randomization stratum treatment (offer of a free place) is random. Hence the lottery losers within each eligibility category and choice profile group provide a clean counterfactual for the lottery winners i.e., the ex-ante probability of admission is identical for applicants within a randomization stratum.

1.4.4 Study design

Given the lottery algorithm, I choose a pair-wise matching design (Bruhn and McKenzie, 2009). I generate matched treatment and control pairs by randomly matching treatment (lottery winners) and control (lottery losers) children within randomization strata. The population of matched pairs is the sampling frame. Two types of applicants get eliminated in this matching scheme: every applicant in randomization strata where all or none of the applicants are treated, and some applicants in randomization strata where the number of treatment and control children are not balanced. The sampling frame is therefore smaller than the overall population of applicants. This has implications only for external validity of the results and does not change their internal validity.

1.4.5 Study cohort and sampling strategy

The sample for this study is drawn from the cohort of children who applied for admission to Grade I in February/ March 2015 and started school in June 2015.[13] I got access to the admissions data in March 2016 after the entire admission process was completed and when the study cohort was already finishing Grade I. Hence I only collected end line data on which this analysis is based.

Karnataka State is divided into 34 districts with significant differences in local education markets (level of private school activity and government school effectiveness), demographics and the number of RTE admissions. For instance, Bangalore Urban district (one part of the capital city) alone accounts for 14 percent of all applications. To explore possible heterogeneities in the impact of the program across these administrative units, I chose the sample from four districts, with two of them belonging to the same region: Bangalore Urban and Bangalore Rural[14] districts from the more developed southern region of the state and Bellary and Gulbarga districts from less developed Northern Karnataka region. Appendix figure A.1.2 is a map of Karnataka with the sample districts identified.

Table 1.1 (panel A) provides a summary of the 2015-16 Grade I admissions database for Karnataka.[15] In total 126,728 children applied for admission into Grade I, of whom about 62,046 (49 percent) got an offer of admission. Panel

[13] The age of entry into Grade I is typically 6 years and the school year in Karnataka is from June to March.

[14] Bangalore Rural comprises suburbs of Bangalore city, Karnataka's capital, and is predominantly an urban area. The rural tag is purely administrative; the district is rural only relative to the heavily urban Bangalore city.

[15] RTE admissions happen at entry grades, which in Karnataka are defined as Lower Kindergarten and Grade I. I study only Grade I applicants, given the difficulty with testing children in Kindergarten.

A summarizes by district, the number of applications and number of treated (offered RTE free place) and control (not offered RTE free place) children. The four sample districts account for 28 percent of the total applications received and 25 percent of all the matched pairs in the state. Panel B presents the proportion of total applicants who claimed social and income disadvantage for participation in the mandate. The proportion of income-disadvantaged claimants for the whole state is around 69 percent, while that in the sample districts is 65 percent.

Table 1.1: Summary of 2015-16 Grade I applications

Panel A: District-wise applicants					
	Applicants	Percentage of total	Treatment	Percent treated	Matched pairs
All 34 districts	126,728	100.0	62,046	49.0	25,123
SAMPLE DISTRICTS					
Bangalore Rural	2,893	2.3	1,150	39.8	759
Bangalore Urban	17,362	13.7	8,819	50.8	2,436
Bellary	9,082	7.2	4,178	46.0	2,022
Gulbarga	6,538	5.2	4,296	65.7	1,076
Total	35,875	28	18,443	51.4	6,293

Panel B: District -wise applicants by eligibility criteria					
				Percent of	
	Applicants	Socially disadvantaged	Income disadvantaged	Income disadvantaged	Socially disadvantaged
All 34 districts	126,728	39,617	87,111	68.7	31.3
SAMPLE DISTRICTS					
Bangalore Rural	2,893	820	2,073	71.7	28.3
Bangalore Urban	17,362	5,760	11,602	66.8	33.2
Bellary	9,082	3,616	5,466	60.2	39.8
Gulbarga	6,538	2,217	4,321	66.1	33.9
Total	35,875	12,413	23,462	65.4	34.6

The planned sample size for the study was 1600 children, 800 matched treatment-control pairs, powered to detect a minimum effect size of 0.1 of a standard deviation and to explore heterogeneous effects. The sampling frame was the population of matched pairs (of treatment and control children) from the four sample districts. Despite the differences in districts in the population of eligible children, number of applications and number of matched pairs, the sample of 1600 children is distributed equally across all four districts: 400 children/200 matched pairs per district.

1.4.6 Study timeline

Figure 1.1 presents the study timeline. As mentioned earlier, this research was conceived in early 2016, almost an year after the relevant RTE lottery in April 2015. Finalizing the design and sampling were done in June-July 2016, after which data collection was done from September-December 2016. The final dataset has 1,616 children/ 808 matched treatment-control pairs.

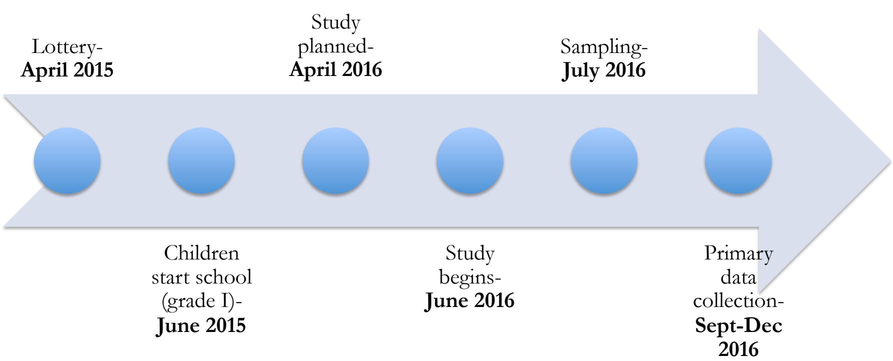


Figure 1.1: Study timeline

1.5 Data

1.5.1 Validity of the design

The study cohort was randomized and assigned to treatment (offer of a free place) almost a year before the start of the research. Therefore, baseline data is unavailable and this entire analysis is based on the endline data. Table 1.2 establishes the validity of the design by presenting treatment-control balance on several observable child and household characteristics. Most of these variables are time-invariant or could not have been impacted by treatment.

Table 1.2: Validity of the design

	Treatment	Control		
	Mean		Difference	p-value
	(1)	(2)	(3)	(4)
Age (in years)	7.33	7.32	0.01	0.62
Gender- male	0.56	0.59	-0.03	0.13
Caste- Scheduled Caste	0.18	0.19	-0.01	0.74
Caste- Other Backward Castes	0.61	0.6	0.01	0.42
Religion- Muslim	0.24	0.24	0	0.89
Mother's age	30.16	30.35	-0.19	0.32
Mother's education (in years)	8.26	8.28	-0.02	0.86
Working mother	0.22	0.24	-0.02	0.34
Father's age	36.69	37.01	-0.32	0.17
Father's education (in years)	8.67	8.62	0.05	0.77
Working father	0.94	0.95	-0.01	0.36
Birth order	1.59	1.63	-0.04	0.23
Number of siblings	1.07	1.09	-0.02	0.61
Attended pre-primary school	0.9	0.91	-0.01	0.65
Asset index	8.92	8.81	0.11	0.52
House ownership	0.5	0.53	-0.03	0.15
Number of rooms in the house	3.01	3	0.01	0.77
N	808	808	1616	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Table presents the treatment and control means on a range of variables. Column (3) is the difference between means, and column (4) is the p-value on the treatment indicator with the balance variable regressed on treatment and sub-district dummies and with standard errors clustered at the pair level. The p-value for an F-test of joint significance of all the balance variables when treatment is regressed on them and sub-district dummies is 0.80.

1.5.2 Attrition

Although I only collect one round of data, there is attrition in the sample: not all of the children in the initially randomized matched pairs could be located. This happened because I recruited participants for the study post-randomization. After randomly sampling 200 pairs per district from the sampling frame of matched pairs, the survey team tried to trace the applicants' parents using their telephone numbers in the admissions database. However, a significant proportion of telephone numbers had become unavailable/non-functional in the 18-month gap between RTE application time (February 2015) and the sampling time (July/August 2016).[16] To arrive at the final study sample, we moved down the randomized sampling sequence[17] and replaced unavailable pairs with pairs lower in the sequence. The survey team called/contacted 2,942 applicants (1471 matched pairs) to finalize the study sample of 1,616 children (808 matched pairs).

Table 1.3 presents district-wise data on the contacted and study sample numbers. Of the 2,942 children contacted, 1,616 (55 percent) are part of the final sample and are surveyed; 856 (29.1 percent) could not be contacted over telephone- I refer to these observations as 'missing.' I dropped another 470 children (16 percent) despite their willingness to participate as their partner in the matched pair was missing. Given the pairwise matching design and the planned use of pair fixed effects in the analysis, collection of data on a lone member in a pair would have been a wastage of resources. I refer to these observations as 'dropped.'

[16] It is not uncommon for Indians to change mobile phone connections in response to continuously changing tariff plans.

[17] I generated a random sampling sequence of matched pairs in each district (using Stata's seed and runiform commands) and then selected the first 200 pairs to be part of the sample. Moving down the sampling sequence involved going beyond the first 200 pairs.

Table 1.3: Sample construction

DISTRICT	Contacted sample	Study sample	Missing	Dropped
Bangalore Rural	698	400	185	113
Bangalore Urban	699	398	203	98
Bellary	695	420	202	73
Gulbarga	850	398	266	186
Total	2,942	1616	856	470
Percent (of 2,942)		54.9%	29.1%	16.0%

The shift from the contacted sample of 2,942 children to the study sample of 1,616 children constitutes attrition: missing and dropped observations are the attritors. The attrition rate, however, is not differential on treatment as both the contacted and study samples are balanced. To check for differential attrition, I regressed an indicator for being in the study sample on all available covariates and covariates interacted with treatment status. I am unable to reject the joint null on an F-test for this regression. I also check if there is any difference in observable characteristics of attritors by treatment status. Table 1.4 shows the regression results for three sub-samples: missing observations, dropped observations and for both missing and dropped observations. These results, demonstrating the balance of attrition on all observable characteristics both individually and jointly for the attritors, read together with the strong treatment-control balance shown in Table 1.2, make it unlikely that the study sample is imbalanced on unobservables that affect outcomes.

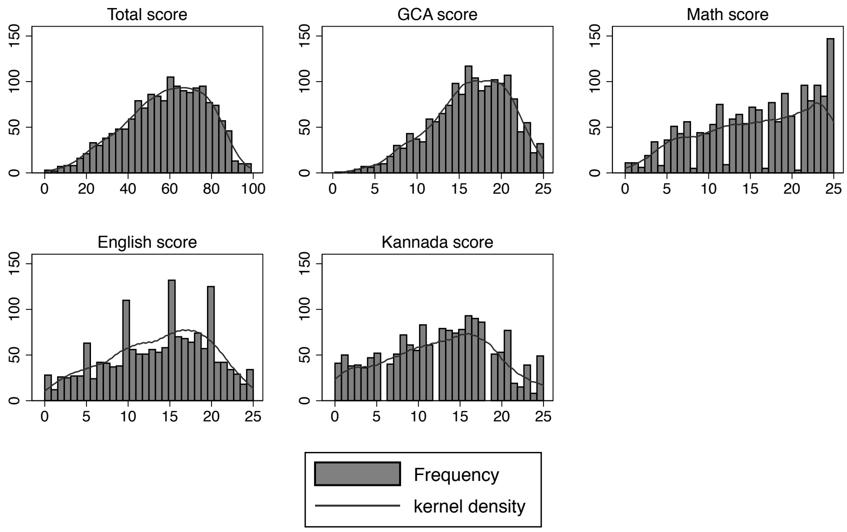
Table 1.4: **Differences in observable characteristics between attritors**

	(1)	(2)	(3)
Sample	Missing only	Dropped only	Both missing and dropped
		Treatment indicator	
Age (in years)	0.00	-0.05	-0.01
	(0.02)	(0.04)	(0.02)
Gender-female	-0.03	0.00	-0.01
	(0.03)	(0.05)	(0.03)
Schools applied to	-0.01	0.01	-0.00
	(0.02)	(0.02)	(0.01)
Caste- Scheduled Caste	0.06	-0.09	0.00
	(0.05)	(0.06)	(0.04)
Caste- Scheduled Tribe	-0.06	0.10	0.00
	(0.10)	(0.13)	(0.08)
Bangalore Urban	0.06	-0.11	0.00
	(0.06)	(0.07)	(0.04)
Bellary	-0.01	0.05	-0.00
	(0.05)	(0.08)	(0.04)
Gulbarga	-0.06	0.06	-0.00
	(0.05)	(0.06)	(0.04)
Constant	0.46***	0.89***	0.60***
	(0.16)	(0.27)	(0.14)
Observations	848	471	1,319
F-stat (without district dummies)	0.678	1.099	0.119
F-stat (model)	1.081	1.758	0.0770

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variable in all regressions is an indicator for being treated. Regression in column 1 uses 'missing' observations only, that in column 2 uses 'dropped' observations and that in column 3 uses both missing and dropped observations. These results show there is no difference in observable characteristics between attritors by treatment status

1.5.3 Outcome Variables

I collected data on two sets of outcomes for all the sample children: test scores and attitudinal (psychosocial) outcomes. Test score data were collected by administering an age appropriate 90-item test that assesses students' ability on four broad skills: General Cognitive Ability (GCA), Mathematics, English, and Kannada (the local language). It was administered one-on-one, at school, on all sample children, taking care that the same enumerator tested the treatment and control children in a matched pair.[18] All four of my test score measures and the total score measure (unweighted sum of the four competency scores) display considerable variation. Figure 1.2 shows their distribution.



Notes:
1.Total score is the sum of the individual scores
2.GCA- General Cognitive Ability
3.N= 1616

Figure 1.2: Histogram of test scores

[18] The testing tool was developed in collaboration with the Center for Early Childhood Development (CECED), New Delhi, and the education department of the Government of Karnataka. The tool is based on the School Readiness Instrument (SRI), developed and validated by the World Bank.

Table 1.5 presents a summary of the raw test scores.

Table 1.5: Summary of test scores

Test	Mean	Std.Dev	Min	Max
General Cognitive Ability (GCA)	16.3	4.6	0.3	25.0
Mathematics	15.8	6.7	0.0	25.0
English	13.6	6.1	0.0	25.0
Kannada	12.5	6.5	0.0	25.0
Total	58.2	19.4	0.0	99.0
Total number of observations-1616				

In addition to test scores, I also collected data on attitudinal items that capture childrens' subjective schooling experience (school experience, interaction with peers, and interaction with teachers) and sense of self-efficacy. The motivation for collecting data on non-cognitive outcomes was two fold. First, Kaufman and Rosenbaum's (1992) 'relative deprivation hypothesis' suggests that admitting ill- prepared poor children into schools where they might feel out of place (due to socioeconomic reasons) would put them at a competitive disadvantage. In a similar vien, Cullen et al. (2006) argue about the importance of match-quality and contend that what is good for one child might not necessarily be good for another. Second, in the context of the RTE 25 percent mandate, there is sizable anecdotal evidence (from media reporting) and a widely accepted popular narrative that children admitted to RTE free places in elite private schools are not well-integrated into their new learning environments. Measuring childrens' responses to attitudinal/ psychosocial items would shed light on whether the relative deprivation hypothesis is in operation in this context and rigorously test the anecdotal claims that RTE beneficiary children are discriminated against.

I developed a psychosocial test based on the instrument used in the Young Lives study (Singh, 2015). My test has 16-binary-response statements that enable the assessment of children on four psychosocial attributes: self-efficacy, peer support, school support and teacher support. Out of 16 statements, only 9 are positively phrased. I recode the yes-no responses to each statement in a way that a one always indicates a better outcome than a zero. Based on these recoded values, I construct four summary indices - self-efficacy index, peer support index, school support index and teacher support index- using two methods. First, I use simple/naïve aggregation by giving equal weight to each statement. In this method, each index ranges from 0-4 with a higher value indicating better outcome on the measure. Second, I follow the inverse covariance weighting approach (Anderson, 2008) and generate indices, each with a mean of zero and standard deviation of one. Table 1.6 presents the summary of the psychosocial measures using both the simple/naïve aggregation and the inverse covariance weighting approach. The indices constructed using the latter approach are used in this analysis.

Table 1.6: Summary of psychosocial/attitudinal outcomes

Panel A-Index construction: simple aggregation (naive approach)				
Index	Mean	Std.Dev.	Min	Max
Self efficacy	3.0	0.8	0	4
Peer support	3.1	1.0	0	4
School support	3.5	0.8	0	4
Teacher support	3.2	0.9	0	4
Panel B- Index construction: Inverse covariance weighting approach				
Self efficacy	0.0	1.0	-4.7	1.2
Peer support	0.0	1.0	-3.6	0.9
School support	0.0	1.0	-6.8	0.5
Teacher support	0.0	1.0	-5.2	0.7

There could be two potential concerns about my outcome variables. First, the average age of the children in my sample is 7.3 years, and one might worry about the reliability of outcomes measured at such a young age. However, evidence shows that test scores measured even at age 7 are significant predictors of both future test scores and labor market outcomes (Currie and Thomas, 1999). Further, the two most cited studies in this literature measure outcomes on children of comparable age groups. While Singh (2015) tests children at ages 5, 8 and 9, Muralidharan and Sundararaman (2015) test them at ages 8 and 10. The second concern is that the outcomes are measured after only 1.5 years post-treatment. It can be questioned if 1.5 years is sufficient time for school effects to manifest. Here again, Muralidharan and Sundararaman (2015) measure impacts both at 2 years and at 4 years post-treatment. Their results broadly remain unchanged between the two test rounds, suggesting that results after 1.5 years of being exposed to treatment are good indicators of the treatment effect.

1.5.4 Independent and control variables

Information on the primary independent variable of interest, the treatment status is available in the admissions database. Information on control variables- household and child characteristics- is collected through a detailed household survey completed with the head of the household.

1.6 Results

1.6.1 Estimation model

I estimate the impact of being offered an RTE free place (i.e., the impact of the policy) on the learning and psychosocial outcomes of children using the following model.

$$Y_i = \beta_0 + \beta_1.T_i + \boldsymbol{\beta}_{Di}.\mathbf{D}_i + \epsilon_i \tag{1.1}$$

Where Y_i is the outcome variable for child i. T_i is a binary treatment variable taking a value of 1 if child i is offered an RTE free place i.e., if the child is a lottery winner. I use sub-district dummies ($\mathbf{D}_i$) to absorb geographic variation and to improve precision.[19] Although use of pair dummies is recommended, given the randomization (Bruhn and McKenzie, 2009), equation (1.1) does not have either pair dummies or pair fixed effects because I lose observations with this approach (especially while estimating heterogeneous effects).[20] However, given the strongly balanced panel, clustering the standard errors at the pair level gives the same results as using pair fixed effects. So equation (1.1) is my preferred estimation model. I will, however, demonstrate the robustness of the main results to the inclusion of controls[21] and pair fixed effects.

[19] The sample is drawn from 4 districts and 28 sub-districts, referred to as 'blocks' in India.

[20] Pair fixed effects only uses within pair variation, so a lot of pairs become unusable for the estimation of heterogeneity.

[21] I use the following control variables and check for robustness of the reults : age of the child, gender of the child, indicator for belonging to Scheduled Caste, indicator for being Muslim, mother's education (in years), indicator for a working mother and two measures of the economic status of the household: a household asset index measure and the number of rooms in the house.

As treatment is randomly assigned, β_1 is an unbiased Intent To Treat (ITT) estimate of the impact of being offered a free place for children whose parents chose to apply to free places. The ITT parameter is not an estimate of the impact of being enrolled into an RTE free place; it is an estimate of the effect of winning the lottery i.e., being offered an RTE free place. Enrollment into a free place is non- random and is determined both by the lottery outcome and parental behavior. For instance, some children who win the lottery might not enroll into a free place.[22] The ITT is, however, the best policy parameter given that parental decisions to apply for RTE places and to enroll their kids when offered admission are not in the direct control of the policy maker (Duflo et al., 2007; Deming et al., 2009). The analysis in this chapter will focus on the ITT estimates, the policy impacts.

1.6.2 Compliance with treatment

Like in most experimental studies, there is a modest non-compliance with treatment in this study. Compliance is defined as being enrollment in a free place if offered admission or not being enrolled in a free place if not offered admission. Non-compliance occurs if lottery winners do not enroll at RTE free places (never-takers) or if lottery losers are enrolled at free places (always-takers). Table 1.7 provides details of the compliance amongst treatment and control groups. That 18.2 percent of the sample are always-takers is puzzling as only the lottery winners should be able to enroll in the free places. Lottery losers' enrollment into a free place could only have happened due to manipulation of the admission system by the grassroots bureaucracy.

[22] The Treatment on Treated (ToT) estimation, which uses the lottery outcome as an instrument for enrollment to an RTE free place, would answer the question of impact of being enrolled at an RTE free place.

Interviews with officials across the bureaucratic hierarchy point to how grassroots functionaries, in collusion with private schools, allotted some of the unclaimed free places to the lottery losers.[23]

Table 1.7: Compliance with treatment assignment

	(1) Total	(2) Enrolled at free place	(3) Not-enrolled at free place	(4) Compliance rate
Treatment	808	723	85	89.50
Control	808	147	661	18.20
	1616	870	746	

Notes: The 85 treated children who did not enroll at a free place are the never-takers, while the 147 control children who did enroll at a free place are the always-takers. Compliance rate for the study (71.3 percent) is the difference between the compliance rate amongst the treated (89.50 percent) and the non-compliance rate amongst the control children (18.2 percent).

The compliance rate for the study, calculated as the difference between the compliance rate of the treatments (89.5) and the percent of always-takers amongst the controls (18.2), is 71.3 percent. This is a reasonably high value. Further, non-compliance and the presence of always-takers and never-takers neither affects the estimation of the ITT nor introduces any bias in the results.

1.6.3 Policy impact on childrens' outcomes

Table 1.8 presents the ITT impact estimates for test scores. The dependent variable, test scores, is converted into standardized z-scores, hence the coefficient estimate should be interpreted as the effect size (standard deviation units). The point estimate of 0.04 in column (1) means that being in the treatment group increases overall test scores by 0.04 standard deviations. The effect is, however, not significant.

[23] Given the setup of the admissions process and the lottery algorithm (where parents are allowed only up to five preferences), it is possible that in some neighborhoods there are lottery losers alongside non-allotted free places (in some schools). This creates an opportunity for grassroots functionaries to assign these free slots to the lottery losers.

The odd numbered columns in the table show the results of the estimation model at equation (1.1), my preferred specification. The even numbered columns show the results for a model that also includes covariates. In both the models the standard errors are clustered at the pair level to account for the pairwise randomization design. None of the treatment effects is significant, demonstrating that the offer of an RTE free place does not have a statistically significant impact on learning outcomes measured across a range of subjects. The sign of the coefficients, however, is positive across all measures. Further, age, being Muslim, being Scheduled Caste and mother's education variables are all statistically significant and have the expected signs.[24]

[24] Existing literature suggests that educational outcomes are positively correlated with age and mother's education and negatively correlated with belonging to socially disadvantaged groups, in this case, being SC or Muslim (Basant, 2007; Magnuson, 2003).

Table 1.8: Policy impact on test scores (ITT estimates)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables	Total		GCA#		Math		English		Kannada	
Treatment	0.04	0.03	0.01	0.01	0.04	0.04	0.03	0.02	0.03	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Age (in years)		0.29***		0.32***		0.28***		0.22***		0.15***
		(0.04)		(0.03)		(0.04)		(0.03)		(0.03)
Gender (boy=1)		-0.07		0.09*		0.01		-0.12***		-0.15***
		(0.04)		(0.05)		(0.05)		(0.05)		(0.04)
Scheduled Caste		-0.37***		-0.21***		-0.26***		-0.32***		-0.30***
		(0.07)		(0.07)		(0.07)		(0.07)		(0.06)
Muslim		-0.45***		-0.33***		-0.25***		-0.17***		-0.70***
		(0.06)		(0.06)		(0.06)		(0.06)		(0.06)
Mother's education		0.03***		0.03***		0.02***		0.05***		0.01
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Working mother		0.00		-0.01		0.03		-0.09		0.10*
		(0.06)		(0.06)		(0.06)		(0.06)		(0.06)
Asset index		0.01		0.01		0.00		0.02**		-0.00
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
House size		0.01		-0.01		0.00		0.03		0.00
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
Constant	0.05	-2.15***	-0.48***	-2.92***	0.29**	-1.79***	-0.24	-2.13***	0.39***	-0.59**
	(0.15)	(0.32)	(0.12)	(0.30)	(0.13)	(0.32)	(0.20)	(0.33)	(0.12)	(0.28)
Observations	1,616	1,616	1,614	1,614	1,614	1,614	1,614	1,614	1,614	1,614
R-squared	0.10	0.20	0.11	0.18	0.10	0.16	0.08	0.17	0.15	0.25

#-GCA is General Cognitive Ability

Notes: *** p<0.01, ** p<0.05, * p<0.1. Regression in columns 1, 3, 5, 7 and 9 are models without covariates, where the outcome variable is regressed on treatment dummy and sub-district dummies. Regressions in columns 2, 4, 6, 8 and 10 have sub-district dummies and controls. Standard errors are clustered at the pair level in all models to account for pairwise matching design. The dependent variables, test scores, are standardized z-scores; hence the coefficient estimate is the effect size. Gender, Scheduled Caste, Muslim and working mother are indicator variables. Age, asset index, mother's education (in years) and house size (number of rooms in the house) are continuous.

Table 1.9 presents the ITT estimates of the impact of policy on the four measured psychosocial outcomes. The self-efficacy index is positive and statically significant across specifications. The point estimate of 0.12 in column (2) indicates that winning the lottery increases the sense of self-efficacy by about 0.12 of a standard deviation (σ). Given that I am testing four indices, I correct for multiple hypothesis by computing sharpened q-values, which are adjusted p-values controlling for False Discovery Rate (FDR) (Anderson, 2008; Benjamini et al., 2006). The q-values are shown in brackets.[25] The treatment effect on self-efficacy is statistically significance even after correcting for multiple hypothesis.

The result on the self-efficacy index alone, unfortunately, does not point to any definitive inference; but the fact that the self-efficacy index is statistically significant and that there is no significant difference on the other three measures (school experience, peer support, and teacher experience) is strong proof against the anecdotal claims and popular views articulated in the mass media that children admitted into free places are at a psychosocial disadvantage due to discrimination/ lack of integration at schools.

[25] I calculate q-values using the code provided by Anderson (2008) online.

Table 1.9: Policy impact on psychosocial outcomes (ITT estimates)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-efficacy		Peer support		School experience		Teacher support	
Treatment	0.12**	0.11**	0.03	0.03	0.03	0.03	-0.03	-0.03
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
	[0.07]*	[0.09]*	[0.67]	[0.70]	[0.67]	[0.70]	[0.67]	[0.70]
Age (in years)		0.09**		0.02		0.02		0.05
		(0.04)		(0.04)		(0.04)		(0.03)
Gender (boy=1)		-0.07		0.02		0.02		-0.01
		(0.05)		(0.05)		(0.05)		(0.05)
Scheduld Caste		-0.13*		-0.11		0.07		-0.15**
		(0.07)		(0.07)		(0.06)		(0.07)
Muslim		-0.26***		0.07		0.02		0.09
		(0.06)		(0.06)		(0.06)		(0.06)
Mother's education		0.02**		0.00		-0.00		0.01*
		(0.01)		(0.01)		(0.01)		(0.01)
Working mother		-0.02		-0.04		-0.10		-0.00
		(0.07)		(0.07)		(0.06)		(0.06)
Asset index		-0.01		-0.01		-0.00		-0.01**
		(0.01)		(0.01)		(0.01)		(0.01)
House size		-0.00		-0.01		-0.02		0.02
		(0.03)		(0.02)		(0.03)		(0.02)
Constant	-0.16	-0.79**	-0.06	-0.12	-0.71***	-0.76**	-0.70***	-1.07***
	(0.13)	(0.31)	(0.11)	(0.31)	(0.23)	(0.32)	(0.22)	(0.33)
Observations	1,553	1,553	1,577	1,577	1,585	1,585	1,557	1,557
R-squared	0.05	0.07	0.12	0.12	0.15	0.16	0.17	0.18

Notes: *** p<0.01, ** p<0.05, * p<0.1. Regression in columns 1, 3, 5 and 7 are models without covariates, where the outcome variable is regressed on treatment dummy and sub-district dummies. Regressions in columns 2, 4, 6 and 8 have sub-district dummies and controls. Standard errors are clustered at the pair level in all models to account for pairwise matching design. The dependent variables, psychosocial outcomes, are standardized z-scores; hence the coefficient estimate is the effect size. Gender, Scheduled Caste, Muslim and working mother are indicator variables. Age, asset index, mother's education (in years) and number of rooms in the house are continuous. Sharpened q-values that correct for multiple hypothesis testing are in brackets for the treatment variable.

1.6.4 Heterogeneous effects

In this section, I investigate if the impact of policy on childrens' outcomes is heterogeneous across four child characteristics: gender, religion, caste, and policy eligibility category (income-disadvantaged versus socially-disadvantaged). The choice of the characteristics is motivated both by existing literature and the context of the policy.

- Gender: There is large literature demonstrating differential household allocation of resources by gender (White et al., 2016; Zimmermann, 2012).
- Religion: Muslims are one of the most economically disadvantaged groups in India (Muralidharan and Sundararaman, 2015); hence I investigate the impact of the policy on Muslims vis-à-vis non-Muslims.
- Caste: Disadvantage based on caste is well- documented in the Indian context. Hence I compare Scheduled Castes (SCs) against non-Scheduled Castes.
- Eligibility criteria: As explained before, policy beneficiaries are either socially or income-disadvantaged. The majority of the current welfare programs target the socially-disadvantaged; the RTE mandate is one of the first large social policies that explicitly supports the income-disadvantaged, hence making eligibility criteria an interesting dimension to explore heterogeneity.

I estimate heterogeneous effects using equation (1.2) by introducing an interaction term between the child characteristic (CC) and the treatment dummy in the ITT estimation model. The coefficient on the interaction term

β_3 is the parameter of interest.

$$Y_i = \beta_0 + \beta_1.T_i + \beta_2.CC_i + \beta_3.T_i * CC_i + \boldsymbol{\beta}_{Di}.\mathbf{D}_i + \epsilon_i \tag{1.2}$$

The results are presented in Table 1.10 for both test scores and psychosocial outcomes. Each cell is a separate regression and the reported value is the coefficient (β_3) and standard error of the interaction term. The interaction term is the difference in the treatment effect between students identified by the covariate and others. The results are positive and statistically significant on one measure for girls (GCA score) and negative and significant on two measures for Muslims (Kannada score and peer support index).

Table 1.10: Heterogeneous impacts by student characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Test scores					Psychosocial outcomes			
	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
Treatment*covariate									
Girl	0.08	0.17*	0.05	0.06	-0.01	0.06	0.08	-0.12	0.06
	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)
Muslim	-0.16	-0.15	-0.11	-0.10	-0.17*	-0.03	-0.29***	-0.02	-0.11
	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)
Scheduled Caste(SC)	0.08	0.10	0.14	0.04	-0.00	-0.03	0.18	0.09	-0.12
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.11)	(0.13)
Income disadvantaged	-0.05	-0.10	-0.13	-0.06	0.10	-0.05	-0.06	0.03	0.16
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.11)	(0.11)
Observations	1,616	1,614	1,614	1,614	1,614	1,553	1,577	1,585	1,557

Notes: *** p<0.01, ** p<0.05, * p<0.1. Each cell is a separate regression, the point estimate and the standard error on the interaction term are reported here. All regressions include sub-district dummies; standard errors are clustered at the pair level

To further unpack these differential effects and to understand the treatment effects for different sub-groups, I estimate equation (1.3), the results of which are presented in Table 1.11. It has four panels of two rows each with every panel showing the result for one child characteristic (CC). For instance, the first two rows present the effects for child characteristic gender, with CC being girl: the first row shows the coefficients γ_2 and and the second row γ_3 .

$$
\begin{aligned}
Y_i = \gamma_0 + \gamma_1.CC_i + \gamma_2.CC_i * T_i + \gamma_3.(1 - CC_i) * T_i \\
+\boldsymbol{\gamma}_{Di}.\mathbf{D}_i + \epsilon_i
\end{aligned}
\tag{1.3}
$$

On test scores, the treatment effects are not significant for any sub-group. On psychosocial effects, the self-efficacy index is positive and statistically significant for girls, non- Muslims and non-SCs.[26] This suggests that the positive treatment effect on self-efficacy for the full sample (Table 1.9) is driven by these three groups. The differences between Muslim and non-Muslim children on the peer support index are also striking, though they do not remain significant after correcting for multiple hypothesis testing.

[26] It is only weakly significant for the income and socially-disadvantaged sub-groups.

Table 1.11: Sub-group impacts by student characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total score					Psychosocial outcomes			
	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
Treatment*covariate									
Girl	0.08	0.11	0.07	0.06	0.02	0.15**	0.08	-0.04	0.01
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Boy	0.00	-0.06	0.02	-0.00	0.03	0.09	-0.00	0.08	-0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Muslim	-0.09	-0.10	-0.04	-0.05	-0.10	0.10	-0.19**	0.02	-0.11
	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)
Non-Muslim	0.08	0.05	0.07	0.05	0.07	0.12**	0.10*	0.04	-0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Scheduled Caste(SC)	0.10	0.09	0.16	0.06	0.03	0.09	0.18	0.10	-0.12
	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.12)	(0.11)	(0.10)	(0.12)
Non-SC	0.02	-0.01	0.01	0.02	0.03	0.12**	-0.00	0.02	-0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Income disadvantaged	0.02	-0.02	0.00	0.01	0.06	0.10*	0.01	0.04	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Socially disadvantaged	0.08	0.08	0.13	0.07	-0.04	0.15*	0.07	0.01	-0.14
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)
Observations	1,616	1,614	1,614	1,614	1,614	1,553	1,577	1,585	1,557

Notes: *** p<0.01, ** p<0.05, * p<0.1. Each panel (comprising two rows) is a separate regression, the point estimate and the standard error on the interaction terms are reported here. All regressions include sub-district dummies; standard errors are clustered at the pair level

1.7 Mechanism analysis

1.7.1 Understanding non-impact

The evidence demonstrates that the policy (offer of an RTE free place) did not have any impact on childrens' outcomes. In this section, I explore the various mechanisms driving the null effect. Given the theory and policy context, there are three potential candidate explanations:

- *General equilibrium effects:* As discussed in the theory section (at page 41), the 25 percent mandate may have improved the overall productivity of the education system, which led to the null results i.e., the schools attended by both the treated and control children responded to competition and improved their academic effectiveness leading to no difference in test scores between the two groups of children.
- *Household input substitution:* Glewwe and Muralidharan (2016) and Das et al. (2013) argue that null impacts of education policies could result when households reoptimize their resources in response to the policy. In this context, household re-optimization could occur if, for instance, parents of lottery-winning children reduce their investments in private tuition and/or in time spent on learning at home or if lottery-losing parents increase their investments.
- *Revisiting the theory of change:* As detailed in section 1.3.4, the theory of change underlying the 25 percent mandate is that it creates a ladder of opportunity for children from disadvantaged/income-constrained households, moves them from low-quality to high-quality educational

settings and thus improves learning outcomes. A necessary condition for improvement in learning outcomes is that there should be a qualitative change in the learning environments: children in the treatment group should be attending qualitatively superior schools than those in the control group.

I investigate each of these explanations in turn.

1.7.2 General equilibrium effects

I contend that the general equilibrium effects argument is not valid here. First, the mandate became fully operation only from 2013 and I measure outcomes in 2016. Three years is probably not enough time for schools to change their production technologies, say by hiring new teachers or changing class sizes. Second, and more fundamentally, the mandate's design and the Indian context do not generate the 'competition effect' as theoretically envisaged:

- As explained in detail in section 1.2.2, public school budgets are not tied to student enrollment and teacher salaries continue to be paid by the government. Hence the incentive to improve does not exist for government schools.
- Elite/high-fee private schools incur a revenue loss through participation in the mandate as their tuition fee (paid by 75 percent children) is higher than the government subsidy. They are being compelled by the government to participate in the mandate and have no inventive to improve outcomes to attract more free-places applicants.

- Low-fee private schools are paid exactly their fee amount for the 25 percent children. They, theoretically, have some incentive to improve performance, assuming they were not at 100 percent capacity prior to the mandate. However, the cost of improving quality (for everyone) might be higher than the revenue gains from the 25 percent children even for these schools.[27]

To summarize, the 25 percent mandate does not create the 'perform or perish' incentives for any of three broad categories of schools, thus making general equilibrium effects extremely unlikely.

1.7.3 Household input substitution

The education production literature views school inputs, home inputs, child characteristics, and household characteristics as the typical inputs into the production of learning (Glewwe and Kremer, 2006; Glewwe et al., 2011). Studies that evaluate the impact of changes in school inputs (or in the extreme, the school itself) on learning are usually estimating the partial derivative of learning with respect to school inputs holding everything else constant. This estimation approach, however, ignores the real possibility of parents changing home inputs in response to policies that change school inputs. Das et al. (2013) demonstrate that households re-optimize resources and inputs in response to policies and argue that naïve impact estimates that do not account for household substitution could potentially be biased. They posit that while the partial derivative is the production-function effect- a technology parameter- the total derivative of learning with respect to school

[27] Schools cannot significantly raise the school fee or size as Government regulation allows only for a 10 percent annual increase.

inputs, accounting for household substitution, is the policy effect, an unbiased estimate of the policy. Therefore, household substitution could potentially bias the estimation of the partial derivative and is a mechanism in understanding the total derivative, the policy effect.

In the context of the RTE 25 percent mandate, household re-optimization is a distinct possibility as explained earlier. To investigate this, I estimate treatment-control differences in home inputs to education on the following five variables: an indicator for attending private tuition, annual private tuition fee, time spent by the child at private tuition, on studying at home and on playing at home.[28] Table 1.12 presents the results.

Table 1.12: Policy impact on home inputs to education

	(1) Treatment	(2) Control	(3)	(4)
	Mean		Difference	p-value
Time spent at private tuition	1.15	1.15	0.00	0.91
Time spent on studying at home	2.22	2.27	-0.05	0.15
Time spend on playing at home	3.06	3.08	-0.02	0.75
Attending private tuition (dummy)	0.45	0.42	0.03	0.15
Annual private tuition fee (in INR)	2220.90	2452.91	232.01**	0.04
N	808	808	1616	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Table presents the treatment and control means on five measured home inputs to education. Column (3) is the difference between means and column (4) is the p-value on the treatment indicator with the home-input variable regressed on treatment and sub-district dummies and with standard errors clustered at the pair level. The p-value for an F-test of joint significance of all the five home-input variables is 0.127. The statistical significance of the private tuition fee variable is driven by 7 outlier observations; the significance disappears when the annual tuition fee data is winsorized at the 0.02 level (high only).

There are no statistically significant differences between treatment and control children in the measured home inputs: the p-value for the F-test of joint significance for all the variables is 0.127. Household substitution of

[28] Measuring private tution attendance and fees is important given the primacy of private coaching (referred to as private tuitions) in the Indian context. Parents seem to view private coaching as a necessary complement to schooling. About 26 percent of all primary school aged children are enrolled at private tuitions (Saha, 2016).

resources, therefore, does not seem to be a mechanism driving the non-impact of the policy.

1.7.4 Policy-induced change in learning environments

To understand policy-induced change in learning environments, I first analyze the types of schools attended by lottery winners and losers in terms of public versus private. Table 1.13 presents the policy impact on private school enrollment. The treatment effect on private school enrollment is only 6 percentage points i.e., the difference between the private school attendance rate of treatment and control children is 6 percent points. So if treatment is defined as private school attendance, the compliance rate is a mere 6 percent. Further, the control mean of 0.93 demonstrates that 93 percent of the applicants would have attended a private school even in the absence of the policy. This inference stems from the fact that the control group provides a counterfactual for a world sans the policy.

Table 1.13: Policy impact on private school enrollment

	(1)	(2)
	Enrolled at private schools	
Treatment indicator	0.06***	0.06***
	(0.01)	(0.01)
Constant/ Control mean	0.93***	0.93***
	(0.01)	(0.07)
Observations	1,616	1,616
R-squared	0.01	0.07
Sub-district dummies	No	Yes
Controls	No	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the pair level. The treatment effect of 0.06 and the control mean of 0.93 imply that 93 percent of the control children and 99 percent of the treatment children attend private schools.

The result that majority of the lottery winners and losers are attending private schools, however, does not mean that there is no difference in the quality of schools attended by lottery winners and losers. As explained in section 1.3.1, there is large heterogeneity in private schooling markets, which is well captured by the school fee measure.[29]

Table 1.14 provides a snapshot of the variation in the private school markets both within and across sample districts. As can be seen, private schooling is available for as little as INR 4,000 at the 10^{th} percentile (p10) of the fees distribution and can go up to INR 32, 761 at p90. Further the highest fee can be 10 or even 25 times the p90 value.

Table 1.14: Summary of school fee data

District	(1) Number of schools	(2) Points on the fee distribution (in INR) Average	(3) Maximum	(4) Median	(5) p10	(6) p90
Bangalore Rural	183	17,087	274,771	13,269	6,011	26,769
Bangalore Urban	1,419	25,346	897,626	14,460	6,925	47,241
Bellary	454	10,625	121,527	8,737	4,196	16,752
Gulbarga	490	9,211	112,050	7,782	1,993	15,706
Total	2,546	19,022	897,626	12,059	4,373	32,761

Notes: All fee values are in Indian Rupees (INR). p10 and p90 are tenth and ninetieth percentiles values of the fee distribution. Fee data is available only for 85 percent of the registered private schools in the Government fee database.

All this indicates that there is a substantial proportion of low-fee schools that even the poor can probably afford.[30] Therefore, despite the majority of the sample children being default private school goers, the 25 percent mandate

[29] In the absence of a more reliable measure of school quality like standardized test scores, I use annual school fees to proxy for school quality as it captures parents' willingness to pay for education and other services provided at school.

[30] Low-fee school in Karnataka context can be defined as schools that charge fee around INR 5,000 or less, which is at the 15^{th} percentile of the fee distribution. This comes from Srivastava (2008) definition of low-fee schools as those with a monthly fee less than the daily minimum wage: the daily minimum wage in urban Karnataka is INR 300-500.

should have improved the quality of schools attended by the lottery winners vis-à-vis losers by moving the winners to higher-fee private schools.

Hence I examine the treatment-control difference in the annual school fee. The fee data is accessed from the Karnataka education department. There is one caveat to this analysis. Fee information is not available for 157 of the 1,616 children in the sample (about 10 percent of the observations). This absence of the fee data could bias the estimates if it is not orthogonal to treatment. Hence I regress a dummy for missing fee on treatment status and rule out any correlation between treatment and missingness.

The Cumulative Density Function (CDF) plots of school fee for treatment and control children is at Figure 1.3. The control CDF first-order dominates the treatment CDF establishing that across the distribution of school fee, treatment children, on average, attend higher-fee schools than control children. However, the gap between the two CDFs is relatively narrow suggesting that the difference is not meaningful from a policy perspective. The same story can be seen in Figure 1.4 where the school fee kernel densities of treatment and control groups are plotted. The right side of the distribution of the density functions clearly show that in the fee range of INR 10,000- 25,000, there are more treatment than control children. The gap between the two curves is again too small to be meaningful.

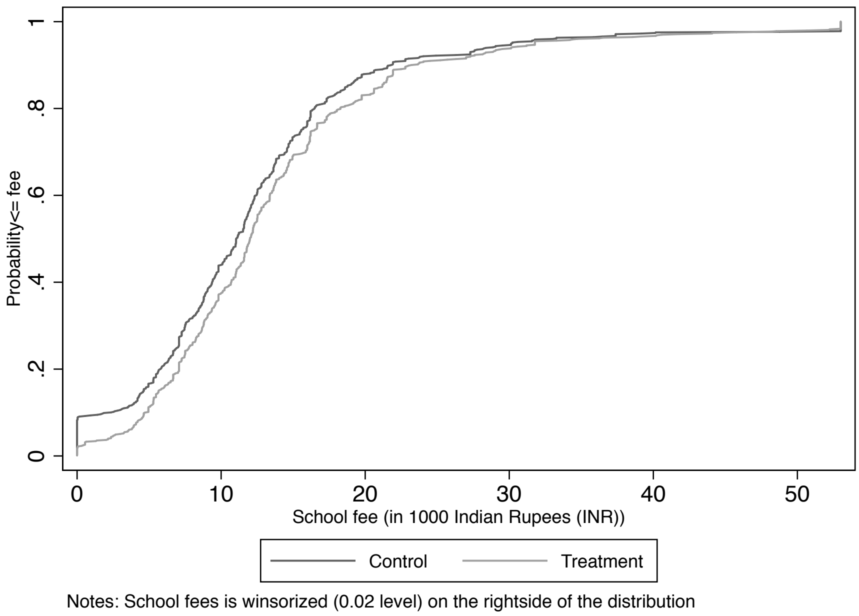


Figure 1.3: CDF of school fee: Treatment versus control

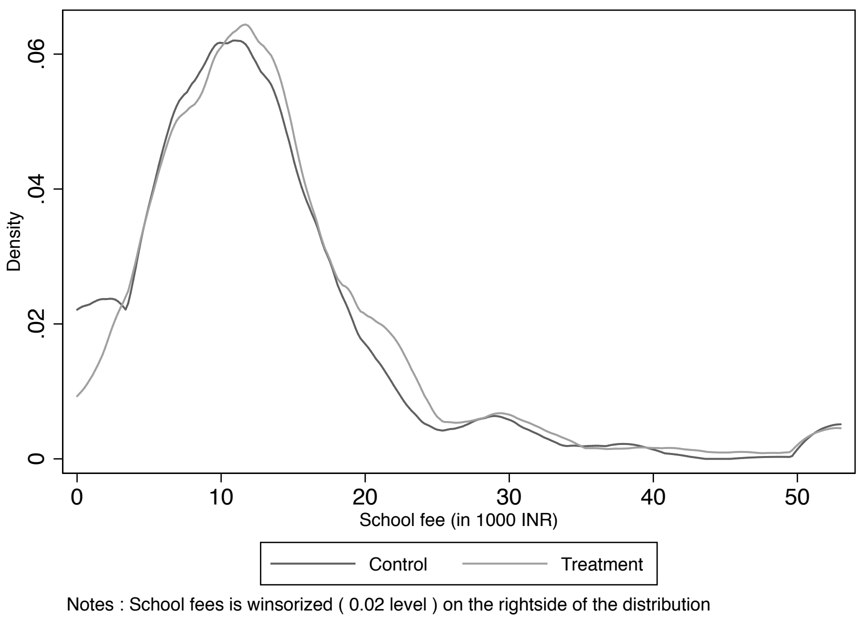


Figure 1.4: Density functions of school fee

Table 1.15 presents regression results estimating the impact of winning the lottery on school fees. While columns (1) and (2) present the ITT estimates for the raw fee data, columns (3) and (4) present results for the winsorized fee data.[31] My preferred results are those with the winsorized data, where I am certain that the results are not driven by extreme outliers in the fee-distribution. The IIT treatment effect is a statistically significant 1,342 (Column (3)). This means that being a lottery winner is associated with attending a school whose average fees is higher by INR 1,342. This is an effect of 0.14 σ (the mean of the winsorized fees is INR 13,166 and the standard deviation is 9,717). So the policy did move treated children to higher-fee schools, but the effect size is small to be meaningful from a policy or economic perspective.

Table 1.15: Policy impact on school fee

	(1)	(2)	(3)	(4)
	Annual school fees in INR			
	Non-winsorized		winsorized	
Treatment indicator	1,725*	1,663*	1,342***	1,270***
	(920.49)	(922.67)	(372.75)	(373.28)
Constant/				
Control mean	13,039***	9,725***	12,472***	11,499***
	(486.17)	(3,432.12)	(361.20)	(2,627.16)
Observations	1,459	1,459	1,459	1,459
R-squared	0.00	0.10	0.00	0.22
Sub-district dummies	No	Yes	No	Yes
Controls	No	Yes	No	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the pair level in all models. The fee data is winsoriozed at the 0.02 level (high only).

[31] I winsorise the top 2 percent of the fee distribution to account for huge outliers in the fee data and to ensure that outliers do not drive the results.

All this evidence demonstrates that even on the school fee measure the policy has not moved children to significantly different schools. Further the control mean is 12, 472 (Column (3)) demonstrating that the average child in the control group, and hence the average applicant in the absence of the policy, attends a school of fee around INR 12,500, the median school fee.

That the policy did not move treated children to better learning environments compared to their default, is therefore, the mechanism explaining the absence of policy impact.

1.8 Discussion

1.8.1 Policy mistargeting

The policy goal of improving outcomes of disadvantaged children has not been achieved due to the failure of a fundamental mechanism inherent in the 25 percent mandate's theory of change. That the treatment effect on private school enrolment is only 6 percent points with a control mean of 0.93, and that the effect on school fee is 0.14 σ indicates that the policy participants/ applicants would have attended broadly similar schools with or without the policy. The default schooling choice for the average policy applicants is a private school with an annual fee of around INR 12,500 (control mean in Table 1.15, Column 3). This value is approximately the median private school fees (Table 1.14, column 4). So the average applicant is a default private school goer, with ability and willingness to pay the fee of the median private school. This description is very different from that of the target applicant who was supposed to be a default government or low-fee private school goer without the ability to afford private school education.

This logically implies one of two things. First, there are no default government/low-fee school going households in the population. Or more plausibly, that the default government/low-fee school going households did not participate in the mandate. Table 1.16 shows the public versus private school attendance proportions in the sample and the population . While only 7 percent of the control applicants are government school goers (93 percent are private school goers), the public school attendance of the population is 41 percent. Hence there is a large proportion of default government school going (poor/disadvantaged) children in the population that has not applied for the policy.

Table 1.16: Comparing private and public school attendance

	Public	Private	Total	Percentage public	Percentage private
Panel A- Sample (by treatment)					
Treatment	13	795	808	1.6	98.4
Control	59	749	808	7.3	92.7
Total	72	1,544	1,616	4.5	95.5
Panel B- Sample(by district)					
Bangalore Rural	15	385	400	3.8	96.3
Bangalore Urban	4	394	398	1.0	99.0
Bellary	28	392	420	6.7	93.3
Gulbarga	25	373	398	6.3	93.7
Total	72	1,544	1,616	4.5	95.5
Panel C- Population (classes 1-8)					
Bangalore Rural	57556	53258	110814	51.9	48.1
Bangalore Urban	106029	569244	675273	15.7	84.3
Bellary	229320	119531	348851	65.7	34.3
Gulbarga	250056	179420	429,476	58.2	41.8
Total	642961	921453	1564414	41.1	58.9

Notes: In three of the four districts the proportion of public school attending children is above 50 percent (Source: Enrollment data from the State education department).

All this points to one definitive conclusion: the target population, default government/low-fee school goers, have not participated in the policy. Majority of the mandate participants are default private school goers who can afford the median school fees. So the program has been mistargeted: the target households have not applied to the program, and infra-marginal households have applied in large numbers. In the targeting literature, these two failures are referred to as errors of exclusion (undercoverage) and inclusion (leakage) respectively (Coady et al., 2004). If the policy were successfully targeted, the point estimates on private school enrollment indicator and the school fees variable would be much higher than the observed values. In the ideal case of perfect targeting the private school effect would be 1, and the fee effect would equal the average school fee of INR 19,022 (Table 1.14, column (2)).

1.8.2 Alternate explanations: John Henry effects

An alternate explanation of the mechanism results could be that the policy is well-targeted, but the RTE admission process and the lottery have altered the behavior of control households. More specifically, lottery-losing households' schooling decisions could have changed just because of participation in the lottery. These type of lottery/experiment-induced behavior changes in the control group are referred to as John Henry effects (Duflo et al., 2007). In the context of the RTE mandate, it can be argued that the lottery-losing households, having gone through the application process and making up their mind on the school they want to enroll their child at, have significantly altered their household spending decisions and reallocated money to the child's education. In that case, the schooling decisions of the control group do not provide the perfect counterfactual.

I agree that there might be some households who would have made this exceptional decision. However, it is hard to imagine that most of the population of lottery-losers have engaged in this behavior. Lottery-induced behavior changes, generally, occur when the control individuals can afford that behavior, say by increasing effort. In the case of the RTE mandate, the behavior change is very expensive and entails reallocation of the household budget for eight years.[32] Credit-constrained low-income households could not have engaged in this behavior. Hence the result that the average control group child is a median-fee private school goer cannot be explained by John Henry effects.

1.8.3 Income profile of policy participants

Having established that the policy is mistargeted, the next interesting question is to get a sense of the degree of targeting failure. I use primary data on household spending on education to estimate the treatment-control spending differences and use this estimate to comment on the income status of the mandate's applicants.

I collected survey data on household spending on education of the sample children. Spending on education involves three broad kinds of costs: first, tuition fee paid to schools; second, other mandatory expenses of education like spending on books, uniforms, transportation and compulsory activities at school; and finally, non-mandatory expenses like spending on private tuitions[33] and other optional activities outside school. I refer to category

[32] The RTE subsidy is for 8 years, from Grade 1-8.

[33] As mentioned before, private coaching/ tuition attendance rates are as high as 26 percent in this context (Saha, 2016).

two and three as non-tuition expenses of private education. Under the RTE mandate, government subsidizes only the tuition fee, leaving households to cover the non-tuition costs. Hence the RTE subsidy is only partial.

Table 1.17 presents the results for the treatment effect on household education expenditure on the sample child. This expenditure involves both tuition and non-tuition expenses. On average, lottery winners who were offered an RTE free place, spent INR 5,610 less than the lottery losers on their child's education. The control mean is 13,159. This implies that household expenditure the child's education was INR 13,159 for the control households and about INR 7,500 for the treatment households. This leads to three inferences. First, despite getting a tuition waiver, treated households spent around INR 7,500 on non-tuition education expenses. Second, the non-tuition expenses are, on average, higher than the tuition costs. Third, RTE applicants have the ability to spend INR 13,159 on their child's education, and thus, the treatment effect of INR 5,610 is a direct transfer to treated households.

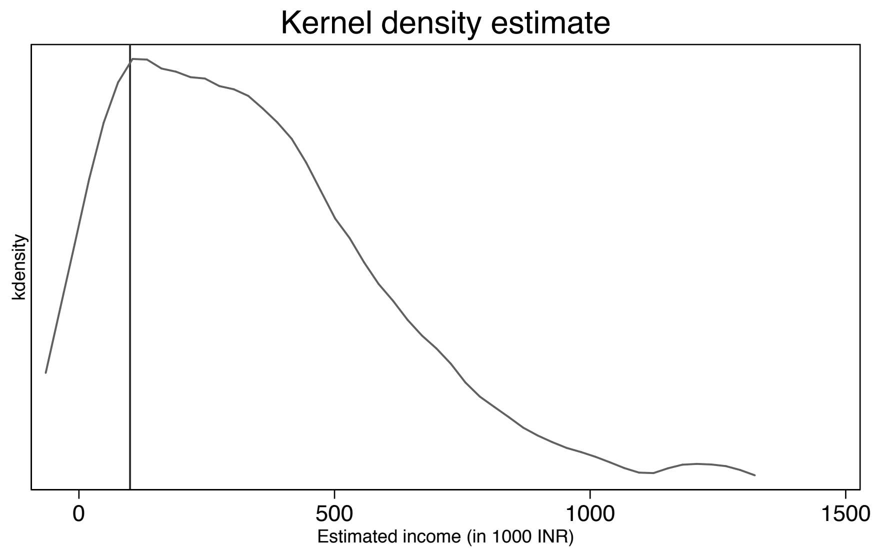
Table 1.17: Policy impact on education expenditure

	(1)	(2)
	Household education expenditure on sample child (in INR)	
Treatment indicator	-5,610***	-5,647***
	(394.31)	(404.72)
Constant/ Control mean	13,159***	6,914**
	(402.93)	(3,001.14)
Observations	1,616	1,616
R-squared	0.08	0.27
Sub-district dummies	No	Yes
Controls	No	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at pair level.

The control group mean of INR 13,159 can be used as a starting point to generate a rough estimate of the income of the applicant households. With an average of two children per household, total household expenditure on education would be about INR 26,320. A national estimate for average percent of household income spent on education in urban India is 7 percent (Tilak, 2009). Therefore the income of an average household spending INR 26,320 on education should be about INR 376,000. This is almost four times the income eligibility cutoff of INR 100,000 in Karnataka and confirms that ineligible households have captured the policy.

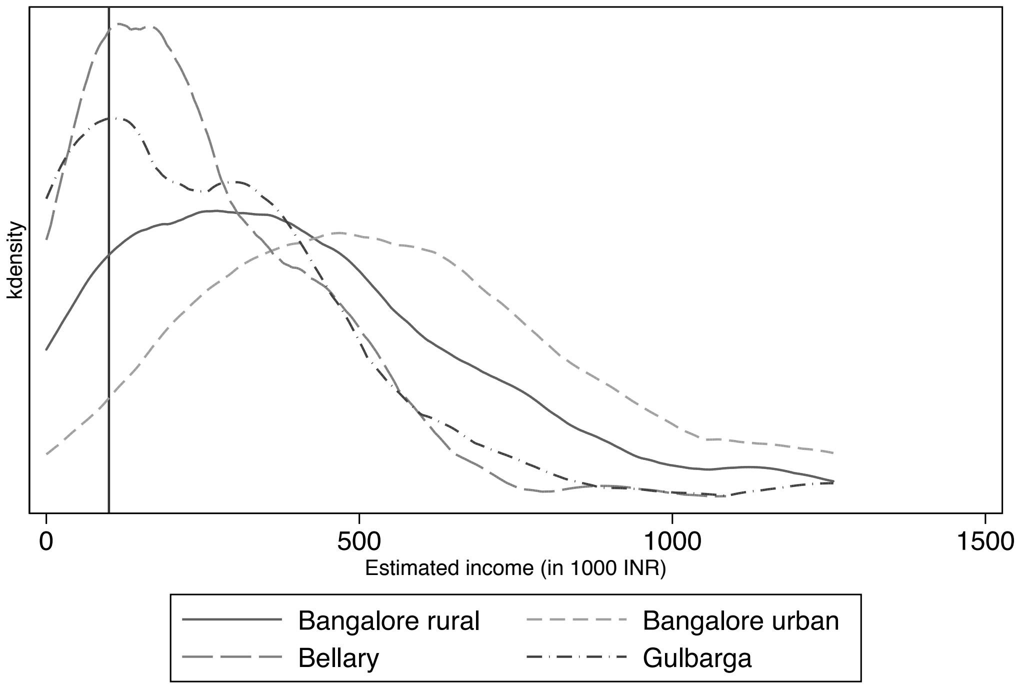
I calculate income estimates for the control group children in the sample using this method and plot the kernel density function for these income estimates. Figure 1.5 shows the distribution along with the income eligibility cutoff line at INR 100,000. Clearly, a huge part of the density of the distribution is above the eligibility cutoff, with a significant percentage of households having incomes 5 times above the cutoff.



Notes:
Estimated income is winsorized (0.02 level) on the rightside of the distribution
The veritcal line at INR 100,000 is the income eligibility cutoff of the program

Figure 1.5: Estimated household income of policy applicants- I

Figure 1.6 plots the income estimates by district. This shows that in the more urban and developed Bangalore area (Bangalore rural and Bangalore urban), the proportion of applicants above the eligibility cutoff is much higher than those in the relatively less-developed Northern Karnataka districts (Bellary and Gulbarga). The failure of targeting, however, is across districts. I acknowledge that this method of estimating income is not very rigorous, but it provides strong suggestive evidence on the income profile of participants and complements the results of Table 1.13 and 1.15.



(Notes:The veritcal line at INR 100,000 is the income eligibility cutoff of the program)

Figure 1.6: Estimated household income of policy applicants- II

1.8.4 Leakages and undercoverage

That the program has been poorly targeted begs two central questions from a policy perspective. First, how are ineligibles able to apply for the program?

Second, why are eligibles not applying for the program? In other words, what explains errors of inclusion and exclusion?

Inclusion errors are a result of deploying an income-based targeting strategy[34] despite the well established challenges with income-eligibility determination (simple means-testing) in developing country contexts (Coady et al., 2004; Banerjee et al., 2009; Alatas et al., 2012). In Karnataka (and in most other states of the country) income-eligibility is determined through a certificate issued by a grassroots government functionary without any independent verification. More often than not, these certificates are issued based on self-declarations by the household head. So the obvious difficulty with establishing income eligibility seems to have opened doors for ineligible households to apply for the program.

Next is the issue of non-participation of eligible households. Anecdotal evidence points to information constraints, transaction costs of application, and partial nature of the RTE subsidy as barriers to participation of the poor. The role of information constraints and transaction costs in excluding the poor from school choice programs has been well documented (Musset, 2012). In this context, the partial nature of the RTE subsidy and the high non-tuition expenses associated with private school education seem to be responsible for exclusion of the poor. As demonstrated in the previous section, the non-tuition costs in my sample are about 1.33 times the tuition

[34] The mandate is one of the first large social policies to use income-based eligibility. Traditionally, the Indian state has used caste-based eligibility (social disadvantage) and proxy-means-testing (defining below poverty line households based on asset and land ownership) approaches for targeting poverty reduction programs. This mandate uses both social and income based targeting. I focus on the income-targeting here as it is the primary targeting mechanism, with 64 percent of the applicants claiming eligibility for being income-disadvanatged. Further, the failure of caste-based targeting owing to capture of benefits by the better off amongst the socially-disadvantaged has been well studied and is an important area of public and political debate in India.

costs. Evidence points to lower socioeconomic status families not benefitting when the government subsidy (voucher) does not cover all the costs of education (Ryan and Watson, 2009). Activists working on the mandate and parents associations have repeatedly raised this issue, pointing out that the high non-tuition expenses could be deterring the most income constrained households from applying to the mandate. This seems a reasonably sound explanation for exclusion errors, though further research is required to precisely understand the binding constraints that prevent poor households from participating in the program.

In summary, the reliance on simple means-testing and the partial nature of the RTE subsidy are the two design flaws in the mandate that are responsible for the failure of targeting.

1.8.5 Cost-effectiveness

Government reimburses the costs of educating the 25 percent children directly to private schools. This cost is fixed as either the lower of the school fee or the average per child expenditure in the government system. The latter amount is INR 11,848 in Karnataka and is around the median of the private school fee distribution.[35] So half the schools (with fee higher than the median value) are reimbursed the ceiling amount of INR 11,848, while those charging fee lees than the median value are reimbursed their actual fee. The average per child reimbursement for the year 2015-16 is INR 6,800[36] for the

[35] The calculation of this amount only uses the recurring costs of running schools (primarily teacher salaries). Other substantial costs incurred by the government like costs of infrastructure, administration and school inputs (text books and free meals) are not included in this calculation. Including all costs increases the per child cost incurred by government to INR 25,500 (Karnataka education budget of INR 165,000 million divided by 6.46 million, the total number of children in government schools from grade 1-10).

[36] Source: Authenticated government data on reimbursements to schools.

entire state: government is, therefore, spending INR 6,800 for procurement of education for disadvantaged children. This amount is only about 60 percent of the cost presently incurred in the government system. Therefore even in the absence of improved test scores, it can be argued that the program is cost-effective as the Government has been successful in procuring same quality education at 60 percent of the cost of state education by leveraging the efficiency of the private sector. This approach to cost-effectiveness is , however, narrow. The calculation only accounts for government spending, and not that of the high-fee private schools, which are co-subsidizing the education at free places.[37] Including the spending by private schools, and hence overall costs to society, will reduce the cost-effectiveness advantage of the 25 percent mandate. So the 25 percent mandate, in its current from, is not as cost-effective as a simple comparison between government spending on the mandate and expenditure on public schools would suggest.

Finally, the treatment effect on school fees is only INR 1,342 (Table 1.15). This means that the government is buying education from private schools worth INR 1,342, by spending 5 times that amount! Crucially, the program beneficiaries are children whose parents can afford private school education. Hence the policy is, in essence, achieving a very inefficient and potentially welfare-reducing income transfer from taxpayers to infra-marginal policy participants.

[37] Half of the private schools charge fees higher than what the government reimburses for free places. These schools are losing revenue through participation in the mandate and are, hence, co-subsidizing the education of free places children.

1.8.6 Spillover effects within beneficiary households

If the net policy effect is to provide an income transfer to unintended households, it is important to evaluate if the transfer led to spillover benefits within beneficiary households. Following the literature on intra-household resource allocation (Das et al., 2013), I investigate if the lottery-winning households invest the income transfer in the education of the other children in the household. Table 1.18 presents treatment effects on household expenditure on all children in the household and all children other than the sample child. As discussed in section 1.8.3, the treatment effect of -5,610 in column (1) implies that treatment households spent INR 5,610 less on the education of the sample child vis-à-vis control households. This saving is the transfer they receive thanks to the program. If this transfer has translated into greater investments on education of other children in the household, the treatment effect on education expenditure of siblings (column (3)) should be positive. However, the estimate is a statistically significant -1,477, implying that treatment households are spending less not only on program participating children, but also on their siblings. An important caveat to the result is that the expenditure data on siblings was not verified (with supporting documents), as was the case for the sample child. Further, treated households had the incentive to under-report expenditures compared to controls in the absence of supporting documents.[38] Hence I treat this result as suggestive evidence pointing to the absence of reallocation of the RTE windfall on education of children within treatment households. A thorough analysis of spillover effects will require

[38] Survey teams consistently reported the discomfort of treated households in reporting expenditures, as several of these households probably knew they were ineligible for the program.

more reliable expenditure data on education and other aspects of human capital development (like health, nutrition and so on).

Table 1.18: Policy impact on education expenditure of siblings

	(1)	(2)	(3)
	Household education expenditure (in INR)		
	Sample child	All children in household	All children excluding sample chdild
Treatment indicator	-5,610***	-6,891***	-1,477***
	(394.31)	(792.01)	(545.01)
Constant/ Control mean	13,159***	23,389***	11,393***
	(402.93)	(741.58)	(450.20)
Observations	1,616	1,616	1,190

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at pair level. Sub-district dummies and controls are not included in the models

1.8.7 Summary of policy implications

I summarize the key policy implications below:

- The RTE mandate did not lead to improved learning outcomes for the policy beneficiaries. This is due to poor program targeting, which caused the non-participation of the most disadvantaged households who could have potentially benefitted from the program.
- The targeting failure owes more to policy design than implementation: the reliance on income-eligibility and the partial nature of the RTE subsidy. The former enabled ineligible households to apply for the mandate, while the latter probably discouraged the eligible.
- Though the mandate appears cost-effective in a very narrow sense, including the overall costs to society reduces the strength of that claim. Government spending on the mandate, in its current form, is an income

transfer from taxpayers to RTE free-place winners.

- The psychosocial treatment effects debunk the prevalent anecdotal claims that children enrolled in free-places are being discriminated against in their new learning environments.

1.9 Replicating the results on another sample

1.9.1 Testing policy mistargeting on a larger sample

Given the study design (pairwise matching) and sampling strategy, the results of this evaluation are valid only for the population of matched pairs from the four districts of Karnataka. Though the four districts account for 28 percent of the RTE applicants (Table 1.1), the results cannot be generalized to the whole state as the districts were not randomly chosen. Further, there is a temporal dimension to generalizing the targeting failure result. It can be argued that the poor targeting performance is primarily due to the nascency of the policy and that participation of eligible households will improve with time. This is the precise argument advanced by policymakers when presented with my results. To address the spatial and temporal concerns on the generalizability of the policy mistargeting finding, I repeated this evaluation on another sample of RTE applicants from the academic year 2016-17 (my primary sample of 1616 children are RTE applicants from the academic year 2015-16, hereafter referred to as sample A. The new sample from academic year 2016-17 is hereafter referred to as sample B). Having conducted an extensive household and child survey on sample A, I realized that the key to evaluating targeting effectiveness was knowing the schools the lottery winners and losers were enrolled at. This information could be obtained through a relatively inexpensive telephone survey. Hence, I repeated the evaluation on a much larger sample: sample B has 15,908 children compared to 1,616 in sample A. The evaluation on sample B, however, does not shed light on the learning outcomes question.

1.9.2 Sampling, survey, and attrition

Table 1.19 presents the summary of the 2016-17 admissions data for the whole state of Karnataka. As mentioned earlier, RTE admissions happen at two entry grades in Karnataka: Lower Kindergarten and Grade I. Though sample A is drawn only from Grade I applicants, I use both the Kindergarten and Grade I applicants for constructing sample B. A total of 184,851 children applied for admission, out of whom 64 percent were offered admission. Like in sample A, 70 percent of the applicants claimed program eligibility owing to being income-disadvantaged.

Table 1.19: Summary of the 2016-17 applications

Panel A: Grade-wise treatment and matched pairs				
Entry grade	Applicants	Treatment	Percent treated	Matched pairs
Lower Kindergarten	107,864	68,520	64	17,294
Grade I	76,987	48,970	64	11,958
Total	184,851	117,490	64	29,252
Panel B: Grade-wise applicants by eligibility criteria				
Entry grade	Applicants	Socially disadvantaged	Income disadvantaged	Percent of income disadvantaged
Lower Kindergarten	107,864	31,334	76,530	71.0
Grade I	76,987	23,218	53,769	69.8
Total	184,851	54,552	130,299	70.5

The 2016-17 RTE admissions process and the place allocation lottery are the same as in the previous year. I generated the sampling frame of 29,252 matched pairs (59,184 children) by using the matching strategy explained in section 1.4.5. This time, however, I deployed a Probability Proportional to Size (PPS) sampling strategy wherein a certain percentage of observations are randomly chosen from each sampling unit. The sampling unit is the sub-district entry grade: with 193 sub-districts and 2 entry grades (Lower Kindergarten and Grade 1) there are 386 sampling units. The survey sample comprised of a randomly selected third of the matched pairs in the sampling frame: 9,612-matched pairs/19,242-children. I conducted a short telephone survey of these applicants in January-February 2017 and collected information on the schools children were enrolled at. As would be expected, some parents did not respond to the survey team's telephone calls. However, the attrition/non-response rate[39] in this sample is much lower than before: only 17 percent compared to 45 percent in sample A.[40] Table 1.20 presents the attrition data that shows the 7 percent point difference in attrition between treatment and control (column (3)).

Table 1.20: Attrition in Sample B

	(1)	(2)	(3)	(4)	(5)
	Randomized sample	Respondents	Attrition rate	Fee Data available	Attrition rate
Treatment	9,621	8,271	14%	6,527	32%
Control	9,621	7,637	21%	4,699	51%
	19,242	15,908	17%	11,226	42%

I obtained information on the second outcome variable for this analysis, school fee, from the education department's school fee database. The attrition problem is more acute for this variable (42 percent) as a large number of

[39] Non-respondents are of three types: those whose telephone numbers were not working, those who were not picking up calls and those who refused to give information.

[40] The higher response rate in sample B owes to the shorter time gap between the RTE admission process and the survey. The gap was 6 months for sample B compared to 18 months for sample A.

schools have not uploaded their information into the government portal. Further the attrition is differential with a 19 percent point difference in attrition rate of treatment and control (Column (5)).[41] I correct for differential attrition using Lee bounds (Lee, 2009).

1.9.3 Results and discussion

Table 1.21 presents the policy impact estimates on the two targeting performance measures: private school enrollment and school fee. The private school effect is 8 percent points, the control mean is 0.91, the Lee bounds for this effect are very tight (lower bound of 8 percent points and upper bound of 9 percent points) and the effect confidence interval does not include zero. This establishes that 91 percent of the applicants in this sample were default private school goers. The fee effect is 2,060 (column (3) winsorized annual school fee variable), and the control mean is 9,513. This implies that the policy has moved children to schools whose fees is INR 2,060 higher, and that applicants come from households who have the ability to pay INR 9,513 (the median school fees for the whole state is INR 9,645). Lee bounds for the fee effect are very wide and the effect confidence interval includes zero (column (6)). This means that the null of zero-impact on fees cannot be rejected. These results confirm the primary inference of policy mistargeting drawn from analysis of sample A. Though the effect sizes in sample B are slightly different, the overall story is unambigious: the majority

[41] The higher attrition rate amongst control children can largely be attributed to the poorer quality of data collected from the control households in the telephone survey. Parents were requested to provide the exact name of the school during the survey; the name and the neighborhood were then used to match schools with the government fee dataset. Control households were generally less forthcoming than the treatment households in providing exact name of the school the child was enrolled at. Not having the exact name led to poorer matching of schools for the control children and hence a higher level of attrition.

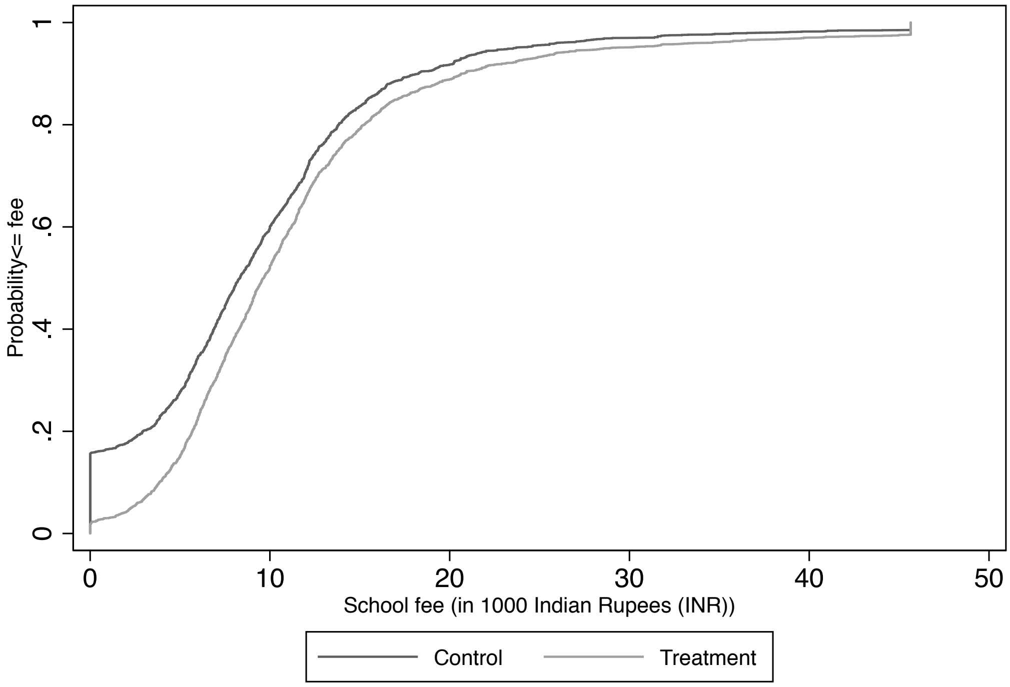
of policy applicants are default private school goers with ability to pay the median school fees.

Table 1.21: Policy impact (ITT estimates): Sample B- I

	(1) Private school	(2) School fee	(3) School fee (winsorized)	(4) Private school	(5) School fee	(6) School fee (winsorized)
Treatment	0.08***	2,253**	2,060***			
	(0.00)	(746)	(143)			
Lower				0.08***	-4,056***	-1,921***
				(0.00)	(653)	(154)
Upper				0.09***	6,007***	4,913***
				(0.00)	(975)	(208)
Constant	0.91***	11,648***	9,513***			
	(0.00)	(646)	(120)			
Observations	15,908	11,226	11,226	19,226	19,226	19,226
Selected observations				15908	11226	11226
Effect CI-lower				0.0710	-5131	-2174
Effect CI-upper				0.0982	7611	5255
Trimming proportion				0.0755	0.279	0.279

Notes: *** p<0.01, ** p<0.05, * p<0.1. Models (1)-(3) report the treatment effect on different outcome variables. The standard errors are clustered at the pair level. Columns (4)-(6) report the Lee bounds for models (1)-(3).

Figure 1.7 shows the treatment and control CDF plots of school fee that demonstrate the lack of a meaningful difference in the fee of schools treatment and control children are enrolled at.



Notes: School fees is winsorized (0.02 level) on the rightside of the distribution

Figure 1.7: CDFs of school fee for Sample B

Table 1.22 presents the results only for Grade I children from the four sample A districts. This subsample of sample B is drawn from the same population as sample A. Hence, comparison of results in Table 1.22 with those from the primary results in Table 1.13 and 1.15 should point to improvements (or the lack of them) in policy targeting across time. The private school effect in Table 1.22 is 4 percent points (compared to 6 in Table 1.13), and the school fees effect is INR 1,478 (compared to INR 1,342 in Table 1.15). The replication of sample A results in this subsample of sample B means that there has been no change in the type of households participating in the policy (default private school goers with ability to pay the median school fees) over the two-year period. This evidence challenges the claim that mistargeting is a result of information constraints alone and that policy targeting would improve with time. The evidence from sample B, therefore, confirms that the mistargeting result is valid for the entire state of Karnataka and illustrates that targeting does not appear to be improving with time. Given this, it is reasonable to argue that the non-impact on test score results can also to be generalized to the whole state of Karnataka.

Table 1.22: Policy impact (ITT estimates): Sample B- II

	(1) Private school	(2) School fee	(3) School fee (winsorized)	(4) Private school	(5) School fee	(6) School fee (winsorized)
Treatment	0.04***	4,465**	1,478***			
	(0.01)	(2,094)	(419)			
Lower				0.03***	-3,719***	-2,639***
				(0.01)	(715)	(495)
Upper				0.05***	8,268***	4,140***
				(0.01)	(2,715)	(634)
Constant	0.95***	13,225***	12,144***			
	(0.01)	(623)	(350)			
Observations	2,755	1,623	1,623	3,363	3,363	3,363
Selected observations				2755	1623	1623
Effect CI-lower				0.022	-4894	-3453
Effect CI-upper				0.0593	12734	5184
Trimming proportion				0.0439	0.219	0.219

Notes: *** p<0.01, ** p<0.05, * p<0.1. These results are for a subsample of Sample B comprising of Grade I children from the four sample A districts. The results are, therefore, comparable to the results of sample A. Models (1)-(3) report the treatment effect on different outcome variables. The standard errors are clustered at the pair level. Columns (4)-(6) report the Lee bounds for models (1)-(3).

1.10 Conclusion

This chapter presents the results of the first empirical investigation of the RTE 25 percent mandate, India's national school choice policy. I take advantage of the lottery-based allocation of RTE free places in Karnataka state and estimate the policy impacts on childrens' learning and psychosocial outcomes. Contrary to expectation of choice proponents, I find no statistically significant difference in the outcomes of RTE beneficiary children relative to the control group. I undertake a detailed mechanism analysis and establish that there is no significant difference in schools attended by treatment and control children and that the average policy applicant is a default private school goer with ability to afford the median private school fee. The implication of my evidence is that the policy is mistargeted. Some of the design elements of the policy make it easier for non-poor households to apply for the policy, while making it harder for low-income parents to apply. Given the importance of the mistargeting finding from the policy design and implementation perspective, I replicate it on a much larger sample of applicant households.

My results and discussion point to several areas for future research. First, the reasons for non-participation of the targeted program beneficiaries need to be better understood. Though I conjecture that the partial nature of the RTE subsidy is the primary culprit, establishing the reasons through collecting survey and qualitative data would be important in improving the policy in the coming years. Second, my results on the spillover effects of the RTE transfer within treated households are rudimentary, at best. Analysis of these effects based on detailed household expenditure data would contribute to understanding the overall welfare effects of the policy. Finally, some Indian states like Madhya Pradesh are using alternate targeting methods to implement the policy. Evaluating the impact of the policy and its targeting performance in these contexts could deepen our understanding of both the design and implementation of this crucial school choice mandate.

References

Abdulkadiroğlu, A. and Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American economic review*, 93(3):729–747.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *The American Economic Review*, 102(4):1206–1240.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Banerjee, A., Duflo, E., Chattopadhyay, R., and Shapiro, J. (2009). Targeting efficiency: How well can we identify the poorest of the poor? *Institute for Financial Management and Research Centre for Micro Finance Working Paper*, 21.

Basant, R. (2007). Social, economic and educational conditions of Indian Muslims. *Economic and Political Weekly*, pages 828–832.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232.

Coady, D., Grosh, M. E., and Hoddinott, J. (2004). *Targeting of transfers in developing countries: Review of lessons and experience*, volume 1. World Bank Publications.

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.

Currie, J. and Thomas, D. (1999). Early test scores, socioeconomic status and future outcomes. Working paper 6943, National Bureau of economic research.

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.

Deming, D., Hastings, J., Kane, T., and Staiger, D. (2009). School choice and college attendance: Evidence from randomized lotteries. *NBER Working Paper*.

Drèze, J. and Khera, R. (2010). The BPL census and a possible alternative. *Economic and Political Weekly*, pages 54–63.

Drèze, J. and Sen, A. (2013). *An uncertain glory : India and its contradictions*. London : Allen Lane.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4:3895–3962.

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*, volume 5, pages 653–743. Elsevier.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., and Ravina, R. (2011). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. Working paper 17554, National Bureau of Economic Research.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management*, 1(1):19–41.

Hoxby, C. M. (2003). School choice and school productivity. Could school choice be a tide that lifts all boats? In *The economics of school choice*, pages 287–342. University of Chicago Press.

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):1477–1503.

Härmä, J. (2011). Low cost private schooling in india: Is it pro-poor and equitable? *International journal of educational development*, 31(4):350–356.

Jha, R., Gaiha, R., Pandey, M. K., and Kaicker, N. (2013). Food subsidy, income transfer and the poor: A comparative analysis of the public distribution system in India's states. *Journal of Policy Modeling*, 35(6):887–908.

Juneja, N. (2014). India's New Mandate against Economic Apartheid in Schools. *Journal of International Cooperation in Education*, 16:55–70.

Kaufman, J. E. and Rosenbaum, J. E. (1992). The education and employment of low-income black youth in white suburbs. *Educational Evaluation and Policy Analysis*, 14(3):229–240.

Kingdon, G. G. (2017). The Private Schooling Phenomenon in India: A Review. IZA Discussion Paper No. 10612.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Magnuson, K. (2003). *The effect of increases in welfare mothers' education on their young children's academic and behavioral outcomes: Evidence from the National Evaluation of Welfare-to-Work Strategies Child Outcomes Study*. Institute for Research on Poverty Madison, WI.

Mehta, C. A. (2015). Elementarty Education in India, Analytical Tables 2014-15. Government report, National University of Education Planning and Administration.

Morgan, C., Petrosino, A., and Fronius, T. (2013). *A systematic review of the evidence of the impact of school voucher programmes in developing countries*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

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.

Musset, P. (2012). School choice and equity: current policies in OECD countries and a literature review. Technical report, OECD.

Noronha, C. and Srivastava, P. (2013). India's Right to Education Act: Household experiences and private school responses. *Education Support Program Working Paper Series*, (53).

Patrinos, H. A., Osorio, F. B., and Guáqueta, J. (2009). *The role and impact of public-private partnerships in education*. World Bank Publications.

Pratham Foundation (2014). Annual State of Education Report 2014. Technical report, ASER Center, New Delhi.

Radhakrishna, R. and Sharma, A. N. (2015). Poverty and social protection in urban india. Working Papers, id:7563, esocialsciences.

Rothstein, J. M. (2006). Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350.

Rouse, C. E. and Barrow, L. (2009). School vouchers and student achievement: Recent evidence and remaining questions. *Annu. Rev. Econ.*, 1(1):17–42.

Ryan, C. and Watson, L. (2009). The impact of school choice on students' university entrance rank scores in Australia. *School Choice and School Improvement: Research in State, District and Community Contexts', Vanderbilt University*, pages 25–27.

Saha, D. (2016). Indians Increasingly Prefer Private Education, 71 Million Take Tuitions. *The Wire*. Available at ://thewire.in/education/indians-increasingly-prefer-private-education-71-million-take-tuitions.

Sarin, A., Kuhn, S., Singh, B. D., Khanghta, P., Dongre, A., Joshi, E., Sengupta, A., Das, R., and Rahman, F. (2016). State of the Nation: Rte Section 12 (1) (c) 2015. Technical report, Indian Institute of Management Ahmedabad and others.

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.

Srivastava, P. (2008). *School choice in India: disadvantaged groups and low-fee private schools*. Symposium Books.

Srivastava, P. and Noronha, C. (2014). Institutional framing of the Right to Education Act: Contestation, controversy, and concessions. *Economic and Political Weekly*, 49(18):51–58.

Svedberg, P. (2012). Reforming or replacing the public distribution system with cash transfers? *Economic and Political Weekly*, pages 53–62.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, pages 416–424.

Tilak, J. B. G. (2009). Household expenditure on education and implications for redefining the poverty line in india. Technical report, Planning Commission of India.

White, G., Ruther, M., and Kahn, J. (2016). *Educational Inequality in India: An*

Analysis of Gender Differences in Reading and Mathematics. IHDS Working Paper 2016-2.

Wolfe, A. (2009). *School choice: The moral debate*. Princeton University Press.

Zimmermann, L. (2012). Reconsidering gender bias in intrahousehold allocation in India. *Journal of Development Studies*, 48(1):151–163.

Appendix A: Theory of change

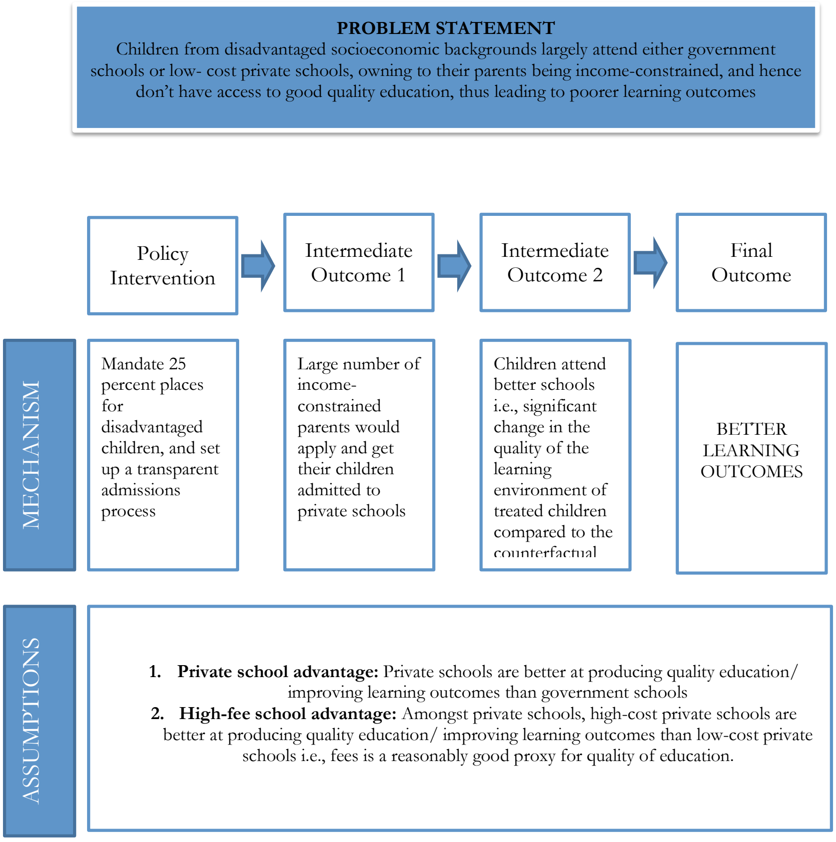


Figure A.1.1: Theory of change underlying the 25 percent mandate

Appendix B: **Sample districts**

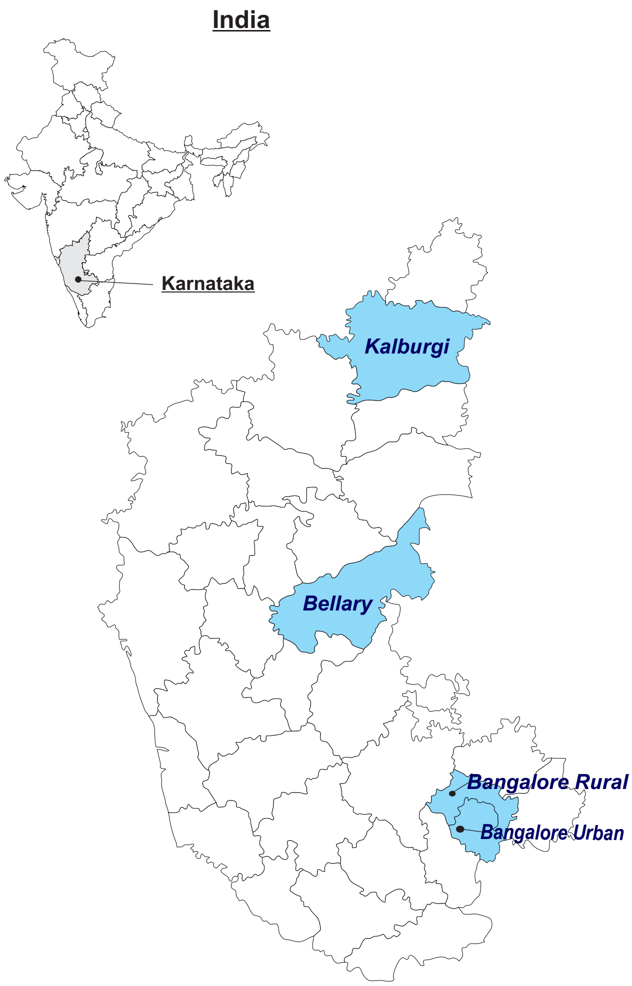


Figure A.1.2: Map of Karnataka with the sample districts

CHAPTER 2

Value added at schools? Experimental evidence on the impacts of private and elite schools

Abstract

I use the context of the RTE 25 percent mandate and the lottery based allocation of places in private schools of Karnataka state, south India to estimate value added at private schools in general and elite private schools in particular. The study sample of 1,616 children were aged 7.3 years, on average, and went through 1.5 years of schooling at end line data collection. The IV estimates show that enrollment at a private school has no statistically significant effect on any of the five test score measures. However, the coefficients are in the range of 0.2-0.6 σ and are not estimated precisely- my sample is only powered to detect very large effects (about 1.3 σ and above). Hence I am not able to comment on the valued added at private schools in learning. Moving beyond test scores, I find large, robust, and statistically significant private school effect of 1.90 standard deviation (σ) on self-efficacy, one of the four measured psychosocial outcomes. On the impact of elite private schools, I find robust effects only on English test scores (0.58 σ). This is unsurprising given the labor market premiums to English proficiency in the Indian context. However,

the lack of test score gains in other subjects is puzzling and points to the importance of non- academic factors in parental school choices.

2.1 Introduction

India's unique school choice policy, the Right To Education (RTE) Act's 25 percent mandate (hereafter the 25 percent mandate), reserves 25 percent of places in all private schools of the country for children from disadvantaged backgrounds. Proponents of the mandate argue that it will lead to improved outcomes for disadvantaged children by providing a ladder of opportunity for such children to move from government schools to private schools and/or from low-fee private schools to high-fee private schools. Inherent in this argument is the assumption that private schools are better than public schools (private school advantage) and that high-fee/elite private schools are better than low-fee private schools (elite school advantage) at producing education outcomes. These assumptions are widely held to be true by policy makers and society at large (Juneja, 2014). However, neither claim is empirically rooted. Existing evidence on the former is mixed, while there is a complete absence of rigorous evidence on the latter.

An ideal method to estimate the valued added at private schools and elite schools, in the context of the 25 percent mandate, is to use the RTE lottery as an instrumental variable for enrollment at private/ elite schools. However, the Intent to Treat (ITT) estimates (in Chapter 1) of the impact of winning the lottery on private school attendance and on school-fee are very modest in magnitude, demonstrating that the RTE lottery did not move lottery-winning children to significantly different schools compared to the lottery losers. Hence if treatment is defined as enrollment at a private school or at an elite school, the compliance rate with the lottery instrument is low and this might lead to large standard errors and imprecise IV estimates.

Acknowledging the challenge of lower compliance rate, I attempt to answer two questions in this chapter. First, are private schools better than government schools at improving childrens' outcomes i.e., is there a private school achievement advantage? Second, are high-fee/ elite private schools better than low-fee private schools at furthering childrens' outcomes i.e., is there a high-fee/elite school achievement advantage? I use the same experimental design, setting and sample from Chapter 1 of the thesis to answer these questions. The study sample comprises of a randomized group of 1,616 children who applied for RTE free places and enrolled in Grade I in mid- 2015. At the time of end line data collection in the last quarter of 2016, the average sample child was 7.3 years old and went through 1.5 years of formal schooling. I rely on the same set of learning and psychosocial variables described in Chapter 1 to measure childrens' outcomes.

To estimate the private school effect, I instrument private school enrollment with the RTE lottery outcome. Despite the small proportion of government school enrolled children in the sample (around 5 percent), the lottery instrument is independent, relevant and exogenous, and hence valid. The IV estimates show that enrollment at a private school has no statistically significant effect on any of the five test score measures. However the coefficients are in the range of 0.2-0.6 σ and are not estimated precisely- the sample is only powered to detect very large effects (about 1.3 σ and above). Hence I am not able to comment on the valued added at private schools in learning. A key finding, however, is on the psyschosocial measures. The private school impact on self-efficacy, one of the four measured psychosocial outcomes, is 1.90 σ. This result is robust to alternate specifications and multiple hypothesis correction and sheds new light on the private school premium question, as previous studies have mostly measured learning outcomes. The one study that did measure psyschosocial outcomes is from rural India and it finds no effects (Singh, 2015).

To estimate the elite school effect, I define elite schools as those charging fee at the 75th percentile (p75) or more of the fee distribution. I find large and statistically significant English effect of 0.58 σ. This effect is robust to alternate definitions of high-fee and to modelling fee as a continuous variable. The English effect probably reflects the significant labor market premiums to proficiency in the language. However, the absence of learning gains in other subjects and overall test scores is still puzzling. It demonstrates that price is not necessarily correlated with quality in the private education markets in urban India. It also underscores the role of non-academic aspects of schooling in parental school choices.

This chapter is primarily related to the empirical literature on school choice in general and to choice in developing countries in particular. Adjudicating on the relative effectiveness of private versus vis-à-vis public schools has been a central focus of this literature. Ashley et al. (2014) review the results of 21 studies from India, Pakistan, Kenya, Ghana, Nigeria and Nepal. They conclude that the evidence in favor of private schools producing better learning outcomes than government ones is moderately strong, though the size of the private school effect is ambiguous. However, barring a few studies (Muralidharan and Sundararaman, 2015; Singh, 2015), most of the literature in this review is chiefly based on cross sectional estimators that are confounded by selection and omitted variables, and hence cannot be treated as causal. Glewwe and Muralidharan (2016) summarize the few rigorous studies from India, Chile and Columbia, and opine that the existing evidence points to private schools being more productive than government schools (producing the same results at much lower costs), but not more effective in raising test scores. They conclude that further evidence is required to understand if the higher productivity of private schools can be translated into better learning outcomes.

Despite not being able to comment on the value-added in learning at private schools, owing to the small government school sample, I contribute to this literature by focusing on a rich set of attitudinal/ psychosocial outcomes in addition to test scores and thus provide a more comprehensive understanding of the impacts of private schools. Singh (2015) points to the economics and child psychology literature and argues that understanding the impact of private schools on psychosocial or non-cognitive skills is equally important, in addition to test scores, as these skills are both desirable in their own right and could play an instrumental role in learning. He measures psychosocial skills on a rural sample, but only along two dimensions: locus of control and self-efficacy. My self-efficacy results, the first evidence from an urban Indian sample, highlight the need for going beyond test scores to arrive at a more comprehensive understanding of the private school effects.

Further, my investigation of the elite school effects is a singular contribution to the literature. Despite the centrality of the question in the context of developing countries (with large and heterogeneous private schooling markets) there is no empirical evidence on the issue. My study setup where the RTE free places of different value are randomly assigned enables an investigation of this question. The finding of absence of value added in overall test scores at elite private schools complements the results from the small global literature on value added at elite public schools. The bulk of the evidence here points to the lack of value added at these schools and suggests that perceptions of school superiority determine parental school choices rather than schools' impact on cognitive development (Abdulkadiroğlu et al., 2014; Zhang, 2012; Lucas and Mbiti, 2014).

This chapter is also related to the empirical literature on price-quality relationships. Following the economic theory of competitive markets that posits a strong correlation between price and quality, empirical research has tested the validity of this relationship in the context of consumer goods. This research finds the correlations to be weak, thus challenging the conventional heuristic of 'you get what you pay for' (Steenkamp, 1988; Judd, 2000; Bartelink, 2016). Beyond consumer goods the evidence is scant; one recent study on health services in German nursing homes finds a positive relationship (Herr and Hottenrott, 2016). My results on the elite school premium and the relationship between school fee and academic quality is the first investigation into 'price-quality correlation' question in education markets and, hence, is an important contribution.

The rest of the chapter is structured as follows: Section 2.2 explains the context, study setting and design, and presents the data and estimation strategy. Section 2.3 presents the results. Section 2.4 is a detailed discussion, and section 2.5 concludes.

2.2 Setting and data

2.2.1 Contextual details

As discussed in Chapter 1, two distinct features of India's private sector are it size and heterogeneity. 49 percent of primary school aged children (6-10 years) and 42 percent of all children (6-18 years) are in the private sector in urban India. The figures for rural India are 21 percent in both categories. Through private schools vary on a range of variables like size, infrastructure, number of teachers and so on, the school fee variable captures this heterogeneity most comprehensively. The private school fee distribution is log-normal with a pronounced rightward skew; the highest annual fee is around Indian Rupees (INR) 200,000 (USD 3,100), 40 times the median fee of INR 5,000 (USD 77). Further, the mean private school fee is only about 9.2 percent of the per capita GDP, and about 26 percent of rural private schools have a monthly fee that is below the daily minimum wage (Kingdon, 2017). Hence the private schooling market serves all socioeconomic groups: the rich, the middle-class, and even a section of the poor.

Given this and the general belief in the 'private school advantage' and the 'elite school advantage,' choice enthusiasts contend that policies like the RTE 25 percent mandate will improve education outcomes by shifting children from low-quality learning environments (government and low-fee private schools) to high-quality learning environments (high-fee/ elite private schools). Low-fee private schools, in this context, are defined as those with fee lower than the daily minimum wage (Srivastava, 2008). As there is no agreed upon definition of high-fee schools, I define schools with fee higher than the 75th percentile (p75) of the fee distribution as high-fee/ elite.

2.2.2 Study setting and design

The setting for this study is same as that for the policy evaluation in Chapter 1. I use the same sample of 1,616 children (808 matched pairs) who applied for the RTE free places from four districts of Karnataka state in South India. The lottery based allocation of free places provides an opportunity for estimating the causal effects of private and elite school enrollment.

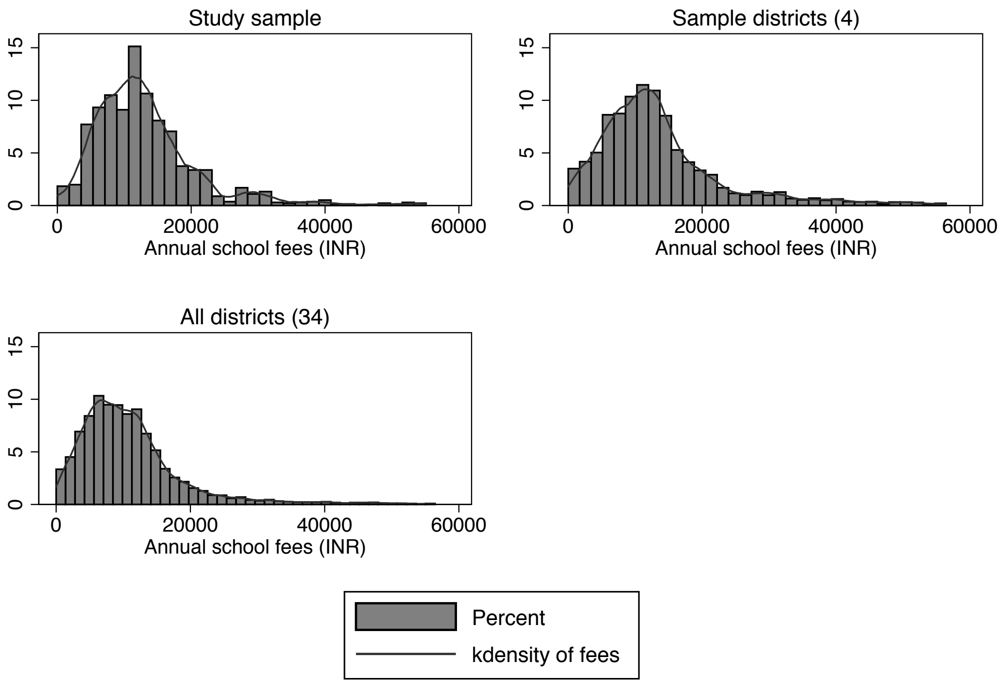
2.2.3 Data

I use three datasets in this analysis all of which have been used in Chapter 1. Data on childrens' outcomes and household characteristics comes from the primary household survey. School tuition fee data, a key variable in estimation of the high-fee school effect, comes from an authenticated school fee dataset of the Government of Karnataka. The dataset has fee data of 10,442 private schools for the whole state (34 districts). Of this 2,546 schools are from the four sample districts. Table 2.1 presents the descriptive statistics of the fee data for the schools in the sample, the four sample districts and the whole state of Karnataka. Figure 2.1 shows the histogram of school fee and the kernel density plots. All the distributions are log normal with a heavy rightward skew, similar to the national pattern reported by Kingdon (2017).

Table 2.1: Summary of school fee data

District	Number of schools	Points on the fee distribution (in Indian Rupees(INR)) Average	Maximum	Median	p10	p90	p99
Sample	587	13,902	586,228	11,625	4,086	22,785	61,389
Sample districts (4)	2,546	19,022	897,626	12,059	4,373	32,761	142,455
All districts (34)	10,442	13,616	1,365,897	9,645	3,396	21,358	82,989

Notes: p10, p90,and p99 are tenth, ninetieth,and ninety ninth percentiles values of the fee distribution.



Notes:School fees data from 0-98th percentile shown here

Figure 2.1: Distribution of annual school fee

2.2.4 Estimation: Private school effect

I start by estimating equation below using Ordinary Least Squares (OLS).

$$Y_i = \beta_0 + \beta_1.Private_i + \boldsymbol{\beta}_{Di}.\mathbf{D}_i + \epsilon_i \tag{2.1}$$

Where Y_i is the outcome variable for child i. $Private_i$ is a binary treatment variable taking a value of 1 if child i is enrolled at a private school. I use sub-district dummies ($\mathbf{D}_i$) to absorb geographic variation and to improve precision.[1] The regressor of interest $Private_i$ is endogenous in this model making β_1 biased. To correct for the bias and to recover a causal estimate of β_1, I perform Two Stage Least Squares (2SLS) estimation by instrumenting private school enrollment with the RTE lottery.[2] The first stage equation of the 2SLS/IV model is as follows.

$$Private_i = \alpha_0 + \alpha_1.Lottery_i + \boldsymbol{\alpha}_{Di}.\mathbf{D}_i + \mu_i \tag{2.2}$$

Where $Lottery_i$ is the RTE lottery outcome for child i taking a value of 1 for lottery winners and zero for lottery losing children. Several assumptions must be satisfied for $Lottery_i$ to be a valid instrument (Imbens and Rubin, 2015; Wooldridge, 2013).

[1] The sample is drawn from 4 districts and 28 sub-districts, referred to as 'blocks' in India

[2] I run the OLS model both with and without household and child related controls: age of the child, gender of the child, indicator for belonging to Scheduled Caste, indicator for being Muslim, mother's education (in years), indicator for a working mother and two measures of the economic status of the household, a household asset index measure and the number of rooms in the house. But my preferred specification for the 2SLS model is the one without controls.

- *Independence* (the instrument should be uncorrelated to the error term): This assumption is valid, by definition, for the lottery instrument.
- *Instrument strength* (the instrument should be a good predictor of the endogenous regressor i.e., α_1 should be statistically significant with an F-stat of at least 10): Table 2.2 (panel A) presents data on the public and private enrollment in the sample. Of the 1,616 children, only 72 (4.5 percent) are enrolled at government schools while the rest are enrolled at private schools. Panel B shows the first stage regressions results of the lottery effect on private school enrollment.

Table 2.2: Private school effects: strength of the lottery instrument

	(1)	(2)	(3)	(4)
Panel A- Public versus private enrollment				
	Public	Private	Total	Percent Private
Treatment	13	795	808	98.4
Control	59	749	808	92.7
Total	72	1544	1616	95.5
Panel B- First stage results				
	Private school enrollment indicator			
Treatment	0.06***	0.06***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.93***	0.89***	0.93***	0.96***
	(0.01)	(0.05)	(0.07)	(0.09)
Observations	1616	1616	1616	1616
R-squared	0.02	0.05	0.08	0.05
Subdistrict dummies	No	Yes	Yes	No
Controls	No	No	Yes	Yes
Pair fixed effects	No	No	No	Yes
F-statistic	32.30	31.82	31.69	32.54

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors clustered at the pair level.

Across specifications the F-statistic is above 30 demonstrating the strength of the instrument, despite the low compliance rate.[3]

- *Exclusion restriction* (the effect of instrument on outcomes should operate only through the endogenous regressor): It is unlikely that winning or losing the lottery would have a direct effect on childrens' test scores; however, it could be argued that winning the lottery might affect attitudinal outcomes, especially self-efficacy. I believe even this is improbable for two reasons. First, I measured childrens' outcomes about 18 months after the lottery results. The time gap rules out the possibility of lottery affecting outcomes. Second, it is the parents that are engaged in the admission system, and it is unlikey that many children are even aware of the RTE lotteries.

- *Monotonicity* (there should be no defiers): A defier, in this case, would be someone that enrolls at a private school if she did not win the lottery and at a government school otherwise. This behavior is implausible and it is safe to ignore the possibility of defiers.

The RTE lottery ($Lottery_i$) is, therefore, a valid instrument for private school enrollment.[4] Another key aspect of the IV strategy is that the IV estimator

[3] While estimating the impact of the policy in Chapter 1, I defined treatment as offer of a free place. Compliance meant being enrolled at a free place conditional on winning the lottery. Here I redefine treatment as offer of place at a private school. Hence compliance would mean being enrolled at a private school conditional on winning the lottery. The compliance rate is the difference between the private school enrollment rate of the treatment and the control children i.e., 98.4- 92.7= 5.7 percent. The IV estimation relies on this small compliance rate.

[4] In addition to the four fundamental assumptions, the lottery instrument also satisfies the Stable Unit Treatment Value Assumption (SUTVA), according to which the treatment status of one unit should not affect the outcome of other units. Following choice theory, it can be argued that a large-scale policy like the 25 percent mandate could trigger competition amongst schools and lead to overall improvements in productivity. That would be a violation of SUTVA. However, as argued in Chapter 1, general equilibrium effects are improbable in this case. First, at the time of my survey the policy is only 3 years old, hardly enough time for schools to change their education production systems (including hiring better teachers, mending pedagogical practices and so on) and to improve outcomes. Second improving effectiveness to attract RTE children is not incentive compatible either for public or private schools.

is the Local Average Treatment Effect (LATE), the average treatment effect for the compliers. The compliers comprise the sub-group that have been induced by randomization to receive the treatment, in this case the sub-group of policy applicants that has been induced to shift to private schools because of the lottery/policy i.e., the population of default government school goers. This is precisely the sub-population that the RTE mandate sought to target. Hence the LATE estimates are very relevant from a policy perspective.

2.2.5 Estimation: Elite/high-fee school effect

As presented in Table 2.2, 95 percent of the study sample is enrolled at private schools. The wide variation in private school fee and its log normal distribution described by Kingdon (2017) for the whole country can also be seen in my sample. I already described the variation in fee for the four sample districts and for the whole state of Karnataka at Table 2.1. Figure 2.1 shows the histogram of school fee and the kernel density plots. Figure 2.2, where the fee density plots of the sample schools, schools in sample districts and those in the whole state of Karnataka are superimposed, identifies the common support of school fee between my sample and the broader population of schools.[5]

[5] The distributions of the sample districts and the study sample are slightly to the right of the distribution of all the private schools in the state. This is because two of three Bangalore districts, which have a greater proportion of high-fee schools, are part of the sample.

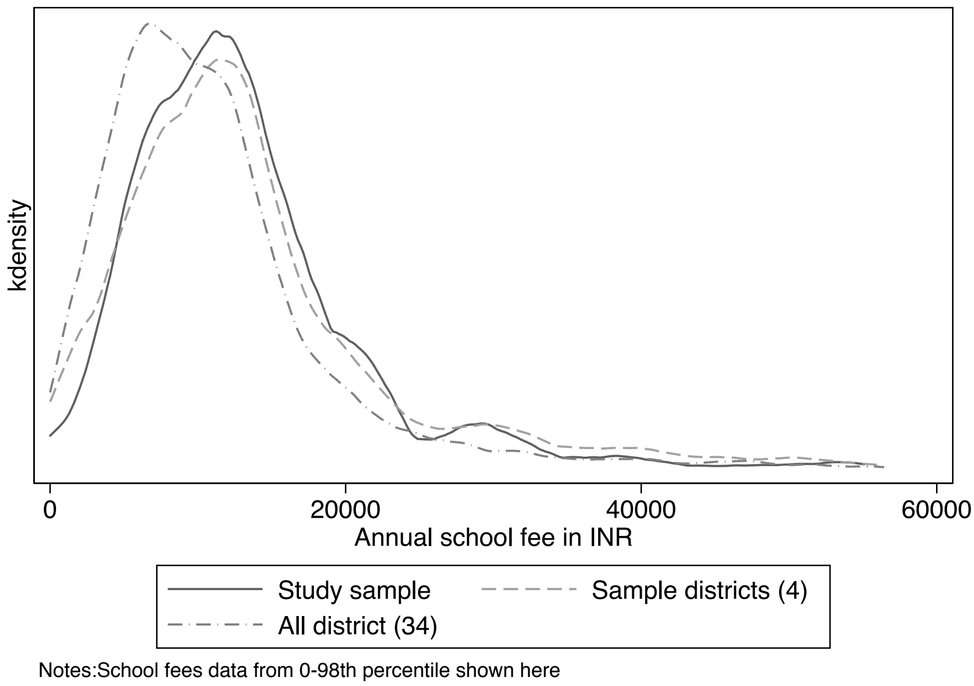


Figure 2.2: Distributional plots of school fee

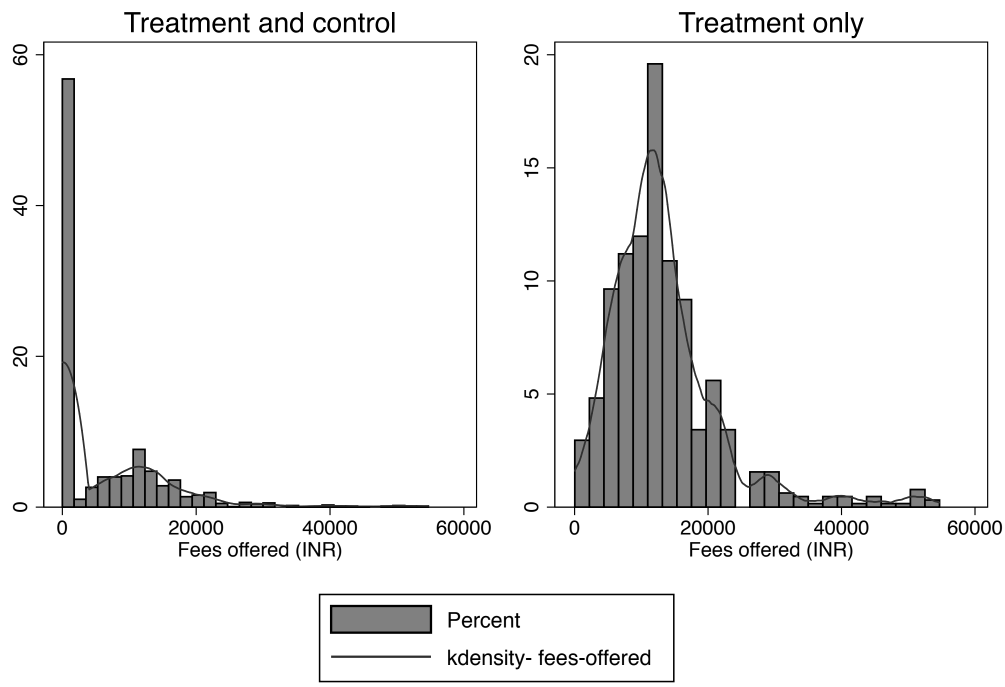
I exploit the variation of school fee in the sample to estimate the elite school effect using the following model.

$$Y_i = \beta_0 + \beta_2.Elite_i + \boldsymbol{\beta}_{Di}.\mathbf{D}_i + \epsilon_i \tag{2.3}$$

Where the variables Y_i and D_i mean the same as in equation (2.1). The endogenous regressor $Elite_i$ is a dummy variable taking on the value 1 if the child is enrolled in a elite school: school with fee higher than the 75$^{\text{th}}$ percentile value of the fee distribution.[6] However, like in the estimation of the private school effect, β_2 would be biased if estimated by OLS given the endogenity of the enrollment at an elite school. I again perform 2SLS estimation and instrument elite school enrollment with two variables. The

[6] Given that my definition of elite schools is somewhat arbitrary, I demonstrate the robustness of elite school effect results to alternate definitions of elite schools, starting from those charging fee greater than the 50$^{\text{th}}$ percentile (p50) till the 90$^{\text{th}}$ percentile (p90).

first instrument is the RTE lottery and the second instrument is the fee of the school that has been offered to the treated child in the lottery (referred to as the fee-offered variable). The fee-offered instrument takes advantage of the varying treatment level for the treated children i.e., each treated child is offered an RTE place at a school with a different fee. Fee-offered for all control children is zero.Figure 2.3 shows the distribution of the fee-offered variable for the full sample (both treatment and control) and for the treatment children only.



Notes: Observations from 0-98th percentile of the fees-offered variable shown here

Figure 2.3: Distribution of fee-offered instrument

Unlike the lottery instrument, however, the fee-offered variable is not exogenous; it is correlated with the schools the parents choose in the admission application (each parent can choose up to five schools in the application). Hence it does not satisfy the independence assumption. To correct this, I include the fee of the first-choice school and that of the highest-fee school in each students' choice profile as covariates. That makes the fee-offered instrument conditionally exogenous.[7] The first stage equation of the 2SLS model, hence, is as follows:

$$\begin{aligned} Elite_i = \alpha_0 + \alpha_2.Lottery_i + \alpha_3.Fee\_offered_i + \alpha_4.Fee\_first_i + \\ \alpha_5.Fee\_highest_i + \boldsymbol{\alpha}_{Di}.\mathbf{D}_i + \mu_i \end{aligned} \tag{2.4}$$

The two instruments, $Lottery_i$ and $Fee\_offered_i$, also satisfy the other assumptions for being valid. The F-stat of the first stage with the two instruments is higher than 10 and the p-value of the Hansen J-statistic (over identification test) is above 0.1. The arguments for the validity of the exclusion restriction and monotonicity are similar to those made at section 2.2.4.

Making the $Fee\_offered_i$ instrument conditionally exogenous by controlling for fee of the first-choice school (Fee_first_i) and that of the highest fee school ($Fee\_highest_i$) in each students' choice profile comes with a cost: missing data. As explained in Chapter 1, fee data is missing for about 15 percent of the schools in the State. In the study sample of 1,616 children, 157 are missing the fee data; introducing three more fee-related variables into the model ($Fee\_offered_i$, Fee_first_i, $Fee\_highest_i$) increases the number of missing fee observations to 382. So I only have 1,232 observations to estimate the elite school effect. Missing observations lead to concerns over differential attrition. Table 2.3 presents regression results confirming that the missingness is orthogonal to treatment. Hence the estimates of the elite school effects

[7] Running a pair fixed effects model would also make the instrument exogenous. But I will have to drop a significant number of observations to estimate pair fixed effects model.

using the sample of 1,232 children would be unbiased.

Table 2.3: Is attrition of fee data random?

	(1)	(2)	(3)
VARIABLES		Missing Fee	
Treatment	0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)
Constant	0.76***	0.97***	0.98***
	(0.02)	(0.03)	(0.11)
Observations	1616	1616	1616
R-squared	0.00	0.20	0.20
Subdistrict dummies	No	Yes	Yes
Controls	No	No	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable in all the regressions is a dummy for missing fee (382 observations). Treatment is not significant in any specification.

I, therefore, estimate the elite school effects using equation (2.5) with the endogenous elite school enrollment variable instrumented by lottery and the fee-offered variables.

$$\begin{aligned} Y_i = \beta_0 + \beta_2.Elite_i + \beta_3.Fee\_first_i + \\ \beta_4.Fee\_highest_i + \boldsymbol{\beta}_{Di}.\mathbf{D}_i + \epsilon_i \end{aligned} \tag{2.5}$$

2.3 Results

2.3.1 Private school effects

The private school effects are presented in Table 2.4. Panel A reports the results of the OLS estimation of equation (2.1): the point estimates (β_1), standard errors in parenthesis, and sharpened q-values in brackets.[8] Private

[8] I correct for multiple hypothesis by computing sharpened q-values, which are adjusted p-values controlling for False Discovery Rate (FDR) (Anderson, 2008; Benjamini et al., 2006).

school enrollment is associated with improved total test score of 0.25 σ, and the effect seems to be driven primarily by gains in English, where the effect size is 0.80 σ. On psychosocial outcomes, the effects are statistically significant on self-efficacy and school experience indices. All the effects are robust to corrections for multiple hypothesis as demonstrated by the significant q-values. These OLS estimates are, however, biased as discussed in the previous section.

The unbiased 2SLS/ IV estimates of the private school impact (with private school enrollment instrumented by the RTE lottery outcome) are presented in Panel B. None of the test score measures is significant in this specification, though the point estimates are in the range of 0.2-0.6 σ. Given the very small sub-sample of government school goers, I am only powered to detect effects of above 1.3 σ. On psychosocial outcomes, only the self-efficacy measure is significant with an effect size of 1.90 σ. This effect is robust to correction for multiple hypothesis and to inclusion of household and child related controls. Finally, I present the Intent To Treat (ITT) / reduced-form estimates (from Chapter 1) of regressing the lottery outcome on childrens' outcomes in Panel C. A comparison of the estimates at Panels B and C illustrates that the IV estimates are scaled up values of the ITT estimates. This is because when the endogenous regressor and the instrument are binary, the IV estimation yields a Wald estimator which equals the ITT estimate scaled by the compliance rate.[9]

[9] The compliance rate is 5.7 percent, so the IV estimates are approximately equal to the ITT values divided by 0.057.

Table 2.4: Private school effects: OLS and IV results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total score					Psychosocial outcomes			
	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
Panel A- OLS results									
Private enrolled	0.25**	0.17	0.01	0.80***	-0.11	0.25***	0.08	0.41***	0.01
	(0.11)	(0.11)	(0.11)	(0.10)	(0.11)	(0.09)	(0.10)	(0.15)	(0.10)
	[0.04]**	[0.16]	[0.60]	[0.00]***	[0.29]	[0.02]**	[0.42]	[0.02]**	[0.81]
Observations	1,616	1,614	1,614	1,614	1,614	1,553	1,577	1,585	1,557
R-squared	0.20	0.19	0.16	0.20	0.25	0.07	0.12	0.16	0.18
Panel B- 2SLS/IV results									
Private enrolled	0.66	0.23	0.72	0.52	0.58	1.90**	0.50	0.56	-0.47
	(0.76)	(0.77)	(0.78)	(0.76)	(0.74)	(0.83)	(0.76)	(0.81)	(0.76)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[0.09]*	[0.67]	[0.09]	[0.09]
Observations	1,616	1,614	1,614	1,614	1,614	1,553	1,577	1,585	1,557
Panel C- ITT/ reduced form results									
Treatment	0.04	0.01	0.04	0.03	0.03	0.12**	0.03	0.03	-0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[0.07]*	[0.67]	[0.09]	[0.09]
Observations	1,616	1,614	1,614	1,614	1,614	1,553	1,577	1,585	1,557
R-squared	0.10	0.11	0.10	0.08	0.15	0.05	0.12	0.15	0.17

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Clustered standard errors (at pair level) are in parenthesis and sharpened q-values in brackets. Panel A is the output of the uninstrumented (OLS) regression with student outcomes regressed on the endogenous private school enrollment variable. This model controls for sub-district (dummies) and for the household and child related covariates. Panel B are the 2SLS estimates with private school enrollment being instrumented by the RTE lottery outcome. The model only controls for sub-district dummies. Results are robust to the inclusion of covariates. Panel C shows the ITT results of regressing student outcomes on lottery indicator and sub-district dummies. The IV estimates are Wald estimators and are equal to the ITT estimates divided by the compliance rate (0.057).

To understand the source of the large self-efficacy effect, I estimate the private school effect on each of the individual statements used in the construction of the psychosocial indices. Table 2.5 presents the results for my preferred specification with sub-district dummies and without controls. The effects are not statistically significant for fifteen out of the sixteen statements. In particular, the effect is not significant on any of the four statements used to construct the self-efficacy index. This suggests that the almost 2.0 σ effect on self-efficacy is not driven by any single statement.

Table 2.5: Private school impact on psychosocial statements

	(1) Coefficient	(2) p-value
Self-efficacy		
Learning English is easy for me	0.10	0.70
Learning Mathematics (additions and subtractions) is difficult for me	0.55	0.11
I am proud of my achievement at school	0.19	0.23
I am good at learning Kannada	0.52	0.10
Peer support		
My classmates/friends tease me	0.27	0.41
Most of my classmates do not want to play with me in leisure hours	-0.07	0.82
I have a lot of friends in my class	-0.04	0.87
I can ask a friend to help me if I have difficulty with class work	0.25	0.38
School experience		
I am bored in class	-0.46	0.16
I feel alone in school	-0.46	0.16
I feel safe at school	0.15	0.22
I am proud to study in this school	0.28	0.13
Teacher support		
My class teacher does not treat me fairly	-0.03	0.89
I can talk to my class teacher freely	-0.09	0.71
I am afraid to ask my teacher to clarify my doubts	-0.78**	0.03
Most teachers at school are concerned about me	0.22	0.19

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Each row in the table presents the results of a separate 2SLS regression: response to a statement on the psychosocial test regressed on private school dummy (instrumented with the lottery) and sub-district dummies.

2.3.2 Elite school effects

Table 2.6 presents the elite school effects. Panel A shows the results of the OLS regression of equation (2.3). The coefficients on GCA, English, Kannada (local language) and school experience are significant. The GCA, English and Kannada effects are also robust to multiple hypothesis correction. That the Kannada effect is negative aligns with the general wisdom that elite schools focus excessively on English to the detriment of the local language learning. The effect sizes are modest with English effect being the highest at 0.2 σ. These results, however, would be biased and are presented only for the sake of comparison with the unbiased IV estimates.

Panel B presents the 2SLS/ IV estimates of the impact of enrollment at an elite school on outcomes. The F-stat of the first stage regression with both the instruments is 13.98. The coefficients on English score and school experience are statistically significant, though only the English effect is robust to multiple hypothesis correction. The effect size is a reasonably high 0.58 σ . This result is robust to the inclusion of household and child covariates. Finally, the result is also robust to using only the sample of private school enrolled children, obtained by dropping the 72 government school going children (Panel C).

Table 2.6: Elite school effects: OLS and IV results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total score					Psychosocial outcomes			
	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
Panel A- OLS results									
High-fee enrolled	0.08	0.16**	0.08	0.20***	-0.11*	0.04	0.05	0.11**	0.08
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.05)	(0.06)
	[0.12]	[0.04]**	[0.12]	[0.00]***	[0.01]***	[0.51]	[0.51]	[0.18]	[0.34]
Observations	1,458	1,456	1,456	1,456	1,456	1,401	1,425	1,429	1,404
R-squared	0.20	0.19	0.16	0.17	0.26	0.07	0.12	0.15	0.17
Panel B- 2SLS/IV results									
High-fee enrolled	0.16	0.29	0.27	0.58***	0.02	0.33	-0.07	0.20**	0.05
	(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.10)	(0.19)
	[0.38]	[0.25]	[0.25]	[0.01]**	[0.58]	[0.20]	[0.63]	[0.20]	[0.63]
Observations	1,232	1,230	1,230	1,230	1,230	1,189	1,201	1,208	1,188
Panel C- 2SLS/IV results (private school sample only)									
High-fee enrolled	0.16	0.29	0.25	0.60***	0.01	0.32	-0.08	0.20**	0.03
	(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.09)	(0.19)
	[0.39]	[0.25]	[0.25]	[0.01]**	[0.60]	[0.24]	[0.79]	[0.16]	[0.79]
Observations	1,172	1,170	1,170	1,170	1,170	1,131	1,141	1,148	1,129

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Clustered standard errors (at pair level) are in parenthesis and sharpened q-values in brackets. Panel A is the output of the uninstrumented (OLS) regression with elite school enrollment as the primary regressor and student outcomes as dependent variables. This model controls for sub-district (dummies) and for the household and child related covariates. Panel B are the 2SLS/IV estimates with elite school enrollment being instrumented by lottery outcome and fee-offered variable. The model only controls for sub-district dummies. The first F-stat is 13. 98 and p-value of the Hansen J statistic (overidentification test of all instruments) is 0.95. Results are robust to the inclusion of covariates. Panel C shows the 2SLS/IV estimates for only the private school attending sample (obtained by dropping the 72 government school going children).

2.3.3 Robustness checks

Given that my definition of elite private school is somewhat arbitrary, I demonstrate the robustness of the elite school effects for different definitions of elite school, starting from schools charging fee greater than the 50th percentile (p50) upto the 90th percentile (p90). The results for 8 alternate definitions of elite school are shown in Table 2.7. Each row of the table presents results for one definition of elite school. The English effect is statistically significant (and robust to multiple hypothesis correction) for all definitions of elite school. The effect size ranges from from 0.58- 0.76 σ.The results also hold for the sample of private school attending children. (Table 2.8).

Table 2.7: Elite school effects: Robustness-I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite	First stage	Test scores					Psychosocial outcomes			
indicator	F-stat	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
>p50	11.24	0.22	0.34	0.37*	0.71***	0.05	0.45	-0.06	0.27**	0.05
		(0.33)	(0.26)	(0.22)	(0.26)	(0.24)	(0.28)	(0.26)	(0.13)	(0.23)
>p55	11.56	0.20	0.32	0.32*	0.65***	0.04	0.40	-0.06	0.24**	0.05
		(0.30)	(0.23)	(0.19)	(0.23)	(0.22)	(0.25)	(0.24)	(0.12)	(0.21)
>p60	11.34	0.19	0.31	0.31	0.64***	0.03	0.38	-0.06	0.23**	0.06
		(0.30)	(0.23)	(0.19)	(0.22)	(0.21)	(0.24)	(0.23)	(0.11)	(0.20)
>p65	11.21	0.17	0.31	0.29	0.61***	0.02	0.35	-0.07	0.21**	0.06
		(0.28)	(0.22)	(0.18)	(0.21)	(0.20)	(0.23)	(0.22)	(0.10)	(0.19)
>p70	13.07	0.16	0.30	0.26	0.58***	0.01	0.32	-0.07	0.20**	0.06
		(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.20)	(0.10)	(0.18)
>p75	13.98	0.16	0.29	0.27	0.58***	0.02	0.33	-0.07	0.20**	0.05
		(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.10)	(0.19)
>p80	16.02	0.16	0.30	0.27	0.59***	0.02	0.33	-0.07	0.20**	0.05
		(0.27)	(0.21)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.10)	(0.19)
>p85	18.19	0.18	0.32	0.31*	0.65***	0.03	0.38*	-0.07	0.23**	0.06
		(0.30)	(0.22)	(0.19)	(0.21)	(0.21)	(0.23)	(0.23)	(0.11)	(0.21)
>p90	27.57	0.18	0.40	0.32	0.76***	0.00	0.39	-0.12	0.24*	0.08
		(0.35)	(0.26)	(0.22)	(0.23)	(0.25)	(0.27)	(0.28)	(0.13)	(0.24)

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Each cell reports the result from an IV regression of one of the outcome variables on an indicator for enrollment at an elite school. Elite school is defined by cutting the school fee distribution at various points, from the 50th percentile (p50) to the 90th percentile (p90). Hence the definition of elite school changes in each row: in the first row the top 50 percent schools in the fee distribution are defined as elite. The F-stat of the first stage regressions is reported in column 1. The F-stat is higher than 10 for all definitions of elite school. The Hansen J- stat p-value (over identification test) is always above 0.1. Hence the instruments are valid. Standard errors are clustered at the pair level in all models.

Table 2.8: Elite school effects: Robustness-II (private school sample only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite	First stage			Test scores				Psychosocial outcomes		
indicator	F-stat	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
>p50	9.42	0.20	0.36	0.33	0.74***	0.03	0.41	-0.10	0.26**	0.03
		(0.35)	(0.26)	(0.22)	(0.26)	(0.24)	(0.28)	(0.26)	(0.12)	(0.23)
>p55	10.53	0.18	0.33	0.29	0.68***	0.02	0.36	-0.09	0.23**	0.03
		(0.31)	(0.24)	(0.20)	(0.23)	(0.22)	(0.25)	(0.23)	(0.11)	(0.21)
>p60	10.72	0.17	0.32	0.28	0.66***	0.02	0.35	-0.09	0.22**	0.03
		(0.30)	(0.23)	(0.19)	(0.22)	(0.21)	(0.24)	(0.22)	(0.10)	(0.20)
>p65	11.03	0.17	0.31	0.26	0.63***	0.01	0.33	-0.09	0.21**	0.03
		(0.29)	(0.22)	(0.18)	(0.21)	(0.20)	(0.23)	(0.21)	(0.10)	(0.19)
>p70	13.36	0.15	0.30	0.24	0.61***	0.00	0.30	-0.08	0.19**	0.03
		(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.20)	(0.09)	(0.18)
>p75	13.85	0.16	0.29	0.25	0.60***	0.01	0.32	-0.08	0.20**	0.03
		(0.27)	(0.20)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.09)	(0.19)
>p80	15.88	0.16	0.29	0.26	0.60***	0.02	0.32	-0.08	0.20**	0.03
		(0.27)	(0.21)	(0.17)	(0.19)	(0.19)	(0.21)	(0.21)	(0.10)	(0.19)
>p85	17.94	0.18	0.31	0.30	0.65***	0.03	0.37	-0.08	0.24**	0.03
		(0.30)	(0.23)	(0.19)	(0.21)	(0.21)	(0.23)	(0.24)	(0.11)	(0.21)
>p90	27.40	0.20	0.40	0.31	0.81***	-0.00	0.39	-0.12	0.23*	0.04
		(0.35)	(0.26)	(0.22)	(0.23)	(0.25)	(0.27)	(0.27)	(0.13)	(0.24)

Notes: *** p<0.01, ** p<0.05, * p<0.1. This table is similar to Table 2.7 except that only the sample of private school attending children is used. Each cell reports the result from an IV regression of one of the outcome variables on an indicator for enrollment at an elite school. Elite school is defined by cutting the school fee distribution at various points, from the 50^{th} percentile (p50) to the 90^{th} percentile (p90). Hence the definition of elite school changes in each row: in the first row the top 50 percent schools in the fee distribution are defined as elite. The F-stat of the first stage regressions is reported in column 1. The F-stat is higher than 10 for most definitions of elite school. The Hansen J- stat p-value (over identification test) is always above 0.1. Hence the instruments are valid. Standard errors are clustered at the pair level in all models.

I conduct a second robustness check on the elite school effects by using fee as continuous variable. This strategy allows the estimation of the impact of an additional Indian rupee of fee on outcomes. The estimation equation is below:

$$\begin{aligned} Y_i = \delta_0 + \delta_1.Fee\_enrolled_i + \delta_2.Fee\_first_i + \\ \delta_3.Fee\_highest_i + \boldsymbol{\delta}_{Di}.\mathbf{D}_i + \epsilon_i \end{aligned} \tag{2.6}$$

$Fee\_enrolled_i$ is the fee of the school the child is currently enrolled at, and as in equation (2.5),it is instrumented with $Fee\_offered_i$ and $Lottery_i$ variables. Table 2.9 presents the results the 2SLS regressions. Only the English and the school experience effects are statistically significant like in the discrete case (Tables 2.6 and 2.7). Further only the English effect is robust to multiple hypothesis correction. The dependent variable, fee, is expressed in 1,000 INR units. Hence a 0.02 English effect means that increasing the school fee by 10,000 INR increases English scores by 0.2 σ. This implies that moving from zero fee school to a school at the 40$^{\text{th}}$ percentile (INR 9,814) of the distribution or moving from 60$^{\text{th}}$ percentile school (INR 12,876) to the 90$^{\text{th}}$ percentile school (INR 22,785) improves English scores by 0.2 σ and does not significantly change other outcomes.

Comparing the English effect sizes of Table 2.6 (fee as a discrete variable) and Table 2.9 (fee as a continuous variable) suggests that the impact of fee on English test scores is non-linear and that the results are driven by the most expensive schools at the top of the distribution.

Table 2.9: Impact of school fee on student outcomes (IV estimates)

	Test scores					Psychosocial outcomes			
	Total	GCA	Math	English	Kannada	Self efficacy	Peer support	School experience	Teacher support
Fee	0.00	0.01	0.01	0.02***	0.00	0.01	-0.00	0.01*	0.00
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
	[0.40]	[0.19]	[0.19]	[0.02]**	[0.62]	[0.30]	[0.61]	[0.30]	[0.61]
Obs.	1,232	1,230	1,230	1,230	1,230	1,189	1,201	1,208	1,188

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Table shows the results estimating equation (2.6) by instrumenting the school fee with the fee-offered and treatment variables. The first stage F-stat is 20.65. The p-value of the Hansen J statistic is 0.82. As the dependent variable fee is expressed in INR 1,000, the coefficient of 0.02 on English implies that an INR 10,000 increase in fee leads to a 0.2 σ increase in English scores.

2.4 Discussion

2.4.1 Private school advantage

My results do not shed any new light on the test score impacts of private schools, given the very small government school attending children in the sample. However, my psychosocial results point to a new dimension of the private school advantage question. The large self-efficacy impact (1.90 σ, robust to multiple hypothesis testing) indicates that private schools improve the sense of self-efficacy of children. This has important implications. Psychology literature has established that self-efficacy influences several aspects of behavior like effort and persistence that are essential for learning (Bandura, 1977, 1982; Jinks and Morgan, 1999; Zhang et al., 2015; Zimmerman, 2000). Jinks and Morgan (1999) provide an excellent overview of the mechanisms through which efficacy improves learning. Hence, a near 2.0 σ improvement in self-efficacy, that I report, may have the potential to increase achievement motivation, and consequently, future academic outcomes.

Given the IV estimation, the population to which the results are valid has to be clearly understood. As explained in section 2.2.4, the private school impacts presented here are not the effects for the average child in the population; they are effects for the compliers, the sub-population of default government school goers. Interestingly however, the compliers are the sub-group targeted by mandate, thus making the results very pertinent from a policy perspective.

Finally, it is crucial to think about the source of the near 2.0 σ self-efficacy effect. Could it be due to private schools consciously practicing 'feel good' teaching, or could it be more subtle, arising from the general societal narrative around the desirably of private schooling relative to public schooling (Juneja, 2014)? There is very little research to inform this issue, but anecdotal evidence suggests that it is more the latter than the former. Regardless, the result is a critical insight to policy makers on what could be fixed in the government system. In addition to emphasizing progress in test scores, incorporating efficacy instruction into the curriculum might go a long way in improving self-efficacy, achievement motivation and learning, and in transforming schools to institutions of human development (Jinks and Morgan, 1999).

From a research perspective, this is only the second study in this literature to measure psychosocial skills after Singh (2015). The large self-efficacy effects point to the need for including non-tests score measures in future evaluations. In addition to self-efficacy, other psychosocial measures like self-concept and locus of control (Zimmerman, 2000) could also be deployed to obtain a more comprehensive assessment of childrens' human development at school.

2.4.2 Elite school advantage

My elite school results establish that the value added at high-fee schools is only in English. These effects are significant for all definitions of elite school and are robust to modelling fee as a continuous variable. The English effect size is also an economically meaningful 0.58 σ. Further, there is no effect on self-efficacy, suggesting that self-efficacy improves owing to shifts from public to private rather than due to shifts within private (from low-fee to high-fee schools).

The English effects are not completely surprising given labor markets returns to proficiency in the language. Munshi and Rosenzweig (2006) use survey data from Mumbai, Indian's largest city, and demonstrate that English education enables lower-caste children to break away from traditional caste networks and benefit from globalization. Similarly, Chakraborty and Bakshi (2016) use data from West Bengal and estimate that there is an 8 percent decline in weekly wages corresponding to a 10 percent reduction in probability of learning English at primary school. Azam et al. (2013) use a nationally representative dataset and estimate the labor market returns are as high as 34 percent for men who speak fluent English, and about 13 percent for men who speak at least some English relative to those who do not speak English at all. The labor market premiums could be even higher for Karnataka, whose capital Bangalore (referred to as India's Silicon Valley) is the center of ICT and the call center economy.

The lack of overall test score effects is still puzzling, challenging the widely held view in the Indian context that high-fee means high quality. This is also the first evidence on price-quality relationship in primary education markets in developing countries. In a perfectly competitive private schooling market, with parents as buyers/consumers and private schools as producers, one would expect that higher price reflects higher objective quality. The failure of this relationship, in this case, has three possible explanations. First is the well-established information asymmetry problem in education markets: schools have more information than parents about the quality of education provided. Empirical evidence has established the weak price-quality relationship in informationally imperfect markets (Steenkamp, 1988). Hence the lack of overall learning effects could be due to the imperfect nature of private education markets in India.

The second explanation relates to understanding of parental preferences as consumers. Parents' might be choosing schools not based on academic value addition, but a host of non-academic variables like peer groups, school environment, facilities and safety. Hanushek (1981) posits that parents are not buying different educational services, but are choosing more compatible social and physical environments for themselves and their children. He further adds: "higher expenditures may have a consumption element and may aid higher income parents in segregating their children." Though this analysis comes from the consumer behavior of American parents, it is as relevant in the Indian context. Finally, though not highly probable, it might be argued that parents care only about gains in English and are getting what they want. Anecdotal evidence suggests that a combination of information asymmetries and parental preferences for non-academic features of schools account for the lack of overall test score effects. Detailed understanding of these mechanisms is an area for future research.

2.4.3 School choice effects

A central concern of this thesis was to estimate the impact of school choice i.e., the impact of expanding the choice set of schools available to parents by relaxing binding income constraints. The policy impact estimates in Chapter 1 would have been the school choice effects if the average policy applicant was an income-constrained default government or low-fee private school goer. However, I demonstrated that the average applicant is not income-constrained and was not exposed to expanded choice. Hence my data does not allow for the estimation of school choice effects.

The results of this Chapter, however, allow me to conjecture on what the school choice effects would have been had the RTE mandate worked as originally conceived. If the mandate led primarily to shifts from government to private sector, I would have observed large self-efficacy effects. Alternately, if the mandate led to shifts from low-fee to high-fee schools, I would have observed large English language gains. In either case, it appears that school choice would have some positive effects on human capital formation in the context of urban India, without being the silver bullet for improving educational outcomes of children from disadvantaged backgrounds.

2.4.4 Policy implications

These results have two key policy implications. First, though private schools improve self-efficacy, which could potentially cause better learning outcomes in the future, they cannot be the definitive solution for improving test scores of disadvantaged children. Hence school choice policies that seek to shift poor children from government to private schools cannot be the primary strategy for improving learning outcomes of disadvantaged children. Second, as government schools seem to be having disastrous consequences on childrens' self-efficacy, a component of efficacy education may be introduced in the government curriculum to correct this failure.

2.5 Conclusion

In this Chapter, I present estimates of the private school and elite school effects for children aged 7.3 years. I am unable to precisely estimate the private school effect on test scores given the small proportion of government school attending children in the sample. I, however, establish that private schools perform much better than government schools in one dimension: boosting self-efficacy of children. I also find that the elite/high-fee private school premium is only in English scores and that enrollment at these schools does not change overall learning outcomes.

These results point to some avenues for future research. First, the self- efficacy results are very large and robust. But given that they are based on a very small sub-sample of government school children, validation of these results in an experiment with higher compliance is necessary. Additionally, some qualitative work is required to understand the source of the large self-efficacy gains that I report. Second, though the evidence on English effects in elite schools is robust, the source of this effect and reasons for the absence of overall effects is unclear. This points to the need for studying the role of information asymmetry and parental preferences for non-academic aspects in choosing schools. Evidence in this area is completely absent in the context of developing countries.

References

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.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Ashley, L. D., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., and Rose, P. (2014). The role and impact of private schools in developing countries. Report, 2206, EPPI-Centre.

Azam, M., Chin, A., and Prakash, N. (2013). The returns to English-language skills in India. *Economic Development and Cultural Change*, 61(2):335–367.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191.

Bandura, A. (1982). The assessment and predictive generality of self-percepts of efficacy. *Journal of Behavior Therapy and Experimental Psychiatry*, 13(3):195–199.

Bartelink, J. A. (2016). Price-quality correlation in tenders. B.S. thesis, University of Twente.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Chakraborty, T. and Bakshi, S. K. (2016). English language premium: Evidence from a policy experiment in India. *Economics of Education Review*, 50:1–16.

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

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management*, 1(1):19–41.

Herr, A. and Hottenrott, H. (2016). Higher prices, higher quality? Evidence from German nursing homes. *Health Policy*, 120(2):179–189.

Imbens, G. and Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences : an introduction*. Cambridge University Press, New York, NY.

Jinks, J. and Morgan, V. (1999). Children's perceived academic self-efficacy: An inventory scale. *The Clearing House*, 72(4):224–230.

Judd, V. C. (2000). The price-quality relationship: an empirical study of food products. *Journal of Food Products Marketing*, 6(1):11–24.

Juneja, N. (2014). India's New Mandate against Economic Apartheid in Schools. *Journal of International Cooperation in Education*, 16:55–70.

Kingdon, G. G. (2017). The Private Schooling Phenomenon in India: A Review. IZA Discussion Paper No. 10612.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics*, 6(3):234–263.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96(4):1225–1252.

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.

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.

Srivastava, P. (2008). *School choice in India: disadvantaged groups and low-fee private schools*. Symposium Books.

Steenkamp, J.-B. E. (1988). The relationship between price and quality in the marketplace. *De Economist*, 136(4):491–507.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. Cengage Learning.

Zhang, H. (2012). The mirage of elite schools: evidence from lottery-based school admissions in China. Working Paper, Chinese University of Hong Kong.

Zhang, Z.-J., Zhang, C.-L., Zhang, X.-G., Liu, X.-M., Zhang, H., Wang, J., and

Liu, S. (2015). Relationship between self-efficacy beliefs and achievement motivation in student nurses. *Chinese Nursing Research*, 2(2):67–70.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, 25(1):82–91.

CHAPTER 3

How do parents choose schools? A discrete choice modelling approach

Abstract

I present evidence on the demand-side of the RTE 25 percent mandate by evaluating the determinants of parental school choices and exploring if parental preferences vary by household type and gender of the child. I estimate a Rank-Ordered Logit (ROL) model that uses data on school attributes, household/ child attributes and rank-ordered school preferences. I find that parents prefer schools with higher fees, size, and infrastructure and those that that are closer to home. These results broadly align with the existing evidence on determinants of choice from high-income countries. On heterogeneity, however, my results diverge. I find no difference in parental preferences, by household socioeconomic status (SES), for home-school distance and composition of the student body. This is contrary to evidence from OECD countries, particularly the US, where lower-SES parents seem to have a stronger preference for school proximity and composition relative to higher-SES parents. Finally, I find that parental preferences for school attributes are same for both boys and girls.

3.1 Introduction

A central assumption of school choice theory is that parents care most about school effectiveness i.e., schools' ability to improve childrens' test scores when choosing schools. From this flows the prediction that expanded choice will drive up the productivity of the education system. However, opponents of school choice have long argued that parents, especially those from lower socioeconomic status (SES) backgrounds, might not prioritize school effectiveness: they either do not have the required information or might prefer other attributes like proximity and school composition over academic effectiveness. This implies that the predicted gains from expanded choice might not materialize, and worse still, choice could have the harmful effect of increasing the achievement gap between children from different socioeconomic groups. Hence understanding the determinants of school choice and estimating if households across different SES have varied preferences for school attributes are central questions in the demand-side debate of school choice.

Several rigorous studies have investigated these questions in the context of OECD countries (Rothstein, 2006; Hsieh and Urquiola, 2006; Hastings et al., 2009; Hastings and Weinstein, 2008; Hanushek, 1981; Musset, 2012; Gallego and Hernando, 2009; Burgess et al., 2015). However, little is known about the demand-side of school choice in low- and middle-income countries (LMICs) where choice policies are becoming increasingly popular.[1]

[1] Carneiro et al. (2016) study on private schooling in rural Pakistan is the only rigorous work on the demand-side of school choice in LMICs. The nascent literature on school choice in these contexts have mostly focused on understanding the impact of choice policies.

In this chapter, I estimate the determinants of parental school choice and explore heterogeneities by household SES in India's third largest city, Bangalore.[2] The implementation of a national school choice policy, the RTE 25 percent mandate, provides the context. The study sample comprises a random set of 636 policy participant households whose children were entering grade I (average age: 6 years). To participate in the policy, parents submit ranked choices for fee-charging private schools within their neighborhoods. I use the stated school choices data and estimate a discrete choice model (McFadden et al., 1973; McFadden, 1978) of parental preferences for different school attributes. The specificity of the RTE application context implies that this investigation differs from the extant literature on the demand-side of school choice in two important aspects. First, I am unable to investigate parental preferences for schools' academic effectiveness, given the unavailability of school-level test scores data.[3] Second, I can only estimate preferences for attributes of fee-charging private schools and not for all the schools in the market.[4] Accordingly, I answer two specific questions in this chapter. First, what are the determinants of parental private schooling choices in the context of the RTE 25 percent mandate? Second, do households across different SES have varied preferences for school attributes?

My modelling strategy and data have several strengths. First, I have rank-ordered data for each parent enabling the estimation of a

[2] The official name of the city has been changed to Bengaluru in 2014. I will, however, use the commonly and widely recognized name Bangalore.

[3] There is no standardized testing of children in elementary grades in Indian private schools. Hence neither the parents nor the researcher observes school-level test scores of private schools.

[4] There are both public and private schools in all neighborhoods of Bangalore city, but the RTE mandate's application process constrains parents to only choose amongst private schools.

Rank-Ordered Logit (ROL) model (Chapman and Staelin, 1982; Hausman and McFadden, 1984). The ROL estimator is more efficient as it relies on a stronger source of variation for identification of model parameters compared to standard conditional logit (Hastings et al., 2009; Beresteanu and Zincenko, 2018).[5] Second, the place allocation algorithm used in Bangalore is based on a Random Serial Dictatorship (RSD) mechanism, wherein the dominant strategy for parents is to reveal their true preferences. Hence the school choices observed in my dataset reflect true parental preferences. Most previous work used data generated in response to non-strategy proof algorithms, like the Boston mechanism, raising concerns around identification of true parental preferences (Hastings et al., 2009). Third, choice is restricted to schools within neighborhoods in my context, thus enabling the unambiguous identification of choice sets for each household. This is critical for the estimation of discrete choice models. Finally, I have information on a range of school and household attributes that allow for building a well-specified model that can account for potential correlation of errors across choices. This implies that the property of 'Independence from Irrelevant Alternatives (IIA),' usually a concern in discrete choice logit models, is not a first order concern in this case.

My results reveal three key findings. First, parents in the urban Indian context use school fees as a marker for quality.[6] I find that parents have a preference for high-fee schools on average: the coefficient on the fees variable is positive for the top two SES terceiles and negative for the bottom one. A positive coefficient implies that price is used a signal for quality in these

[5] The only other study that used rank-ordered data in investigating the demand-side of school choice is Hastings et al. (2009) study on the efficacy of a public school choice policy in Charlotte- Mecklenburg in North Carolina.

[6] By quality I do not mean academic effectiveness alone, but a combination of academic and non-academic aspects of schooling that parents seem to value.

markets. Second, home-school distance is an key determinant of school choice and households, across SES terciles, prefer proximity equally. The lack of heterogeneity is contrary to evidence from US and Chile, where households' sensitivity for proximity decreases with increasing income. A final noteworthy finding is that parental preferences are not heterogeneous by gender. This result strengthens the evidence on absence of discrimination against young girls in intra-household resource allocation in the Indian context.

This chapter is primarily related to the literature on the demand-side of school choice. Earlier work in this literature focused on the choice between public and private schools (Alderman et al., 2001; Checchi and Jappelli, 2004), while more recent work examines whether parents care about test scores and investigates heterogeneity in preferences across parent types. Rothstein (2006) studies the demand side of school choice by evaluating the distribution of student outcomes across schools within metropolitan housing markets in the US and concludes that school effectiveness is not a primary determinant of parental choices. Hsieh and Urquiola (2006) study the impact of Chile's nation wide school choice program and argue that parents might prioritize peer groups over academic effectiveness of schools. On a similar note, Abdulkadiroğlu et al. (2014) find that the highly competitive and coveted exam school education in Boston and New York has very little effect even for the highest marginal ability applicants compared to regular public education. They conclude that parents seem to care about non-learning related attributes while choosing exam schools, and hence, the likelihood of parental choice having strong demand side effects on the production of human capital in schools is reduced. Gallego and Hernando (2009) and Neilson (2013) study the determinants of parental choice in Chile and find that parents care both about academic and non-academic attributes of schooling, implying that they face trade-offs between their preference for test scores and for other school attributes. Musset (2012) summarizes several other studies from OECD countries and argues that the determinants of parental school choices are complex and go beyond simple academically rational reasons.

On heterogeneous preferences by household SES, Hastings et al. (2009) find that higher-SES parents (in North Carolina (US)) prefer high-performing schools, while lower-SES families face trade-offs between their preference for schools' academic performance and composition of the student body. Similarly, Burgess et al. (2015), in their study of parental choices in England, find significant heterogeneity across SES groups in their demand for academic effectiveness. Their demand elasticities show that the highest SES group responds much more significantly to improvements in academic standards than the other SES groups. Musset (2012) argues that lower-SES parents are at a disadvantage in acquiring information on school effectiveness and hence underweight it.

My results, despite coming from the unique institutional context of the RTE 25 percent mandate, contribute to and enrich this strand of the school choice literature. Further, this is only the second study after Carneiro et al. (2016) from a non-OECD country context and the first study from the Indian context, where both the role of the private sector in elementary education and demands for introduction of school choice are on the rise.

The chapter also contributes to the small literature on gender discrimination in the Indian context. Several studies have reported higher educational expenditures for boys relative to girls and have indicated a pro-male bias in household resource allocation (Zimmermann, 2012; White et al., 2016). However, more nuanced studies have established the absence of bias against young children conditional on school enrollment (Azam and Kingdon, 2013; Kingdon, 2005). My result of homogeneous preferences for both boys and girls, therefore, contributes to and strengthens the existing evidence.

The rest of the chapter is structured as follows: Section 3.2 presents the institutional setting in which this study is conducted. Section 3.3 describes the empirical framework, while section 3.4 presents the sampling strategy and describes the data. Section 3.5 presents the results, while section 3.6 lays out a detailed discussion of the key findings. Section 3.7 concludes.

3.2 Institutional setting

3.2.1 School choice in Karnataka

The implementation of the India's national school choice policy, the RTE 25 percent mandate, in Karnataka state of south India provides the setting for this study. The mandate sets aside 25 percent of places in all private schools of the country for children from disadvantaged backgrounds. Government pays the tuition fee for these 25 percent places directly to private schools. A detailed discussion on the mandate and its implementation are in Chapter 1 of this thesis. I will restate some aspects of the implementation in Karnataka that are specifically relevant to the analysis in this chapter.

- Households belonging to certain disadvantaged social and economic groups are eligible to apply for free places under the mandate, subject to a neighborhood restriction.

- Implementation of the 25 percent mandate started from 2013 and majority of the private schools[7] in the state have been enlisted to participate in the mandate.
- A centralized online application and admission system was set up in 2015 to overcome the inefficiencies associated with a decentralized system, where parents had to physically visit every school they wanted to apply to.
- As demand for free places is higher than supply, a centralized lottery is conducted to allocate free places to applicants.
- The school year in Karnataka is from June- April. So the annual admission process for free places starts in February (when applications open) and the lottery is conducted in April.
- Parents were allowed to give a ranked preference order of a maximum of five schools within their neighborhoods in 2015 and 2016. In 2017, however, this restriction was relaxed and parents could nominate any number of schools within their neighborhoods.[8]

I use the parental school choices from the 2017 application cycle for this analysis. The setting allows me to observe the complete preference ranking of parents. This is a huge advantage as ranked choices give much more information about the underlying parental preferences compared to just knowing either the first choice of parents or the school the child is currently enrolled at, which is typically the case in many studies. Burgess et al. (2015),

[7] The only private schools that do not participate are schools run by religious and linguistic minorities (less than 10 percent of all the private schools in the state), which are exempt from the mandate.

[8] The Karnataka free places online application system first prompts parents to choose their neighborhood, after which a drop down list of all the schools in that specific neighborhood are presented for parents to choose from.

for instance, use only the information on the first choice of parents in estimating the determinants of choice in England. Similarly Carneiro et al. (2016), Gallego and Hernando (2009), and Neilson (2013) use information on the school attended by the child. To my knowledge, Hastings et al. (2009) work in North Carolina is the only school choice study to use complete preference ranking data.

3.2.2 School choice and choice sets

A fundamental requirement in estimating discrete choice models is to have a well defined choice set i.e, the choices in the set must be mutually exclusive, exhaustive and finite. In typical school choice policies (in developed country contexts) where parents are allowed to apply for schools outside their neighborhood, identification of the choice sets becomes tricky as the researcher does not observe the set of schools over which parents are making choices.[9] Burgess et al. (2015), for instance, faced with this situation conduct a detailed exercise using spatial data to define the feasible choice sets for each parent. The application system in my setting, by only allowing choices within the neighborhood, delineates the choice set for each parent and enables estimation of discrete choice models.

3.2.3 Lottery mechanism and potential for strategic choices

Modelling parental preferences based on stated school choices requires that the stated choices, as observed by the researcher, are parents' true

[9] Theoretically, the choice set could be the set of all schools both within and outside a parent's neighborhood. But practically, parents might not consider several schools while making choices. For instance schools that are either far from their home or those where the probability of admission is low might not be part of the parents' choice set.

preferences. Student assignment mechanisms used in the implementation of choice policies play a definitive role in determining if parents reveal their true preferences. Mechanisms in which revealing the true preferences is the dominant strategy are referred to as strategy proof. On the other hand, manipulable mechanisms, like the commonly used Boston mechanism[10] are not strategy proof, and school choices elicited under these mechanisms are not ideal for investigating parental preferences. Studies that use Boston mechanism based school choice policies to model parental preferences have to present additional arguments and/or data to strengthen their case that the nominated preferences are indeed the true preferences (Hastings et al., 2009).

Karnataka's RTE admissions lottery algorithm is based on the Random Serial Dictatorship (RSD) mechanism. The algorithm orders all applicants by a random lottery number, the first applicant is assigned her top choice school, the next her top choice amongst available school places and so on (Abdulkadiroğlu and Sönmez, 1998, 2003). Under the RSD mechanism, with no limit on the number of choices, revealing true preferences is the dominant strategy. Abdulkadiroğlu and Sönmez (2003) contend that the

[10] The Boston mechanism, also referred to as the first- choice maximizer, operates in multiple rounds as follows: First, students submit a preference ranking of schools. Schools also have rules for prioritizing certain students. For instance, in many cities in the US, students with a sibling attending the same school are given first priority and those living within walking distance of the school are given second priority. In round 1 of the lottery, the first choices of the students are considered and all places are filled based on the priority order of the school. In round 2, only the second choices of the remaining students are considered and so on till round k or till all places are filled (k is the highest number of choices expressed by any student). Glazerman and Meyer (1994) make the following observation about the manipulablity of the Boston mechanism: "It may be optimal for some families to be strategic in listing their school choices. For example, if a parent thinks that their favourite school is oversubscribed and they have a close second favourite, they may try to avoid "wasting" their first choice on a very popular school and instead list their number two school first."

mechanism is both Pareto efficient and strategy-proof.[11] Hence there is no scope for strategic choice in my context, and I can be confidant that the nominated ranked choices observed in the 2017 RTE applications data represent true parental preferences.[12]

3.3 Empirical framework

In this section I present an econometric model for estimating parental school preferences and lay out the strategy for estimating the model.

3.3.1 Economic Model

I model school choice of a household/parent as a discrete choice from amongst a defined set of schools. I use the classical discrete choice modelling approach that assumes utility maximizing behavior by the parent and is based on Random Utility Models (RUMs) developed by McFadden (1973; 1978) and as elucidated by Train (2009). Let U_{ijn} be the expected utility

[11] The version of the RSD mechanism used in Karnataka in 2015 and 2016 with restrictions on the number of choices is, however, not completely strategy proof. In RSD with restrictions on number of choices, the dominant strategy is to nominate atleast one safe school even if it is not highly preferred. The restriction on the number of choices was, however, eliminated in 2017 making the mechanism for that year strategy-proof. Further, the RSD mechanism with no restrictions on number of choices is, infact, Obviously Strategy Proof (OSP), meaning that it has an equilibrium in obviously dominant strategies. Li (2017, p. 3257) explains that "a strategy is obviously dominant if and only if a cognitively limited agent can recognize it as weakly dominant."

[12] Despite being strategy-proof the RSD mechanism is not typically deployed for allocating school places as local laws prescribe different priority orders at schools. For instance, in the US context, students with siblings already enrolled at a school are given higher priority at that school. Such local laws also exist in some Indian states that are implementing the RTE mandate. For instance, in Delhi, the national capital, local laws prescribe that students living within one kilometer radius of a school should be prioritized in admission to that school. These priority requirements, coupled with expanded choice outside the neighborhood, preclude the use of RSD mechanism. Karnataka, however, has no such priority constraints and choices are restricted to within neighborhoods, hence allowing the use of strategy-proof RSD mechanism.

of parent $i \in \{1, 2, ..., I\}$ from enrolling their child at school $j \in \{1, 2, ..., J\}$ in neighborhood $n \in \{1, 2, ..., N\}$.[13] This utility is known to the parent but not to the researcher. The behavioral model for the parent is to chose school j if and only if $U_{ijn} > U_{ikn} \ \ \forall j \neq k$.

Parental utility can be decomposed into an observed component/ representative utility, V_{ijn}, and a random unobserved one, ε_{ijn}, which includes idiosyncratic tastes of the parent and school characteristics not captured in the representative utility.

$$U_{ijn} = V_{ijn} + \varepsilon_{ijn} \tag{3.1}$$

The researcher does not observe parental utility but observes P attributes of the schools $\boldsymbol{X}_{jn} = \{x_{j1}, x_{j2}, ..., x_{jP}\}$, Q attributes of the households $\boldsymbol{Z}_{in} = \{z_{i1}, z_{i2}, ..., z_{iQ}\}$ and the home-school distance, d_{ijn}. Representative utility, V_{ijn}, is the function that relates these observed factors to the parental utility. It can be denoted as follows:

$$V_{ijn} = V(\mathbf{X}_{jn}, \mathbf{Z}_{in}, d_{ijn}) \tag{3.2}$$

V depends on unknown parameters that should be statistically estimated. But I conceal the dependence for now. Given the behavioral model, the probability that parent i chooses school j is:

$$P_{ijn} = Prob(U_{ijn} > U_{ikn}) \qquad \forall j \neq k \tag{3.3}$$

[13] I index neighborhoods in the notation because parents can only choose schools within a particular neighborhood.

3.3.2 Econometric Model

Following McFadden(1973; 1978) and assuming that each ε_{ijn} in equation (3.1) is independent and identically distributed (i.i.d) with a type I extreme value distribution, the choice probability of equation (3.3) reduces to a logit choice probability and has a closed form solution.[14]

$$P_{ijn} = \frac{e^{V_{ijn}}}{\sum_{l=1}^{J} e^{V_{iln}}} \qquad j = 1, ..., J \tag{3.4}$$

Representative utility is modelled as linear in parameters:

$$V_{ijn} = \boldsymbol{\alpha}.\mathbf{X}_{jn} + \beta.d_{ijn} + \boldsymbol{\theta}.\mathbf{Z}_{in} * \mathbf{X}_{jn} + \boldsymbol{\gamma}.\mathbf{Z}_{in} * d_{ijn} \tag{3.5}$$

The household attributes $\mathbf{Z}_{in}$ enter equation (3.5) only as interaction terms because equation (3.4) is estimated as a conditional logit model with fixed-effects at the household level. The vector of coefficients $\boldsymbol{\alpha}$ and the distance coefficient β are assumed to be common for all households. The coefficients on the interaction terms $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ capture the heterogeneity in preferences across households. Substituting equation (3.5) into equation (3.4) produces the conditional logit model:

$$P_{ijn} = \frac{e^{\boldsymbol{\alpha}.\mathbf{X}_{jn} + \beta.d_{ijn} + \boldsymbol{\theta}.\mathbf{Z}_{in} * \mathbf{X}_{jn} + \boldsymbol{\gamma}.\mathbf{Z}_{in} * d_{ijn}}}{\sum_{l=1}^{J} e^{\boldsymbol{\alpha}.\mathbf{X}_{ln} + \beta.d_{iln} + \boldsymbol{\theta}.\mathbf{Z}_{in} * \mathbf{X}_{ln} + \boldsymbol{\gamma}.\mathbf{Z}_{in} * d_{iln}}} \qquad j \in \{1, 2, ..., J\} \tag{3.6}$$

This model can be estimated if parents preferred school from amongst the

[14] If V is modelled only as a function of household attributes $\mathbf{Z}_{in}$, the estimator is a multinomial logit. However, if V is modelled as a function of school attributes $\mathbf{X}_{jn}$ or as a function of school and household attributes as described in equation (3.2), the appropriate estimator is a conditional logit.

choice set is known. However, if complete rank-ordered data of parental choices is available, instead of just a single choice, the model can be extended to exploit the additional information in the data and to estimate the parameters efficiently (Chapman and Staelin, 1982; Hausman and McFadden, 1984; Beresteanu and Zincenko, 2018). Chapman and Staelin (1982, p.291) illustrate how rank-ordered choices can be decomposed into multiple statistically independent choice observations. They formalize the rank order "explosion rule" definition:

> Given rank ordered choice set data, the original rank ordered observation that decision maker i prefers the alternatives in the order $1, 2, ..., J_i$ - i.e., that $U_{i1} \geq U_{i2} \geq ... \geq U_{iJ_i}$ - can be"exploded" (decomposed) into $J_i - 1$ statistically independent choice observations of the form $(U_{i1} \geq U_{ij}, j = 1, 2, ..., J_i)$, $(U_{i2} \geq U_{ij}, j = 2, 3, ..., J_i)$,..., $(U_{iJ_{i-1}} \geq U_{iJ_i})$.

Hence a ranking of J alternatives can be represented as $J - 1$ independent choices. Train (2009) refers to these independent choices generated by the explosion process as pseudo-observations. In Table 3.1, I illustrate the process of explosion of a full rank-ordered choice set. Consider the rank ordered data of one student whose choice set comprises five schools: A, B, C, D and E. If the student's rank order is as shown in column (2), then four pseudo-observations can be generated as shown in columns (3)- (5).

Table 3.1: Explosion strategy with complete rank-ordered data

(1) School	(2) Rank order	(3) Pseudo-observation	(4) School	(5) Choice
A	1	1	A	1
B	2	1	B	0
C	3	1	C	0
D	4	1	D	0
E	5	1	E	0
		2	B	1
		2	C	0
		2	D	0
		2	E	0
		3	C	1
		3	D	0
		3	E	0
		4	D	1
		4	E	0

The explosion strategy can also be used for partially rank-ordered choice sets. The unranked choices cannot be exploded, while the ranked ones can be in the usual way following the explosion rule (Chapman and Staelin, 1982). The number of pseudo-observations is equal to the number of choices expressed (the depth of the rank-ordered information). Table 3.2 illustrates the explosion strategy for a partially rank-ordered situation: a student with five schools in her choice set only chooses 3.

Table 3.2: Explosion strategy with partially rank-ordered data

(1) School	(2) Rank order	(3) Pseudo-observation	(4) School	(5) Choice
A	1	1	A	1
B	2	1	B	0
C	3	1	C	0
D	0	1	D	0
E	0	1	E	0
		2	B	1
		2	C	0
		2	D	0
		2	E	0
		3	C	1
		3	D	0
		3	E	0

The psudo-observations thus generated provide a stronger source of variation than in cases where just the first choice is observed. As explained by Hastings et al. (2009, p.17):

> Intuitively, when only a single (first) choice is observed for each individual, it is difficult to see whether an unexpected choice was a result of an unusual error term (ε_{ij}) or an unusually high weight placed by the individual on some aspect of the choice (β_i). However, when an individual makes multiple choices that share a common attribute (e.g., high test scores) we can infer that the individual has a strong preference for that attribute, because independence of additive error terms across choices would make observing such an event very unlikely in the absence of a strong preference.

Given that I use rank-ordered data in the analysis, the econometric model I estimate should more appropriately be called a Rank-Ordered Logit (ROL) model or an Exploded Logit model, where rank-ordered/exploded refer to the conditional logit model incorporating rank-ordered data.

3.3.3 Model Assumptions

The logit model of equation (3.4) is derived from the assumption of errors being distributed i.i.d and extreme value, with the key assumption being that the errors are independent across alternatives (schools).[15] The independence assumption can be stated formally as follows: $Corr(\varepsilon_{ijn}, \varepsilon_{ikn}) = 0,\ \forall j \neq k.$

This assumption leads to the Independence from Irrelevant Alternatives (IIA) property, also known as Luce's choice axiom (Duncan, 1959), according to which the relative odds of choosing between two alternatives j and k depend only on the attributes of those alternatives and not on the availability or attributes of the other alternatives. IIA implies proportional substitution, which is unrealistic in many circumstances.[16] Hence IIA could be a concern while using conditional logit models, and its violations would produce biased estimates. However, as elucidated by McFadden et al. (1977) and Train (2009), whether IIA is valid or not depends on how well the representative utility (V_{ijn}) is specified. McFadden et al. (1977) note:

> IIA property is or is not valid for a particular specification of representative utility in a logit model of a particular choice situation, not for the choice situation itself.

[15] As explained in Train (2009), the difference between extreme value and normal errors is empirically indistinguishable. Hence the key logit assumption is not about the shape of the distribution, but that the errors are independent across choices.

[16] The idea of proportional substitution and its implausibility were elucidated by McFadden et al. (1973) with the celebrated 'red-bus/blue-bus' example. Suppose two modes of travel are available to an individual, a car and a red bus, and that the probability of choosing each is 0.5. So the odds of taking a car rather than a red bus is 0.5/0.5= 1. IIA requires that when a new alternative becomes available, the probabilities should adjust in exactly the same amount to retain the original odds. If a new alternative, say a blue bus, becomes available, the only way to retain the original odds (of taking a car rather than a red bus) is for the probability of each of the three modes to be 0.33. This is what the logit model would estimate. However, it is intuitive to see that the actual probabilities will be 0.5 for car, and 0.25 each for the red bus and the blue bus. These actual probabilities violate IIA as the odds of car versus red bus now becomes 0.5/0.25= 2.

Intuitively, IIA implies that variables omitted from the model are independent across alternatives in a way analogous to independent error terms in OLS regression (McFadden et al., 1977; Cheng and Long, 2007). Hence if representative utility is well specified to capture all sources of correlation between the error terms, IIA would not be a concern.[17]

The rich data on school attributes that I assembled (details in next section) allow representative utility, V_{ijn}, to be sufficiently well specified. I am, hence, convinced that the errors across alternatives, ε_{ijn} and ε_{ikn}, are uncorrelated, and that IIA is not a concern for this analysis.[18] Further, violations of the logit assumption are less problematic when estimating average preferences than when forecasting substitution patterns (Train, 2009, p.40). Given that I am only attempting the former, even the skeptical reader can be assured that the ROL estimator is appropriate in this case.

A second concern is that of covariate-independence or the classic endogeneity issue: the logit model assumes no correlation between the school attributes specified in the model and the error term. More formally, the assumption is: $Corr(X_{ijn}, \varepsilon_{ijn}) = 0$. This endogeneity was ignored in traditional discrete choice studies. However, starting with Berry et al. (1995), strategies to overcome it have been developed in the industrial organization literature. The Berry, Levinsohn and Pakes (BLP) approach deals with the endogeneity by introducing an additional alternative specific characteristic

[17] Another way to deal with the IIA issue is to use alternate modelling strategies like the nested logit model, multinomial probit model or the mixed logit model that relax the independence assumption (McFadden, 1984). However, estimation of these models involves more complicated data structures and intensive calculations.

[18] Tests like the Hausman-McFadden test (Hausman and McFadden, 1984) and Small-Hsiao test (Small and Hsiao, 1985) have been proposed to evaluate the validity of IIA empirically. However, latest evidence shows that these test are unsatisfactory for applied work (Cheng and Long, 2007).

in the representative utility.[19] The estimation involves using instrumental variables that are usually hard to find. The BLP method has recently been used in school choice studies focusing on demand estimation (Gallego and Hernando, 2009; Neilson, 2013; Carneiro et al., 2016). The goal of the analysis in this chapter, however, is neither to do demand estimation nor to build a causal model. I am estimating a predictive model, where minimizing a combination of bias and estimation variance is the primary goal, in contrast to causal models where the focus is on minimising the bias (Shmueli, 2010). Further, the ROL estimator has a smaller variance than regular logit, hence, increasing the empirical precision of my estimates (Beresteanu and Zincenko, 2018).

3.4 Study sample and data

3.4.1 Study population

The study population comprises of the cohort of children (average age: 6 years) that applied for entry into RTE free places in 2017 in Bangalore city, the capital of Karnataka state. Free places are available at entry grades of schools, which in Karnataka are kindergarten and Grade I. I only use data of Grade I applicants as I expect parents to invest more effort in choosing schools at Grade I than kindergarten. Locating this study in Bangalore city is appropriate for many reasons. The city has the largest number of schools per neighborhood compared to any other region/ district of the state, reasonable within neighborhood heterogeneity in schools and large

[19] The representative utility at equation (3.2) becomes: $V_{ijn} = V(\mathbf{X}_{jn}, \mathbf{Z}_i n, d_{ijn}, \lambda_{jn})$, where λ_{jn} is the school characteristic that is unobserved by the econometrician, but observed by both households and schools, and valued exactly the same by all households.

enough neighborhoods to estimate parental preference for proximity. Table 3.3 provides details of the study population.

Table 3.3: Study population details

Item	Value	Comments
Applicants to Grade I	21,295	25 percent of all applicants in the state
Neighborhoods	193	
Schools	1,359	About 18 percent of all schools in the state
Schools per neighborhood (mean)	7	Standard deviation (SD) = 5.05
Schools per neighborhood (median (p50))	6	90^{th} percentile (p90)=14
Preferences (p50)	4	p90=7
Neighborhoods with only one school	11	
Applicants per neighborhood (mean)	110	SD= 78, Max= 351, Min=2

As the study population comprises only those households applying to the 25 percent mandate, a very specific sub-group of the population, it is necessary to describe how similar or otherwise this population is from the wider population of households with 6-year old children. The 25 percent free places are reserved for socially and economically disadvantaged groups, defined as households with income below INR 100,000 (USD 1,541). This income eligibility cutoff is slightly lower than the state per capita income of INR 126,976 (USD 1,956). So if the mandate were implemented in strict conformity with the income eligibility cutoff, my study population would only comprise of households with income below the state average per capita income.[20] However, I demonstrated in Chapter 1 of this thesis that the policy is mistarteged in Karnataka leading to many ineligible households applying for the program. I estimate the income of applicant households to be orders of magnitude higher than the eligibility cutoff. For Bangalore city in particular, the estimated average income of the applicant households is INR

[20] Given that I do not have the income distribution of all households with 6-year old children, this is the best guess of where the applicants fall in the overall income distribution.

561,443, five times the eligibility cutoff, four times the state per capita income and almost twice the per capita income of Bangalore city (INR 271, 378). I also established that the average policy applicant is a default private school goer with ability and willingness to afford the median private school fee. All this suggests that the income/ SES distribution of RTE applicant population is broadly similar to that of the the wider population of households with 6-year old children, sans the tails.[21]

3.4.2 Study Sample

To estimate the parameters of the ROL model, data on a sample of parents from the study population is required. As parents live in neighborhoods with identical choice sets, I adopted a two-stage cluster sampling strategy with neighborhoods as the primary sampling units. I used Probability Proportional to Size (PPS) strategy for first stage sampling as the number of applicants per neighborhood/ cluster is varied. PPS sampling allows that larger clusters have bigger probability of being sampled. For the second stage, I sampled exactly the same number of children per cluster, thus ensuring that the probability of being sampled is same for each child in the study population. Table 3.4 gives details of the planned and realized study sample.

[21] The left tail of the income/SES distribution of the wider population are the default government school-going, extremely poor households that did not apply to the mandate, and the right tail is the ultra wealthy households who also would not apply for the mandate.

Table 3.4: Study sample details

Item	Value	Comments
Clusters in Study population	193	11 neighborhoods have only one school
Clusters in the sampling frame	182	Single school neighborhoods not included
First stage: sampled clusters	60	PPS sampling
Second stage: students per cluster	10	Simple random sampling
Planned sample size	600	Additional households surveyed in few clusters
Realized sample size	636	

The neighborhood map of Bangalore city with the sampled neighborhoods highlighted is at Figure A.3.1 in the Appendix to this Chapter.

3.4.3 Data

The following data is required on the study sample to estimate the parameters $\boldsymbol{\alpha}$, β, $\boldsymbol{\theta}$, and γ of the ROL model at equation (3.6).

- The set of choices available to each parent.
- The rank order of parental choices.
- The values of $\mathbf{X}_{jn}$, $\mathbf{Z}_{in}$, and d_{ijn}.

I use several data sets. The rank-ordered data of parental choices, the dependent variable in the analysis, and data on choice sets for each neighborhood are obtained from Karnataka Education department's RTE admissions database. Data on household/ student level variables ($\mathbf{Z}_{in}$) is obtained from a primary household survey conducted on all sample households in January and February 2018. Data on school level variables ($\mathbf{X}_{jn}$) is obtained from multiple government data sources. I geo-coded all sample households and schools in the 60 sampled neighborhoods and calculated home-school distance (d_{ijn}) using Stata's *georoute* command.

Table 3.5 lists all the variables used in the analysis and presents the descriptive statistics.

Table 3.5: Descriptive statistics

Variable	Mean	Standard Deviation	Median	Percentiles 10th	90th	95th
School Level Variables						
School fees (in INR)	23730.30	30562.49	14508.00	7757.00	46943.00	63660.00
Size (of student body)	471.87	430.47	342.00	74.00	1078.00	1290.00
Rooms	19.54	14.04	16.00	6.00	35.00	45.00
Computers	15.52	17.79	10.00	1.00	36.00	45.00
Graduate teachers	6.87	10.02	3.00	0.00	18.00	27.00
Proportion of SC/ST students	0.14	0.11	0.12	0.02	0.27	0.32
N	555					
Household Level Variables						
Age of child (years)	6.25	0.27	6.22	5.91	6.66	6.73
Gender (dummy, boy=1)	0.54	0.50				
Mother's education (years)	9.50	3.41	10.00	5.00	13.00	14.00
Father's education (years)	8.61	4.03	10.00	0.00	13.00	14.00
Children in family	2.05	0.80	2.00	1.00	3.00	4.00
Family size	4.37	1.20	4.00	3.00	6.00	7.00
Asset index	9.84	4.10	9.00	5.00	15.00	17.00
SES index	-0.00	1.26	0.18	-1.78	1.48	1.83
N	636					
Home- School Distance (in kilometres)						
Distance (first choice)	2.20	2.90	1.28	0.35	4.91	7.05
Distance (all choices)	2.52	3.31	1.50	0.43	5.16	7.73

3.4.4 Description of School Attributes

I now explain the construction of and the rationale for each of the seven school attributes used in the estimation.

- *School fees*: This variable is the value of the annual tuition fee charged by the school in Indian Rupees (INR). All schools in the sample are private schools charging a non-zero fee. Further, the fee distribution is log normal with a large number of low to medium fees schools and a smaller number of high-fee schools.[22] Figure 3.1 shows the fee kernel density plot of 555 schools in the 60 study neighborhoods. Data on this variable is missing for 50 of the 555 sampled schools (9 percent).[23] The missingness is not concentrated in specific neighborhoods: 50 missing schools are spread across 29 neighborhoods. Neither is it correlated with specific preferences. Hence the missingness is random, and the estimated parameters are valid for the study population.

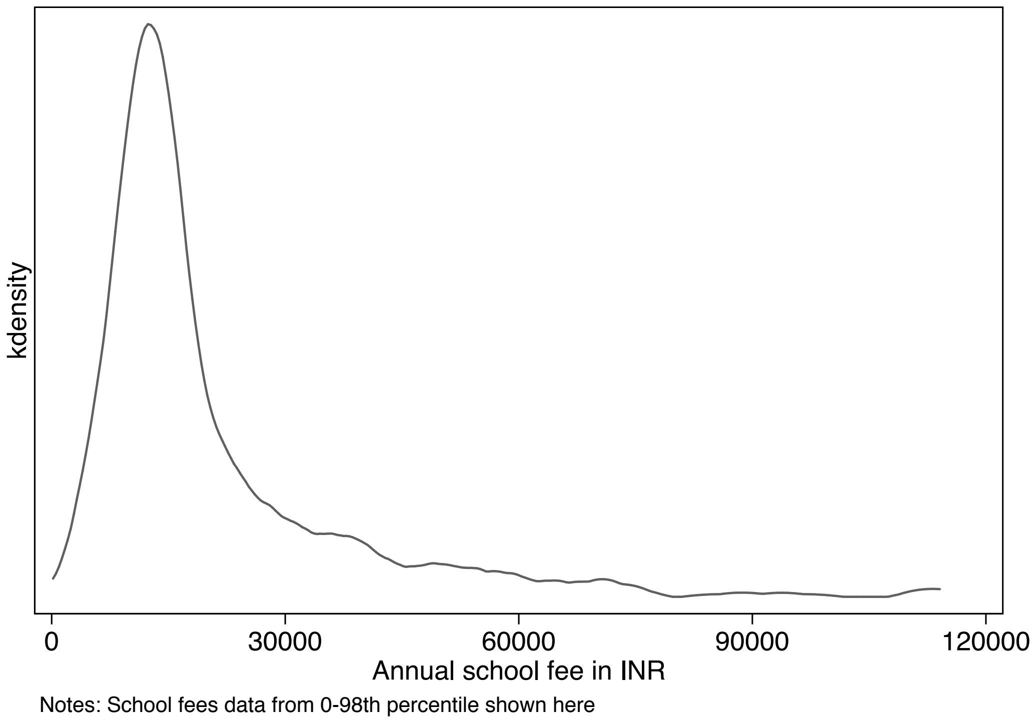


Figure 3.1: Distributional plot of annual school fee

- *Size, Rooms, Computers, Graduate teachers*: These variables measure the total number of children attending school and the number of rooms,

[22] A detailed discussion on the heterogeneity in the fee structure across private schools and definitions of high and low-fee schools is in Chapter 2 of the thesis.
[23] Some schools did not report their fee information on the government fee portal.

computers and graduate teachers available in the school respectively. While size could be an important preference variable by itself, rooms, computers and graduate teachers are proxies for school infrastructure/ quality. The median values of these variables are smaller than the means (Table 3.5) indicating that the distributions are log-normal with long right tails similar to the fee distribution.

- *Proportion of SC/ST*: This is a measure of the proportion of children from Scheduled Caste (SC) and Scheduled Tribe (ST) backgrounds. SCs and STs are socially and economically disadvantaged and I use this variable as a proxy for proportion of lower-SES children at school.

I, therefore, have data on a rich set of school attributes that are important to explore the determinants of choice amongst private schools in Karnataka. In Table 3.6, I present the cross-correlations of the different school level variables. Almost all the correlations are statistically significant. Except the correlations involving the proportion of SC/ST students, all pairs of variables are positively correlated. The magnitude of some of the correlations between size and other variables is large as would be expected.

Table 3.6: Cross-correlations between school attributes

Variables	Fees	Size	Rooms	Computers	Graduate teachers	Proportion of SC/ST students
Fees	1.000					
Size	0.245 (0.000)	1.000				
Rooms	0.135 (0.003)	0.661 (0.000)	1.000			
Computers	0.302 (0.000)	0.616 (0.000)	0.591 (0.000)	1.000		
Graduate teachers	0.135 (0.002)	0.415 (0.000)	0.426 (0.000)	0.333 (0.000)	1.000	
Proportion of SC/ST students	-0.160 (0.000)	-0.225 (0.000)	-0.112 (0.009)	-0.174 (0.000)	-0.014 (0.748)	1.000

Notes: Pair-wise correlations are presented here. Statistical significance of each correlation is in parenthesis

3.4.5 Description of Household Attributes

As discussed in the section 3.4.2, the study sample is broadly representative of the wider population of households with primary school aged children. Hence, I expect sufficient variation in the sample along different household attributes to enable the investigation of heterogeneity by household type. The first six household variables in Table 3.5 are self-explanatory. Below is a discussion on the asset index and SES index.

- *Asset index*: Opponents of school choice have argued that it might not help lower-SES children, as their parents might not prioritize school performance relative to higher-SES parents. To study heterogeneity, a robust measure of SES of households is necessary. Reliable data on income or SES of households is unavailable in developing countries like India.[24] Researchers have typically relied on assets as a good proxy for household income. I collected data on asset ownership for 15 assets and constructed an index by weighting each asset with the inverse of the probability of ownership.[25] The asset index measure has enough variation to be useful for differentiating households. Figure 3.2 shows the histogram and kernel density plot of the asset index variable.

- *SES index*: The asset index only captures one aspect, albeit a very important one, of the SES of households. Typically SES indices include information on income, education and occupation of household members. In the context of parental school choices, education of

[24] In fact, measuring household income is almost impossible as more than 90 percent of the labor force works in the informal sector, and less than five percent of the population has tax records.

[25] Commonly owned assets are given a lesser weight than rarer and high value assets. For example, a laptop has a higher weight than a television.

parents is of particular importance along with income (for which I use asset index as a proxy). I, therefore, create an SES index combining asset index, mothers' education and fathers' education using principal component analysis.[26] Figure 3.2 shows the histogram and kernel density plot of the SES index variable.

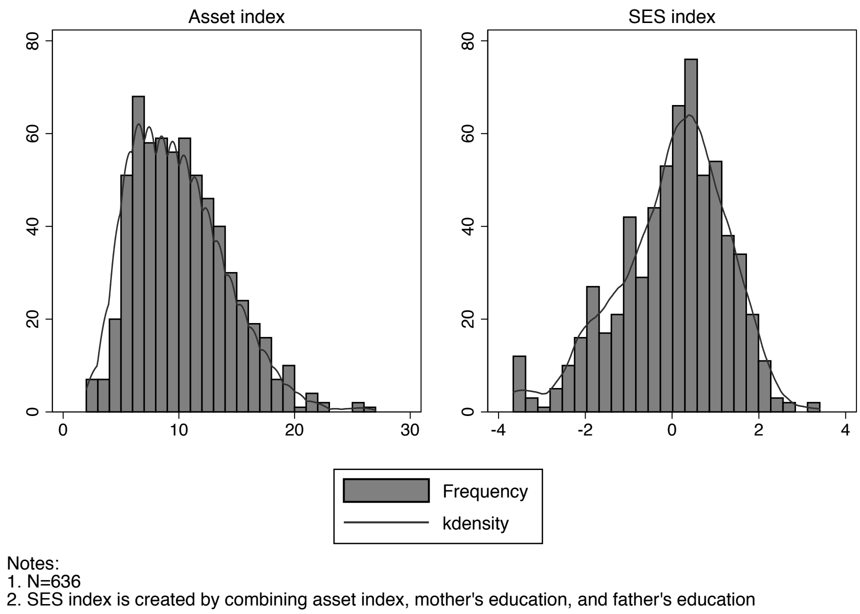


Figure 3.2: Distributions of asset and SES indices

I will present heterogeneity analysis along two dimensions: household SES and gender, and do robustness checks with the asset index measure. The reason for expecting heterogeneous preferences across SES has already been discussed. The gender dimension has not been adequately emphasized in the literature from OECD countries as there is no salient gender bias in those contexts. However, gender bias remains a significant issue in South Asia. Carneiro et al. (2016), for instance, in their work on school choice in Pakistan estimate the choice models separately for girls and boys.

[26] The PCA has three components and the percent of variance explained by the first component is 0.53.

3.5 Results

3.5.1 Parsimonious model: only school attributes

I first estimate a parsimonious ROL model using only school attributes as explanatory variables. In all the models the standard errors are clustered at the (more conservative) neighborhood level rather than the child/ household level.[27] I estimate the model on the full-sample[28] using seven school attributes and dummies for school type indicating the highest grade at school.[29]

The results, regression coefficients and marginal effects, are presented in Table 3.7. All the explanatory variables shown in the table are in standardized z-score units. The results indicate that home-school distance and size of the school (number of children enrolled) are the strongest determinants of school choice. The direction of the effects is unsurprising, the coefficients on these variables are significant at the one percent level and the marginal effects are

[27] McKenzie (2018) explains that when inferences are to be made for the population based on a sample from few clusters (neighborhoods in this case), the right level for clustering standard errors is the neighborhood. Without neighborhood level clustering, the estimands would still be correct for the sample but not for the population.

[28] The dataset for this analysis comprises 2,391 psudo-observations obtained by expanding the rank-ordered data of the 636 children in the sample as explained at Table 3.2. The dataset is in long form as required for the estimation of clogit model and has 20,478 observations (rows in the dataset). The clogit model imposes fixed effects at the pseudo-observation level; given the missing data problem (fee data is missing for about 10 percent of the schools) about 13 percent of the psudo-observations are not usable. That leaves 2,076 pseudo observations/ 16,773 observations/ 606 children, which comprise the full estimation sample.

[29] Schools in Karnataka (and across the country) can be broadly divided into four categories based on the the highest grade taught. Primary schools go from grades 1-5, Upper primary schools from grades 1-8, Secondary schools from grades 1-10, and Higher secondary schools from grades 1-12. Of the 555 schools in the sampled neighborhoods, 50 percent are secondary schools, 22 percent are higher secondary schools, 22 percent are upper primary, and 6 percent are primary. I use school type dummies in the model because the highest grade at school could be an important determinant of choice- parents will probably prefer schools that have grades till 10 or 12 over schools whose highest grades are only 5 or 8.

meaningfully large: a one standard deviation increase in size leads to a 10.4 percentage point (pp) increase in probability of a school being chosen, and a one standard deviation increase in distance leads to a 8.5 pp reduction in probability of a school being chosen. The coefficient on school fee is positive and weakly significant and the marginal effect is small.

Table 3.7: Determinants of school choice- parsimonious model

	(1) Coefficients	(2) Margins
Fees	0.067*	0.015*
	(0.038)	(0.009)
Size	0.453***	0.104***
	(0.062)	(0.014)
Rooms	-0.040	-0.009
	(0.063)	(0.015)
Computers	0.025	0.006
	(0.072)	(0.016)
Graduate teachers	-0.026	-0.006
	(0.037)	(0.009)
Proportion of SC/ST	-0.077	-0.018
	(0.070)	(0.016)
Distance	-0.367***	-0.085***
	(0.126)	(0.028)
Observations	16,773	16,773
Clusters	58	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parenthesis and are clustered at the neighborhood level to account for the sampling design. The model also includes dummies for school type (based on the highest grade at school) in additional to the school attributes shown in the table. All the explanatory variables, except the indicator variables for school types, are standardized. The ROL coefficients are at column (1), while the average marginal effects are at column (2).

3.5.2 Main model: both school and household attributes

I estimate the main model- ROL model at equation (3.6)- using the seven school attributes and one household attribute: household SES. I divide the sample into three SES terciles, create dummies for each of them and interact the dummies with the (standardized) school variables. Table 3.8 presents the results of a single regression, allowing the preferences of each school characteristic to vary by household SES. Column (1)- (3) show the coefficients of the interaction terms, and columns (4)- (6) show the marginal effects for each SES tercile. Column (7) presents the p-value of the F-test of joint significance of the three coefficients at columns (1)-(3). And column (8) presents the p-value of the F-test of joint significance of the two coefficients at columns (2)-(3). A p-value less than 0.1 in column (7) indicates that the school attribute is significant, and a p-value less than 0.1 in column (8) indicates heterogeneity across SES on that school attribute.

In addition to school fees, size and home-school distance, which were significant in the parsimonious model, number of computers is also a significant determinant of school choice. Further, the preferences for fees, size, number of computers and number of graduate teachers is heterogeneous across SES types, while there is no heterogeneity in household preference for distance.

Table 3.8: ROL model, heterogeneity by household SES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Coefficients			Average marginal effects		p values of F-tests	
	Ref group Top Tercile	Interaction: Middle Tercile	Interaction: Bottom Tercile	Ref group Top Tercile	Interaction: Middle Tercile	Interaction: Bottom Tercile		
Fees	0.114**	-0.044	-0.124**	0.026**	-0.010	-0.028**	0.025	0.016
	(0.054)	(0.057)	(0.049)	(0.012)	(0.013)	(0.011)		
Size	0.325***	0.304***	0.122	0.075***	0.070***	0.028	0.000	0.006
	(0.088)	(0.099)	(0.114)	(0.020)	(0.023)	(0.026)		
Rooms	-0.083	-0.014	0.116	-0.019	-0.003	0.027	0.432	0.278
	(0.103)	(0.103)	(0.116)	(0.024)	(0.024)	(0.027)		
Computers	0.206***	-0.261***	-0.303**	0.047***	-0.060***	-0.069**	0.001	0.006
	(0.055)	(0.082)	(0.120)	(0.013)	(0.019)	(0.028)		
Graduate teachers	0.036	-0.130**	-0.058	0.008	-0.030**	-0.013	0.163	0.083
	(0.050)	(0.058)	(0.063)	(0.012)	(0.013)	(0.014)		
Proportion of SC/ST	-0.135	0.135	0.002	-0.031	0.031	0.001	0.242	0.343
	(0.101)	(0.097)	(0.112)	(0.023)	(0.022)	(0.026)		
Distance	-0.253***	-0.227	-0.175	-0.058***	-0.052	-0.040	0.012	0.239
	(0.084)	(0.157)	(0.234)	(0.019)	(0.036)	(0.053)		

N= 16,773; Number of children= 606

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are in parenthesis and are clustered at the neighborhood level to account for the sampling design. The model also includes dummies for school type (based on the highest grade at school) in additional to the school attributes (standardized) and interactions of school attributes and the SES tercile dummies. Columns (1)-(3) present the regression results of a single regression and columns (4)-(6) present the average marginal effects of the same regression. Column (7) and (8) present p-values of the F-test that check for joint significance of coefficients in columns (1)-(3) and (2)-(3) respectively. A p-value less than 0.1 in column (7) indicates that the school attribute is significant, and a p-value less than 0.1 in column (8) indicates heterogeneity across SES on that school attribute. The values of the coefficients and the marginal effects for the top tercile are those in columns (1) and (4) respectively; for the other two terciles, the coefficient in the respective column should be added to that in the reference group.

I then estimate demand elasticities (shown in Table 3.9) for the variables that explain parental preferences: fees, distance, size and computers. The calculations assume equal market shares for schools and show the percentage change in a schools demand/ market share that is associated with a one-percent change in the attribute.[30] These elasticities enable consistent comparison across households that have different types of schools in their choice sets (Burgess et al., 2015, p.1281). I elaborate on the magnitudes and the implications of these elasticities in the discussion section.

Table 3.9: Elasticity of demand with respect to school attributes

	(1) Initial share of demand Q	(2) Mean school characteristic X	(3) Average	(4) Top Tercile	(5) Middle Tercile	(6) Bottom Tercile
			Elasticity of demand			
Share of population				0.35	0.33	0.32
Fees (1000 INR)	0.105	28.98	1.55	2.96	1.83	-0.25
Size (100 children)	0.105	7.4	3.07	2.15	4.16	2.96
Computers	0.105	21.98	0.45	4.04	-1.08	-1.92
Distance (km)	0.105	2.89	-0.99	-0.97	-1.24	-1.11

N= 16,773; Number of children= 606

Notes: Average schools per neighborhood is 9.48, so Q=1/9.48= 0.105. Elasticity is $\frac{\partial Q}{\partial X} * \frac{X}{Q} = \beta * X * (1-Q)$ for the top terceile, which is the reference interaction group. For the other two groups the elasticity is $(\beta + \beta_k) * X * (1-Q)$, where k is the tercile number of the SES group. The average is the weighted average of the elasticities of each tercile weighted by their population share.

[30] Details on elasticity calculations are in notes to the table.

I further estimate the main model with the seven school attributes and two household attributes: household SES and child gender. The explanatory variables in the model are the seven school attributes (standardized), the interactions between school attributes and SES tercile dummies and interactions between school attributes and the gender dummy.[31] As discussed in section 3.4.5, gender is a critical dimension to explore heterogeneity in the Indian context. Table 3.10 presents the results. Distance, fees, size, computers and test scores are still the determinants of choice. The heterogeneity pattern is also similar to Table 3.8, except for the graduate teachers variable, where the preferences are no longer varied by household SES. The most striking result is that none of the coefficients in column (4) are significant, and the p-value of the joint significance test of all the gender interactions is 0.41. This implies that there is no difference in parental preferences for school attributes by the gender of the child.

3.5.3 Robustness checks

The main results presented at Tables 3.8 and 3.10 are robust to replacing the household SES variable with the asset index variable. These results are at appendix Tables A.3.1 and A.3.2.

[31] I make the reasonable assumption that there is no additional interaction effect between SES and child gender.

Table 3.10: ROL model, heterogeneity by household SES and gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficients				p values of F-tests	
	Reference category	Interaction: SES (Middle Tercile)	Interaction: SES (Bottom Tercile)	Gender (Girl)		
Fees	0.103*	-0.044	-0.127**	0.026	0.056	0.020
	(0.058)	(0.055)	(0.050)	(0.041)		
Size	0.329***	0.299**	0.107	0.011	0.000	0.008
	(0.094)	(0.100)	(0.115)	(0.086)		
Rooms	-0.110	-0.022	0.102	0.074	0.436	0.305
	(0.115)	(0.104)	(0.116)	(0.071)		
Computers	0.217**	-0.263**	-0.305**	-0.023	0.003	0.005
	(0.076)	(0.081)	(0.119)	(0.063)		
Graduate teachers	0.064	-0.119**	-0.044	-0.076	0.203	0.127
	(0.052)	(0.059)	(0.064)	(0.050)		
Proportion of SC/ST	-0.149	0.139	0.002	0.025	0.363	0.321
	(0.111)	(0.097)	(0.113)	(0.079)		
Distance	-0.301**	-0.258	-0.210	0.139	0.007	0.179
	(0.107)	(0.163)	(0.230)	(0.124)		
N= 16,773; Number of children= 606						

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parenthesis and are clustered at the neighborhood level to account for the sampling design. The model also includes dummies for school type (based on the highest grade at school) in additional to the school attributes, interactions of school attributes and the SES tercile dummies and interactions of school attributes with the gender dummy. Columns (1)-(4) present the regression results of a single regression. Column (5) and (6) present p-values of the F-test that check for joint significance of coefficients in columns (1)-(4) and (2)-(3) respectively. A p-value less than 0.1 in column (5) indicates that the school attribute is significant, and a p-value less than 0.1 in column (6) indicates heterogeneity across SES on that school attribute. The results in column (4) indicate heterogeneity along the gender dimension. The p-value of the F-test of joint significance of all the gender interactions in column (4) is 0.41.

3.6 Discussion

I set out to identify the determinants of private school choice and the heterogeneity in preferences for school attributes across household types against the context of the implementation of the RTE 25 percent mandate. Of the seven school attributes used in the models, four explain parental choices: home-school distance, fees, size and computers. Additionally, preferences are heterogeneous by household SES for fees, size and computers, and not for home-school distance.[32] Further, school composition is not a determinant of choice, and finally, there is no heterogeneity in preferences by the gender of the child.

On the direction of the effects, school size have a positive effect across SES terciles, while home-school distance has a negative effect. Fees has a small positive effect for the top two terciles and a very small negative effect for the bottom tercile. Finally, parents in the top terceile have a large positive preference for computers compared to those in the other two terciles whose preference for computers is negative. These results enable me to comment on the various theoretical and policy concerns surrounding the demand-side of school choice.

3.6.1 Fees is a marker for school quality

Parents' preferences for school fees are positive, statistically significant and heterogeneous in the main model (Table 3.8). The price elasticity is positive and elastic for the top two SES terciles, while being negative and inelastic

[32] There is heterogeneity in preferences for graduate teachers in Table 3.8, but it is not robust to inclusion of gender interactions in Table 3.10. Further the p-value of the F-test, even in Table 3.8 is 0.083, meaning that the significance is only at the 10 percent level.

for the bottom SES tercile (Table 3.9). Abstracting from the context these results are puzzling as the natural expectation is that of negative preference for fees, given the downward sloping demand curve. Carneiro et al. (2016) and Gallego and Hernando (2009) find fees to be a determinant of choice and demonstrate this negative relationship in Pakistan and Chile respectively. Both these studies, however, use data on actual schools attended by children. My study context varies critically in two aspects: First, I am using data on parental preferences elicited in the context of a full tuition waiver policy. Second, despite the tuition waiver, enrolling children in RTE free places involves considerable expenditure. As I demonstrate in Chapter 1 of this thesis, education at a private school involves both tuition and non- tuition expenses and the non- tuition expenses are approximately 1.3 times the tuition expenses. The 25 percent mandate only covers the tuition fee and parents are to bear the non-tuition expenses. So even in my context, parents should not have a preference for high-fee schools.

The preferences I observe can only be explained if parents view fees as a marker for school quality, either academic or otherwise.[33] This is completely plausible in the Indian private schooling context where the proliferation of low-fee private schools in the last decade has generated a general perception that fees is correlated with quality (Juneja, 2014).[34] That fees is a marker for quality, therefore, explains the positive average preference for fees and the heterogeneity: the tuition waiver is sufficient to induce higher-SES parents to apply for higher-fee schools but not lower SES ones.

[33] Alternately, it can be argued that the positive coefficient is due to omitted variable bias i.e., there is some key explanatory variable that parents observe and I do not. This is unlikely given that I have a rich data of school characteristics, some of which are not even observed by parents.

[34] Detailed argument in Chapter 2 of the thesis.

The heterogeneity in preference for fees, combined with the evidence from Chapter 2 that high-fee schools lead to significant value addition in English language, suggests that expanded choice under the RTE act is probably widening the childrens' learning gap in English learning across the SES divide. Hence my results seem to lend some support to choice-skeptics claims; however, more credible evidence based on schools' test scores data is required to establish if expanded choice could lead to widening of learning gaps as richer and poorer households choose schools differently.

3.6.2 No heterogeneity by distance and school composition

Unsurprisingly, home-school distance is a significant determinant of choice and parents across all three income terciles prefer proximity. Interestingly however, I find no statistically significant difference across household SES terciles in their preference for proximity. This heterogeneity result is contrary to the evidence from US and Chile (Hastings et al., 2009; Neilson, 2013; Gallego and Hernando, 2009) where proximity has been found to be an inferior good.[35] Further, school composition (proportion of children from disadvantaged backgrounds) is not a determinant of choice and there is no heterogeneity amongst household types along this dimension. This is also contrary to evidence from US, where Hastings et al. (2009) find that the average African American parent places a relatively higher weight on racial composition of the school than an average white parent.

The contextual differences between the operation of classic school choice models like vouchers in the US and the RTE 25 percent mandate in Bangalore

[35] A good whose demand increases with income is a normal good and that whose demand decreases with increasing income is an inferior good.

probably explain the lack of heterogeneity on home-school distance and school composition. Poor households in the US prioritize proximity as commutes to distant schools are costly and favor minority dominated schools, given the salience of race and race-based stratification of the schooling system. In contrast, Bangalore city has myriad affordable and reliable home-school commute options, both public and private (school buses). Further, choices in this context were to be made within neighborhoods, where the home-school distances are not particularly large: 90^{th} percentile of distance is around 5 kilometers, which is not an unusual distance for children to travel in any metropolitan city in India. This presumably explains the absence of heterogeneity in preferences for distance. Finally, despite the salience of caste in India, there is no caste based stratification of private schools comparable to the race-based stratification of public schools in the US.

3.6.3 Gender neutral preferences

The third takeaway from the results is that parental preferences are gender neutral. This seems puzzling given the widely reported phenomenon of discrimination against the girl child in educational investments in the Indian context (Zimmermann, 2012; White et al., 2016). However, studies that carefully explore the variation in this discrimination (by age-group and region) have not found a pro-male bias for young children. Kingdon (2005), using a nationally representative sample of rural Indian households, demonstrates the absence of a bias against younger girls (5-14 years) in intra-household resource allocation, conditional on school enrollment. More relevant is the latest evidence from Azam and Kingdon (2013) who find no statistically significant pro-male bias in household resource allocation in most Indian states for younger children (5-9 years).[36] My results are, therefore, consistent with the latest evidence on this issue.

3.6.4 Higher SES households have information advantage

The impact of school size on preferences is positive and statistically significant, on average, and for each of the three SES groups. The reason for this is straight forward: size is one of the most visible school characteristics, and in the absence of other information, parents might be using this as a proxy for school quality. The heterogeneity (lower-SES parents value size more than higher-SES ones) suggests that higher-SES parents probably have better (and additional) proxies for school quality than size.

On number of computers, the contrast in preferences by SES is most stark. While the top-SES parents seem to place a very high premium on the number of computers at school, lower-SES parents do not prefer schools with large number of computers. Computers, in my view, are a finer metric of school infrastructure (compared to coarser metrics like number of rooms); heterogeneity on this dimension suggests that higher-SES parents seem to have a better sense of the school facilities than lower-SES ones.

The heterogeneity on size and computers, therefore, suggests that higher-SES parents do have some informational advantage in assessing school quality and the finer aspects of infrastructure in schools.

[36] They find a statistically significant pro-male bias only for Chattisgarh and Madhya Pradesh, two central Indian states.

3.6.5 Limitations

Despite its many strengths, this analysis has a few limitations that I list below.

- *Absence of test scores data*: Given that standardized test scores data is not available in the Indian context, I am not able to comment on the central question in the demand-side debate of school choice: whether parents prioritize test scores while choosing schools?
- *Choice amongst private schools*: My policy setting constrains parents to choose amongst private schools in their neighborhoods. Hence I am unable to comment on parental preferences for attributes of public schools. Having both public and private schools in the choice set would have been ideal and enabled the estimation of preferences for all schooling options in the market.
- *Study Population*: My study sample is drawn from the population of RTE applicants. So the results are valid for this population. I argued in section 3.4.1 that the RTE applicant population is broadly representative of the wider population of households with 6-year old children (excluding the tails of the income/ SES distribution). However, in the absence of income/ SES distribution of the wider population, I cannot be certain that my study population is representative of this wider population.

3.7 Conclusion

I explored the determinants of private school choice under the RTE 25 percent mandate in Bangalore city, south India. I find that parents prioritize school fees, size, infrastructure and home-school distance while choosing schools. The positive average preference for school fee suggests that fee is a marker for quality in urban India; the heterogeneity across SES on this dimension indicates that the tuition-fee waiver under the RTE mandate is not sufficient to induce poorer households to choose higher-fee schools. The effects on school size, infrastructure and home-school distance are broadly in accordance with existing evidence from other contexts like the US, England and Chile. My results diverge from the extant literature on the heterogeneous preferences for school proximity across household types and on the importance of school composition. I do not find any significant differences in preferences for school proximity by household SES type, and school composition is not a determinant of choice. This divergence in results can be accounted for by the contextual specificity of the RTE 25 percent mandate and urban India.

References

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. and Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American economic review*, 93(3):729–747.

Alderman, H., Orazem, P. F., and Paterno, E. M. (2001). School quality, school cost, and the public/private school choices of low-income households in pakistan. *Journal of Human resources*, pages 304–326.

Azam, M. and Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, 42:143–164.

Beresteanu, A. and Zincenko, F. (2018). Efficiency Gains in Rank-ordered Multinomial Logit Models. *Oxford Bulletin of Economics and Statistics*, 80(1):122–134.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289.

Carneiro, P. M., Das, J., and Reis, H. (2016). The value of private schools: evidence from Pakistan. IZA Discussion Paper No. 9960.

Chapman, R. G. and Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, pages 288–301.

Checchi, D. and Jappelli, T. (2004). School choice and quality. Technical report, CEPR Discussion Paper No. 4748.

Cheng, S. and Long, J. S. (2007). Testing for IIA in the multinomial logit model. *Sociological methods & research*, 35(4):583–600.

Duncan, L. R. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Gallego, F. and Hernando, A. (2009). School choice in Chile: Looking at the demand side. Working Paper No. 356, 2010, Pontificia Universidad Catolica de Chile.

Glazerman, S. and Meyer, R. H. (1994). Public school choice in Minneapolis, midwest approaches to school reform. In *Proceedings of a Conference Held at the Federal Reserve Bank of Chicago*, pages 26–27.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management*, 1(1):19–41.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, 123(4):1373–1414.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, pages 1219–1240.

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):1477–1503.

Juneja, N. (2014). India's New Mandate against Economic Apartheid in Schools. *Journal of International Cooperation in Education*, 16:55–70.

Kingdon, G. G. (2005). Where has all the bias gone? Detecting gender bias in the intrahousehold allocation of educational expenditure. *Economic Development and Cultural Change*, 53(2):409–451.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers of Econometrics*.

McFadden, D., Tye, W. B., and Train, K. (1977). *An application of diagnostic tests* *for the independence from irrelevant alternatives property of the multinomial logit model*. Institute of Transportation Studies, University of California.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of econometrics*, 2:1395–1457.

McKenzie, D. (2018). When should you cluster standard errors? New wisdom from the econometrics oracle. Available at ://blogs.worldbank.org/impactevaluations/when-should-you-cluster-standard-errors-new-wisdom-econometrics-oracle.

Musset, P. (2012). School choice and equity: current policies in OECD countries and a literature review. Technical report, OECD.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Mimeo, Yale University.

Rothstein, J. M. (2006). Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350.

Shmueli, G. (2010). To explain or to predict? *Statistical science*, pages 289–310.

Small, K. A. and Hsiao, C. (1985). Multinomial logit specification tests. *International economic review*, pages 619–627.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

White, G., Ruther, M., and Kahn, J. (2016). *Educational Inequality in India: An Analysis of Gender Differences in Reading and Mathematics*. IHDS Working Paper 2016-2.

Zimmermann, L. (2012). Reconsidering gender bias in intrahousehold allocation in India. *Journal of Development Studies*, 48(1):151–163.

Appendix C: **Bangalore city map**

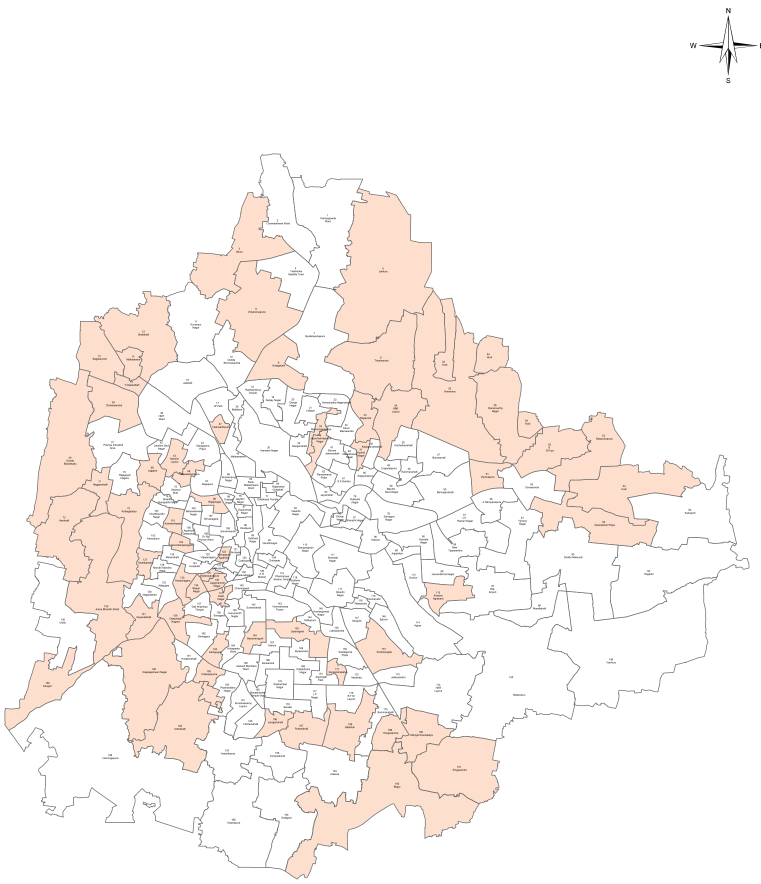


Figure A.3.1: Neighborhood map of Bangalore city

Appendix D: **Additional results**

Table A.3.1: Heterogeneity by household asset index

	(1)	(2)	(3)	(4)	(5)
		Coefficients		p values of F-tests	
	Ref group Top Tercile	Interaction: Middle Tercile	Interaction: Bottom Tercile		
Fees	0.104*	-0.043	-0.097**	0.117	0.095
	(0.053)	(0.076)	(0.045)		
Size	0.548***	-0.167*	-0.075	0.000	0.146
	(0.109)	(0.092)	(0.114)		
Rooms	-0.088	0.049	0.064	0.887	0.844
	(0.114)	(0.107)	(0.118)		
Computers	0.077	0.052	-0.182**	0.014	0.074
	(0.049)	(0.071)	(0.091)		
Graduate teachers	-0.036	0.044	0.001	0.795	0.627
	(0.062)	(0.057)	(0.068)		
Proportion of SC/ST	-0.116	0.039	0.004	0.661	0.887
	(0.106)	(0.102)	(0.093)		
Distance	-0.238**	-0.158	-0.187	0.024	0.476
	(0.100)	(0.157)	(0.162)		
N= 16,773; Number of children= 606					

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parenthesis and are clustered at the neighborhood level to account for the sampling design. The model also includes dummies for school type (based on the highest grade at school) in additional to the school attributes and interactions of school attributes and the asset tercile dummies. Columns (1)-(3) present the regression results of a single regression. Column (4) and (5) present p-values of the F-test that check for joint significance of coefficients in columns (1)-(3) and (2)-(3) respectively. A p-value less than 0.1 in column (4) indicates that the school attribute is significant, and a p-value less than 0.1 in column (5) indicates heterogeneity across asset index groups on that school attribute. The values of the coefficients for the top tercile are those in columns (1); for the other two terciles, the coefficient in the respective column should be added to that in the reference group.

Table A.3.2: Heterogeneity by asset index and gender

	(1)	(2)	(3)	(4)	(5)	(6)
		Coefficients			p values of F-tests	
	Reference category	Interaction: SES (Middle Tercile)	Interaction: SES (Bottom Tercile)	Gender (Girl)		
Fees	0.106*	-0.041	-0.099**	-0.005	0.224	0.096
	(0.056)	(0.079)	(0.046)	(0.039)		
Size	0.538***	-0.169*	-0.085	0.046	0.000	0.153
	(0.108)	(0.092)	(0.115)	(0.074)		
Rooms	-0.117	0.044	0.054	0.077	0.818	0.887
	(0.126)	(0.108)	(0.120)	(0.068)		
Computers	0.096	0.048	-0.191**	-0.032	0.022	0.048
	(0.076)	(0.074)	(0.086)	(0.061)		
Graduate teachers	0.003	0.057	0.021	-0.088*	0.317	0.575
	(0.052)	(0.059)	(0.064)	(0.050)		
Proportion of SC/ST	-0.117	0.043	0.043	0.002	0.790	0.886
	(0.119)	(0.102)	(0.095)	(0.081)		
Distance	-0.279**	-0.175	-0.182	0.090	0.042	0.476
	(0.137)	(0.166)	(0.171)	(0.155)		
N= 16,773; Number of children= 606						

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parenthesis and are clustered at the neighborhood level to account for the sampling design. The model also includes dummies for school type (based on the highest grade at school) in additional to the school attributes, interactions of school attributes and the asset tercile dummies and interactions of school attributes with the two gender dummies. Columns (1)-(4) present the regression results of a single regression. Column (5) and (6) present p-values of the F-test that check for joint significance of coefficients in columns (1)-(4) and (2)-(3) respectively. A p-value less than 0.1 in column (5) indicates that the school attribute is significant, and a p-value less than 0.1 in column (6) indicates heterogeneity across asset index groups on that school attribute. The results in column (4) indicate heterogeneity along the gender dimension.

Conclusions and Policy Implications

In this thesis I use the context of India's national school choice policy, the 25 percent mandate, in the south Indian state of Karnataka and investigate the following three sets of questions, each of which is presented in a substantive chapter.

1. What is the impact of the 25 percent mandate on childrens' learning and psychosocial outcomes? Has school choice led to improved student outcomes as predicted by theory?

2. What is the value added at private and elite (private) schools?[1] Are private schools in general, and elite private schools in particular, more effective at improving student outcomes as contended by choice enthusiasts?

3. What are the determinants of parental school choices? Do preferences for schools vary by household and child characteristics?

Summary of evidence

In Chapter 1, I exploit the lottery based allocation of free places under the 25 percent mandate and evaluate its impact on a sample of 1,616 children drawn from 4 districts of Karnataka. I find no difference between the test

[1] There is no agreed upon definition of elite schools in the literature. I define schools with fee higher than the 75th percentile (p75) of the fee distribution as high-fee/elite.

scores of lottery winning children relative to the losing ones on a range of subjects. I find statistically significant difference on one of the four measured psychosocial outcomes: self- efficacy. I conduct a detailed mechanism analysis to understand the non-impact. The evidence establishes that both the lottery winners and losers attend similar types of private schools, and that in a counterfactual world without the policy, the average policy applicant would have attended a private school charging the median tuition fee. This implies that the policy applicants are not income-constrained and default government school going children as envisaged in the policy's theory of change. Majority of the policy applicant are from non-poor households that do not need government support to attend a private school. The 25 percent mandate has been mistargeted with ineligible households applying in large numbers and eligible households not participating. I argue that design flaws in the mandate are responsible for mistargeting. I replicate the mistargeting result on a second and bigger sample drawn from across Karnataka state.

In Chapter 2, I use the same experimental set up and data from Chapter 1 to estimate the value addition at private schools vis-à-vis public ones and at elite/high-fee private schools vis-à-vis low-fee ones. I exploit the RTE lottery and use Two Stage Least Squares (2SLS)/ Instrumental Variables (IV) approach to estimate private and elite school effects on childrens' learning and psycosocial outcomes. I find that enrollment at a private school has no statistically significant effect on any of the five test score measures. However, the coefficients are in the range of 0.2-0.6 σ and are not estimated precisely- my sample is only powered to detect very large effects (about 1.3 σ and above). Hence I am not able to comment on the valued added at private schools in learning. On psychosocial outcomes, I find that the effect on self-efficacy, one of the four measured psychosocial outcomes, is large, statistically significant and robust (effect size: 1.90 σ). Moving on to the value addition at elite schools, I find significant, robust and economically meaningful effects only on English test scores (effect size: 0.58 σ). I argue that the English effects are unsurprising given the labor market premiums to English proficiency. However, the absence of overall learning gains is puzzling.

In Chapter 3, I present evidence on the demand-side of the RTE 25 percent mandate by evaluating the determinants of parental private school choices and exploring if parental preferences vary by household type and gender of the child. I deploy a discrete choice modelling approach and use rich rank-ordered data of parental preferences to estimate a Rank- Ordered Logit (ROL) model. I find that home-school distance, school-fees, size and infrastructure are the primary determinants of choice. The positive average preference for school fee suggests that fee is a marker for quality in urban India; the heterogeneity across SES on this dimension indicates that the tuition-fee waiver under the RTE mandate is not sufficient to induce poorer households to choose higher-fee schools. The effects on school size, infrastructure and home-school distance are broadly in accordance with existing evidence from other contexts like the US, England and Chile. My results diverge from the extant literature on the heterogeneous preferences for school proximity across household types and on the importance of school composition. I do not find any significant differences in preferences for school proximity by household SES type, and school composition is not a determinant of choice. This divergence in results can be accounted for by the contextual specificity of the RTE 25 percent mandate and urban India.

The evidence presented in the three chapters has sizeable implications for the 25 percent mandate and school choice programs in urban India. These implications would also be relevant beyond the Indian context to several developing countries, which have similar primary schooling markets.

Policy implications- 25 percent mandate

This research is the very first empirical investigation of the RTE 25 percent mandate, the largest school choice policy in the world.[2] The evidence presented in this thesis has several implications directly relevant to the mandate's design and implementation in Karnataka and in other Indian states.

Improving the mandate's targeting

The central result of non-impact of the policy owing to failure of targeting implies that the mandate, in its current form, has not accomplished the stated goals. It has led to a potentially welfare reducing income transfer from taxpayers to non-poor households. Government, therefore, has to fix the targeting of the mandate if it has to work as envisaged and improve learning outcomes of children. I identified that mistargeting occurred due to two reasons. First, income-based eligibility determination, in a context where income determination is infeasible, has allowed ineligible non-poor households to apply in large numbers to the mandate. A second reason is the partial nature of the RTE subsidy. My results show that the non-tuition expenses of private school education are 1.3 times that of the tuition expenses, on average. These high non-tuition costs probably deterred poor

[2] When fully implemented the mandate would impact 16 million children.

households from applying to the mandate. Both the income-based targeting of the program and the partial nature of the RTE subsidy are design elements of the policy, and unless they are addressed it is unlikely that targeting and hence the effectiveness of the policy will improve.

Can income based targeting be improved through administrative effort? In Karnataka and in other Indian states that use income-based targeting, grassroots administrative functionaries issue income certificates based on a declaration by the parent. Though corruption and incompetence at the grassroots could be partly responsible for false income certificates, there is no systematic administrative fix for this problem. Even a well meaning and competent administrative functionary cannot assess a household's income in the absence of supporting administrative records like tax documents or pay stubs. Income based targeting is, therefore, unworkable in a context where less than five percent of the population has tax records and more than 90 percent of the labor force work is in the informal sector. Hence administrative effort alone cannot significantly improve the situation.

This implies that the income-based targeting should be replaced with an alternate model. Madhya Pradesh, another front runner state in the implementation of the mandate, uses the existing poverty census to target the program.[3] However, using the poverty census is not without challenges (Drèze and Khera, 2010; Alkire and Seth, 2008). Further, there are legal concerns to shifting from the current income-based targeting to poverty census based approach, as experienced by Rajasthan, another Indian state

[3] India has a system of conducting regular poverty census, where households are scored on different criteria that proxy income poverty to generate a numerical poverty score. Households below a particular cutoff on the scale are classified as Below Poverty Line (BPL) and are eligible for social programs. A significant number of benefits, including food subsidies, rural housing programs and self-employment programs are currently targeted based on this approach.

(Rajasthan High Court, 2017, p. 32). Hence choosing an appropriate targeting strategy is a complex issue with administrative, legal and political dimensions. Pointing to a concrete solution is beyond the scope of this thesis. The recommendation that emanates from my evidence and analysis is that government should devote its energies toward ensuring right targeting of the mandate. Unless this is fixed, the mandate will not achieve its goals as envisaged in the theory of change.

Can the size of the RTE subsidy be increased to cover non-tuition expenses? I demonstrate that, on average, the non-tuition costs of education are 1.3 times the tuition costs and argue that these high costs could be deterring the truly income constrained households from applying to the mandate. Demands for increasing the RTE subsidy to cover the non-tuition costs have been made repeatedly by activists working on the mandate in Karnataka. Realizing the importance of non-tuition costs, Delhi, another keen implementer of the mandate, subsidizes a large part of the non-tuition expenses as well. Following this route will be costly for the government, but my evidence and anecdotal claims by activists suggest that it could increase the participation of the poor in the program.

Cost-effectiveness of the mandate

Given the evidence that the per child expenditure in private schools is almost a third of that in the public sector and that private schools are more productive than government ones [4] (Muralidharan and Sundararaman, 2015), some commentators and policy makers argue that the 25 percent mandate is a cost-effective way of providing education to children from

[4] Private schools produce similar levels of test score gains at at a much lower cost.

disadvantaged backgrounds.

The data on cost-effectiveness in Karnataka shows that the average government spending on the mandate was INR 6,800 per child for the year 2015-16. The average cost of schooling in the government system is INR 11,848.[5] Juxtaposing this cost data with the ITT policy impacts on learning from Chapter 1 shows that government has been able to procure similar quality education from private schools by paying only 60 percent of the costs currently incurred. It is therefore contended that 25 percent mandate has been successful in leveraging the higher productivity of the private sector for procuring education to disadvantaged children. However, this naive cost-effectiveness calculation only accounts for spending by the government and not that of the high-fee private schools, which are co-subsidizing the education at free places.[6] Including the spending by private schools, and hence overall costs to society, will reduce the cost-effectiveness advantage of the 25 percent mandate. So the 25 percent mandate, in its current from, is not as cost-effective as a simple comparison between government spending on the mandate and average per child expenditure in public schools would suggest. However, an alternate school choice policy that reserves free places only at low and median-cost private schools or even the classic school

[5] The calculation of this amount only uses the recurring costs of running schools (primarily teacher salaries). Other substantial costs incurred by the government like costs of infrastructure, administration and school inputs (text books, free meals) are not included in this calculation. Including all costs rises the per child cost incurred by government to INR 25,500 (Karnataka education budget of INR 165,000 million divided by 6.46 million, the total number of children in government schools from grade 1-10).

[6] Half of the private schools charge fee higher than what the government pays for free places. These schools are losing revenue through participation in the mandate and are, hence, co-subsidizing the education of free places children.

voucher system could be cost-effective.[7]

Policy implications- school choice

I cannot estimate the school choice effects using the data from the implementation of the mandate as it did not lead to expanded school choice for the average applicant. However, the results on the private school effect and elite school effect in Chapter 2 allow me to conjecture on what the school choice effects would have been had the RTE mandate worked as originally conceived. If the mandate led primarily to shifts from government to private sector, I would have observed large self-efficacy effects. Alternately, if the mandate led to shifts from low-fee to high-fee schools, I would have observed large English language gains. In either case, it appears that school choice would have some positive effects on human capital formation in the context of urban India without being the silver bullet for improving educational outcomes of children from disadvantaged backgrounds.

In summary, my results demonstrate that the 25 percent mandate (if correctly targeted) in particular, and school choice in general, has the potential to provide some learning gains to poor disadvantaged students. It cannot, however, promote significant improvements in learning. Hence reliance on private schools cannot be the default option for improving

[7] A fundamental goal of the 25 percent mandate is to create inclusive learning environments, in addition to improving test scores of disadvantaged children. The cost-effectiveness calculus for the 25 percent mandate will, hence, depend on the value society places on creating inclusive learning environments relative to improving test scores. For instance, spending money to send poor children to elite private schools, that do not produce any significant learning gains, could still be welfare enhancing if society places high value on creating inclusive learning spaces.

education outcomes. Reforms to strengthen public schools must be put in place alongside market based reforms like choice. This conclusion resonates with the Drèze and Sen (2013, p.138) who contend:

> Private schools are likely to remain a part of the picture of Indian school education, but the need for carrying out major reforms of state schools as well as for relying on them is not going to disappear merely because of the existence of the private option.

Strengths and Limitations

The methodologies and the data used in this thesis have several strengths; a few of the key ones are listed below:

- The evidence in Chapters 1 and 2 comes from a cleanly identified randomized experiment and hence is causal. The study is powered to detect even small policy (ITT) impacts of up to 0.1 standard deviations.
- The central result of Chapter 1 that the policy is mistargted is replicated on a larger sample of about 16,000 children drawn from across Karnataka state and from another year of the policy implementation, thus increasing its temporal and spatial validity.
- The childrens' outcome measures deployed here are more comprehensive than those used in existing research and hence provide a more complete picture of the impact of 25 percent mandate and school choice on outcomes.
- In Chapter 3, I have rank-ordered data for each parent. This generates a stronger source of variation for identification of model parameters compared to just having information on the first choice, which is

typically the case.[8] Further, the place allocation algorithm used in Bangalore is based on a Random Serial Dictatorship (RSD) mechanism, wherein the dominant strategy for parents is to reveal their true preferences. Hence the school choices observed in my dataset reflect true parental preferences. Most previous work used data generated in response to non-strategy proof algorithms like the Boston mechanism, thus, raising concerns around identification of true parental preferences.

Some of the limitations are below:

- The dataset used in the analysis does not allow estimation of school choice effects as the average policy applicant has not been exposed to expanded choice.
- The compliance rate[9] of the study is very low for the estimation of private school effects. Hence the IV estimates for the private school effect on test scores have large standard errors are not estimated precisely. I am,therefore, not able to comment on the private school premium in learning.
- I do not have test scores data in estimating the determinants of choice in Chapter 3. Hence I cannot comment on a fundamental question in the demand-side debate of school choice: the relative weight that parents' place on test scores compared to non-academic aspects of schooling.
- My study sample in Chapter 3 is drawn from the population of RTE applicants. So the results are valid only for this population. The

[8] The only other study that used rank-ordered data in investigating the demand side of school choice is Hastings et al. (2009) study on the efficacy of a public school choice policy in Charlotte- Mecklenburg in North Carolina.

[9] Compliance is defined as attending a private school conditional on winning the RTE lottery.

evidence that this population is broadly similar to the wider population of households with 6- year old children is only suggestive.

Areas for future research

I now present potential areas for future research prompted by evidence in this thesis.

The 25 percent mandate

I demonstrate that the 25 percent mandate is mistargeted in Karnataka because of the income-based targeting and the partial nature of the RTE subsidy. Other Indian states are implementing the mandate using alternate targeting strategies (Madhya Pradesh) and are subsidizing the non-tuition costs of private school education for poor children (Delhi). Evaluating the impact of the policy and its targeting performance in these contexts could deepen our understanding of both the design and implementation of this crucial school choice mandate. Second, my conclusion that the current spending on the mandate is a transfer from taxpayers to RTE free place winners is robust, but I could not estimate the welfare impacts of the policy owing to data constraints. Estimating the within household spillover effects of the RTE transfer and the welfare implications from a societal perspective are important areas for future research.

Value added at schools

On the impacts of private schools, I report large and robust self-efficacy effects. But given that they are based on a small sub-sample of government school children it is necessary to validate these results in an experiment with higher compliance. Additionally, some qualitative work is required to understand the source of the large self-efficacy gains that I report. Second, my elite school effects demonstrate the significant gains in English from attendance at an elite school, but the source of this effect and reasons for the absence of overall effects is unclear. This points to the need for studying the role of information asymmetries and parental preferences for non-academic aspects in choosing schools. Evidence in this area is completely absent in the context of developing countries.

Demand side of school choice

I estimate a classic discrete choice model that does not allow school choices to depend on unobserved school effect common to all students. This could potentially lead to biased estimates. Improving upon my estimation model using the BLP specification (Berry et al., 1995), allowing for school fixed effects, is a probable direction to extend my work.

References

Alkire, S. and Seth, S. (2008). Determining BPL Status: Some Methodological Improvements. *Indian Journal of Human Development*, 2(2):407–424.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Drèze, J. and Khera, R. (2010). The BPL census and a possible alternative. *Economic and Political Weekly*, pages 54–63.

Drèze, J. and Sen, A. (2013). *An uncertain glory : India and its contradictions*. London : Allen Lane.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46.

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.

Rajasthan High Court (2017). Annual Report 2016, Rajasthan High Court. Government report, Rajasthan High Court.