

This is an Accepted Manuscript of an article published by Taylor&Francis in [Oxford Review of Education](#) on 13/01/2014, available online:
<http://www.tandfonline.com/doi/pdf/10.1080/03054985.2013.873529>

Published details:

Singh, Abhijeet (2014) "Test score gaps between private and government sector students at school entry age in India", *Oxford Review of Education* 40 (1): 30-49. DOI: 10.1080/03054985.2013.873529

This article is reproduced in accordance with the self-archiving policies of Taylor&Francis.

Test score gaps between private and government sector students at school entry age in India

Abhijeet Singh
University of Oxford¹

October 2013

Abstract

Various studies have noted that students enrolled in private schools in India perform better on average than students in government schools. In this paper, I show that large gaps in the test scores of children in private and public sector education are evident even at the point of initial enrolment in formal schooling and are associated with previous attendance in private and public preschools. Children in the sample were aged between 4.5 and 6 years at the time of the survey and were tested on receptive vocabulary and quantitative skills. Most (albeit not all) children in the sample had attended private and/or government preschools and at the time of the survey, about 44% had made a transition into formal schooling. Attending private pre-schools is associated with substantially, and significantly, higher test scores relative to attending public preschools. A considerable portion of this difference can be removed by controls for parental background and selected child characteristics but the gap remains significant. Possible implications of these results are discussed.

¹ *I am particularly grateful to Francis Teal for helpful discussion and guidance over several years. I am also grateful for detailed and useful feedback to the Editors of the special issue, Angela Little and Caine Rolleston, and two anonymous referees for detailed comments which helped improve this manuscript greatly. Helpful comments were also received at early stages from Stefan Dercon, Caroline Knowles, Geeta Kingdon, Pieter Serneels and Martin Woodhead. Young Lives is core-funded by UK aid from the Department for International Development (DFID) and co-funded from 2010 to 2014 by the Netherlands Ministry of Foreign Affairs. www.younglives.org.uk*

I. Introduction

The private sector in Indian education has been growing rapidly in the last two decades (Kingdon 2007) and it is now well-known that there are significant gaps between the average learning scores of children in private and public schools (e.g. Muralidharan and Kremer 2008).

Studies that evaluate gaps in learning between students in the private and state sectors in India thus far have focused overwhelmingly on private and public schools without any reference to prior preschool education. In this, they have missed the first two years of the child's education. Given the widespread recognition of the importance of early childhood factors on future cognitive outcomes, this omission is a major limitation of the literature as it stands today. A major reason behind this constraint in our knowledge has been the lack of suitable data which links the type of preschool/school attended to various household factors. As Kingdon (2008) discusses in depth, reliable data on schooling in India is scarce and this scarcity has severely hampered econometric research on education in India. While this situation has improved markedly for school-based data in the past decade, most notably through the annual ASER Surveys conducted by Pratham as well as various in-depth and detailed quantitative surveys², the possible importance of preschools in determining inequalities in test scores remains understudied.

An understanding of the skills that children develop even before they first enter school is important. Firstly, in the context of test score gaps between different groups of children, understanding when gaps emerge, whether these gaps subsequently persist, shrink or are magnified by experiences in later childhood, is central to formulating policy responses to these gaps. In the context of the US, for example, understanding when gaps emerge and what proportion of the variation in test scores is explained by home factors and pre-school factors has been a central theme in the literature on racial test score gaps (e.g. Fryer and Levitt 2004, 2006 and Todd and Wolpin 2007) and gender gaps in mathematics achievement (e.g. Fryer and Levitt, 2010). Secondly, as a vast body of work in education, psychology and economics documents, early childhood investments play an important role in determining later life outcomes and preschool investments in particular may have large persistent impacts: see, for instance, the influential work arising from the Perry Preschool Project in the US (Schweinhart 1993, Schweinhart et. al. 2005), the Effective Provision of Preschool Education project in the UK (Sylva et. al. 2004 and papers in Sylva et. al. (Eds., 2010)) and recent work by economists Flavio Cunha, James Heckman and co-authors (Cunha and Heckman 2008, Cunha, Heckman and Schennach 2010) using longitudinal data from the US. For the most part, however, similar evidence has not been available for developing countries³.

² For example, the SchoolTELLS survey carried out by Geeta Kingdon and coauthors and the repeated surveys carried out by the Andhra Pradesh Randomized Evaluation Studies have been very useful in facilitating detailed analyses of factors affecting the rate of learning in different schools.

³ See, however, the influential Lancet Studies on the importance of preschool factors especially Engle et. al. (2011) who pay particular attention to the potential effects of preschool education to the development potential of children and Walker et. al. (2011) who discuss the risk factors and protective factors affecting early childhood development in developing countries.

In this paper, I attempt to partially address the gaps in our knowledge that I have identified above. Specifically, using data collected in 2006/7 by the Young Lives Project in the state of Andhra Pradesh, I focus on cognitive outcomes for children between 4 ½ and 6 years; 5 years is the modal age of enrolment into formal schooling in India and this is exactly the age window in which most children are transitioning to formal schools. I investigate descriptively the extent of gaps in receptive vocabulary and quantitative skills in this sample between children who have been enrolled in different combinations of public and private preschools and schools until the time of the survey. Furthermore, I analyze the correlates of test score gaps between different educational paths and demonstrate that whereas much of the gap is explained by controlling for differences in parental and socio-economic characteristics including household expenditures, a significant gap persists between children who attended private preschools or schools and other children, most of whom are enrolled in state-run institutions. The objective in this paper is not to focus on why students in the private sector outperform those in the government sector but on when; causal attribution of the sources of these gaps is beyond the scope of this paper. I discuss possible implications of the gaps in cognitive skills I describe, which are large even at this early age, in the final section of this paper.

In the next section I describe the nature of the schooling available to young children in India. Section 3 sets out the data which is analysed in section 4. A final section concludes.

2. Schooling for Young Children in India

The preschool system in India, as elsewhere, is important as a feeder system for the formal school system into which children later graduate. However, the quality and take-up of different types of systems is highly variable and it may reasonably be expected that the preschool trajectories of children affect their school-readiness and will contribute to future learning. In the Indian context, the implications of this choice may be even starker because of important institutional differences between different types of preschools as well as a large variation in the age at which children first enrol in both preschools and schools.

Public pre-schools refer mostly to *anganwadis/balwadis* which are early childcare centres supported by the government which administer a package of six services provided under the Integrated Child Development Services (ICDS) Scheme: supplementary nutrition, non-formal pre-school education, immunization, health check-up, referral services and nutrition and health education. The quality of services provided by *anganwadis*, especially with regard to preschool education which lags behind immunization and supplementary feeding as a focus for ICDS implementation in many states, is frequently seen as poor. For example the FOCUS Report (Citizens' Initiative for the Rights of Children Under Six, 2006), which surveyed 203 *anganwadis* across six states noted that "preschool education activities were quite limited in most sample *anganwadis* and virtually non-existent in large parts of Chattisgarh and Rajasthan" and that "where early childhood education is provided at the *anganwadi*, such activity is sporadic and limited". This is despite the fact that preschool education is reported by parents as the most important reason that they want their children to go to *anganwadis*.

Private pre-schools are fee-paying institutions in general, consisting of kindergarten classes, and have a narrower focus on learning than *anganwadis*; they are also more likely to impart English language instruction, a major draw for parents from all backgrounds⁴. Like private primary schools, private preschools are highly variable in quality and outreach. The diversity in the preschool sector in India,

⁴ In many cases, private preschools refer simply to the kindergarten sections of the private schools and therefore the close correspondence in characteristics between these institutions is not surprising.

and indeed the primary education sector, has been documented in detail by Woodhead et al (2009) who use qualitative data for selected children in the Young Lives study to explore the implications of entry into these institutions.

Primary schools are the official entry point into formal education although children in preschools, especially private kindergarten pre-schools, already receive instruction on numeracy and the alphabet. While public schools do not charge tuition fees, private schools are mostly fee-paying; recent estimates suggest that about 30% of children of primary school age in rural areas are enrolled in private schools (Pratham, 2013) with corresponding figures much higher in urban areas. The differences between public and private schools have been widely studied and it has been found that private schools are associated with student achievement that are as high or higher even after accounting for all pre-existing differences in socio-economic background. This conclusion is borne out by a now large literature which employs data from different states in India and from urban and rural areas and a range of econometric strategies including controlling cross-sectionally for observed characteristics of children and their households (Muralidharan and Kremer 2008, Kingdon 1996, French and Kingdon 2010, Desai et. al. 2009), controlling for village and household fixed effects to isolate the effects of unobserved differences between children in public and private schools (French and Kingdon, 2010), propensity score estimates using cross-sectional data (Chudgar and Quin, 2012), value-added estimates using panel data on individual children (Singh 2013) and experimental estimates using variation induced through lottery-based school vouchers (Muralidharan and Sundaraman 2013). Most studies in the literature find that students in private schools significantly outperform students in government schools (at least along some dimensions); coupled with the fact that per-pupil expenditures in private schools are frequently much lower than in the state sector, this suggests that private schools are considerably more productive on average than state schools⁵. These studies have also documented that private schools pay teachers substantially less than state schools and that teachers in these schools are more likely to be women, younger, less experienced, less likely to hold a teaching qualification, less likely to absent from school and more likely to be teaching while in school compared to their state school counterparts. While these insights have been greatly boosted by the host of studies in the past decade, much of the stylized picture of comparison between state and private schools was already apparent at the time of the influential PROBE report which presented survey results from the mid-nineties (PROBE Team, 1999).

A related but distinct literature has also emerged in the past decade evaluating various interventions for the low productivity of Indian schools, for example through incentives paid to teachers for greater attendance (Duflo et. al. 2012), computer-assisted learning (Banerjee et. al. 2003), para-teachers (Atherton and Kingdon 2010, Muralidharan and Sundararaman 2012, Banerjee et. al. 2003), performance-based incentives to teachers (Muralidharan and Sundararaman 2011), diagnostic feedback to teachers (Muralidharan and Sundararaman 2010) and block grants to schools (Das et. al. 2013).

In contrast to the vibrant literature that has grown around schooling in India - in describing the gaps in the relative effectiveness of private and state schools, in investigating the sources of this gap, and in evaluating possible interventions for remedying the low productivity of schools in general - much less is known about private and state preschools especially with regard to cognitive outcomes even at

⁵ This is also a point that has been stressed by James Tooley and coauthors in a range of different work focused on urban, peri-urban and rural areas in different parts of the country (See, e.g., Tooley and Dixon 2006 and Tooley, Dixon and Gomathi 2007) .

a descriptive level. As I noted previously, much of this neglect is caused by data paucity: school data collection exercises naturally do not observe children before they have entered school and are frequently hampered by their limited access to information from the household; at the same time large data collection exercises at the household level, even when they do collect any test data from children, only focus on children of school-going age⁶.

It is in the context of this literature that I attempt to extend the analysis further by accounting for the years in preschool that many of the children spent before they were enrolled in schools. I believe this is important: in the absence of any information on previous preschool attendance, it is *a priori* possible that the differences in average scores that these studies measured at a later date are caused entirely by the different preschool attendance of the children causing divergences that persist. Also, having ignored the evolution of cognitive skills prior to school entrance, the extant literature is unable to comment on whether investing in effective preschool education programs might be a possible mechanism for the remediation of social disadvantage in test scores⁷.

3. The data

The data I use in this study were collected by the Young Lives Project in 2002 and 2006/7 in the State of Andhra Pradesh. The surveys cover two cohorts: the first of 2011 children born between January 2001 and June 2002, and the second of 1008 children born between January 1994 and June 1995. In the second round (2006/7), 1950 children of the younger cohort and 994 children of the older cohort could be traced and resurveyed highlighting the exceptionally low attrition rates in the dataset. In this paper, I will be using data from the younger cohort from Round 2 of the survey. Of the 1950 children re-surveyed in 2006/7, 1941 had complete background information required for the regression analysis presented later in the paper; 9 children for whom the full set of controls was unavailable have been excluded from all analysis in the paper.

The major reason that I confine our attention to children in the younger cohort is that they are at a critical age with regard to school choice. Aged between 4 ½ and 6 years of age in 2006, by Round 2 about 44% of the cohort is enrolled in primary schooling, with the rest likely to soon follow suit: 5 years is the modal age of enrolment in Andhra Pradesh⁸. Thus our data were collected at the very time that questions about school choice were being decided for this cohort; this is a major advantage of the data. By concentrating exclusively on this cohort I can well document if divergences have arisen even at this early stage.

The survey administered detailed household questionnaires in both rounds. In round 2, the household questionnaire collected extensive information on a range of household characteristics including their consumption expenditures, social backgrounds, parental education and household composition. Additionally, the survey asked several questions about schooling of the children including, in the younger cohort, details on their preschool enrolment and attendance, whether they have begun

⁶ Notably the India Human Development Survey 2005 only collected test data for children between 8-11 years of age

⁷ In Argentina, for example, Berlinski, Galiani and Gertler (2009) report that a one year increase in pre-primary school increases average third grade test scores by 23% of a standard deviation. If similar relationships exist in India, then inequalities in preschool access likely contribute to inequalities in educational attainment observed later in the educational trajectory.

⁸ In the older cohort, born in 1994/5, about 70% of children started schooling at the age of 5 years; 98% of the children were observed to be enrolled at the age of 8 years in 2002.

primary schooling, the age at which they started or are expected to start primary schooling and the type of school (whether public or private) that they are enrolled in or are likely to be enrolled in if not yet enrolled. Since the survey only covers Andhra Pradesh, it may be erroneous to generalize findings to other states. Nonetheless, the findings will be suggestive at least for what we may expect to find in other states and can certainly provide evidence applicable to Andhra Pradesh which in itself is one of the larger Indian states⁹.

Children in this sample were given two learning tests which are used as outcome variables in our analysis. The first test, the Peabody Picture Vocabulary Test (PPVT), is a widely-used test of receptive vocabulary which attempts to measure vocabulary acquisition in persons from 2.5 years old to adulthood. The test is orally administered individually and is un-timed. The task of the test taker is to select the picture that best represents the meaning of a stimulus word presented orally by the examiner (Dunn and Dunn 1997, Cueto et al. 2009). The test, originally designed for testing individuals in English, was adapted into local languages; in our analysis I focus exclusively on children who were administered the test in Telugu, which accounts for about 95% of the sample¹⁰. The second test is the Cognitive Development Assessment (CDA). The CDA was developed by the International Evaluation Association (IEA) during the second phase of the Pre-Primary Project in order to assess the effect of attending a pre-school center in the cognitive development of 4 year old children (Cueto et. al. 2009). The CDA consists of different sub-sections measuring children's grasp of spatial relations, quantities and time. Of these, the Young Lives study administered only the quantitative sub-section, which is what I use in this paper¹¹. Test scores were generated for this paper using the 3-PL Item Response Theory model which posits a mathematical relationship between latent ability and observed responses to test scores; these models are routinely employed in educational assessment, for example by international testing programs like PISA, TIMSS, GRE and SAT, but are less common in development economics¹². Scores were normalized to have a mean of zero and a standard deviation of 1 in the sample.

Despite being close in age, there is a wide dispersion in their current enrolment status in educational institutions and in their test scores. At the time of the survey, the children in the sample were observed dispersed into the following categories by current enrolment: Public school (n=689), private School (n=160), public preschool (n=413), private preschool (n=549) and not enrolled (n=130). The children currently enrolled in primary schools, whether private or public, can further be categorized by their prior preschool experience. The summary statistics for the younger cohort for their achievements on

⁹ Andhra Pradesh had a population of about 84 million in 2011, larger than the population of Germany, and an area of 276,754 sq. km., which is larger than the United Kingdom. The data are relatively clustered and interview children in 100 communities (villages or urban wards) in 20 sub-districts (mandals); Kumra (2009) compares the Young Lives sample to representative household data for Andhra Pradesh and finds that the sample does contain similar variation that found in representative surveys.

¹⁰ This restricts the sample used for the analysis of PPVT scores to 1829 children. The exclusion of children who took the test in other languages is necessitated by possible differences in the complexity of the same words across languages (thus leading to Differential Item Functioning) and has been recommended by the original adaptors of the test in Telugu for Young Lives (Cueto et. al. 2009).

¹¹ This subscale requires children to indicate which one of a set of pictures fits the description provided by the examiner. Notions such as *a few*, *most*, *half*, *many*, *equal*, *a pair*, etc. are assessed with statements such as: '*Point to the plate that has a few cupcakes*'. See the detailed Young Lives technical note by Cueto et. al(2009) for further elaboration on the content and characteristics of the two achievement tests.

¹² For a more detailed explanation of Item Response Theory (IRT) models please see Das and Zajonc (2010) and Van der Linden and Hambleton (1997). It is worthwhile to note that key results of the analysis are not driven by the choice of score aggregation method; results are similar when using raw scores instead of IRT scores.

both the PPVT and the CDA test are set out in Table 1 by the path they have followed to date through the educational system. I identify nine possible combinations of public and private preschools and schools in our data beginning with those who have, at the time of the survey, attended neither preschool nor school. Reflecting the age of the children, more than half the sample has only attended preschool thus far. A similar number of children are enrolled in private preschools as in public preschool at the time of the survey; the data suggest that the age at which children transition from preschools to schools in the public sector is lower than in the private sector.

As is evident from Table 1, there is a wide gap in the test scores of children across different trajectories; the median score in public preschool is about 0.5 standard deviations (SD) lower in CDA and about 0.75 SD lower in PPVT than the median score in private preschool students. This is a large difference similar in magnitude to the raw gap measured between students in private and government schools at later ages (Singh 2013). In Figures 1 and 2, I present the distribution of test scores in the PPVT and CDA in our sample for children by their enrolment status at the time of the survey. There is a clear pattern evident in both figures. In both tests, children who are in private institutions score significantly more than children in public educational institutions and children in government schools display a distribution of test scores very similar to children who are not enrolled in any institution. The descriptive statistics provided above do not provide any evidence regarding the source of the gaps. It is conceivable that the entire gap is a product of home environments and differing socio-economic backgrounds of the children and unrelated to the quality of preschool or school education that they receive. However, it is apparent from these descriptive measures that the gap noted in later childhood descriptively between children in private and public schools is not entirely a product of their school history but emerges earlier in childhood.

4. Empirical Specifications and Results

In this Section I investigate the relation of the observed test score gaps between children in different educational institutions and their background characteristics.

Before proceeding with the econometric analysis, it is useful to look descriptively at the difference in the socio-economic background of students in different trajectories. Table 2 presents the mean and standard deviation of a vector of background characteristics for each trajectory identified in the previous section. The vector includes parental education, child's age (in months), household size, wealth index (a proxy for household socio-economic status), dummy variables for caste, gender, being the first-born child and for female-headed households. As is immediately obvious, there are important differences between children enrolled in different institutions: children in government institutions are more likely to be from poorer households, with less educated parents, living in rural areas and much more likely to belong to deprived caste groups (Scheduled Castes and Scheduled Tribes). *A priori* it is conceivable, indeed plausible, that most or even all of the observed test score gaps between the different institution types could be a reflection of these background characteristics and not a reflection on the effectiveness of the institutions themselves.

In order to investigate this possibility further, I estimate multivariate (Ordinary Least Squares) regressions which control for the background variables presented in Table 2 and dummy variables for the educational institutions/trajectories that the children were enrolled in. This allows for inspecting the partial correlations between test scores and these variables as well as to see how much the test score gaps decline upon inclusion of the set of background variables.

Specifically, I propose to estimate two sets of two regressions each for both the test scores available in the data. The equations are specified as follows:

$$(1) \quad Y_i = \alpha + \beta_1 \cdot \mathbf{enrol}_i + \beta_2 \cdot \mathbf{site}_i + \varepsilon_i$$

$$(2) \quad Y_i = \alpha + \beta_1 \cdot \mathbf{enrol}_i + \beta_2 \cdot \mathbf{site}_i + \beta_3 \cdot \mathbf{X}_i + \varepsilon_i$$

In the first specification, the test scores (Y_i) are related to the current enrolment of the child and a vector of site fixed effects which are included as a dummy variable for each cluster excluding one as the base¹³. In the second specification, I additionally add a vector of control variables \mathbf{X} which include all the variables presented in Table 2¹⁴. Standard errors are clustered at the site level in the estimation.

It is possible to define the current enrolment of the child in different ways. In the empirical analysis, I first run the two regression specifications presented above with dummy variables only for the current enrolment status of the child i.e. a dummy variable each for public preschool, private preschool, public school and private school with children not enrolled in any educational institution as the base category. Results from this exercise are presented in Table 3.

Regardless of the institution a child is enrolled in, there is a significant (positive) difference in the test scores from the base category of non-enrolled children, both with and without control variables added in; the only exception is the PPVT score for children enrolled in government preschools where the coefficient, albeit positive, is statistically insignificant. The second pattern to note is that the point estimates across the four institution types differ importantly in magnitude from a 0.25 SD coefficient on public preschools for the CDA scores to about a 1 SD coefficient for private schools in the specification without controls (Col. 1). I use F-tests to assess whether these differences are statistically significant. Two things are important to note: first, for both preschools and schools, the coefficient on the private sector institution is significantly higher than the corresponding public sector category; second, in both sectors, private and public, the coefficient for primary school are significantly higher than for preschools¹⁵. Finally, the third pattern to note is that whereas the magnitude of the gap between the coefficients on the public and private sector institutions does decline with the inclusion of controls (as expected), the gaps remain statistically significant and important in magnitude: for the CDA z-score, while the regressions not including the background variables report a difference of about 0.5 SD between the coefficient of private and public sectors at both preschool and primary school level, this declines by half to about 0.25 SD upon inclusion of the controls; for PPVT z-scores, the decline is rather more from about 0.5 to 0.2 SD for primary schools, upon inclusion of the controls but these gaps remain significant.

¹³ Site fixed effects, included here by including a vector of site dummies, allow for removing any levels differences between different sites (mandals). This is particularly important in this case since the take-up of different institution types varies much across the sites in the sample. In effect, these regressions compare children who are in different types of education but living in the same cluster.

¹⁴ The wealth index is entered with dummy variables for the terciles of the wealth index (with the poorest third being the base category). This procedure is adopted to allow for non-linearities in the relationship between wealth and test score, and to explicitly recognize that wealth indices based on asset ownership are not inherently cardinal.

¹⁵ This pattern should not, of course, be interpreted causally: it is, for instance, possible that this pattern is caused by the transition into formal schooling being brought forward by parents for children with higher academic achievement or potential.

Subsequently, instead of controlling for the current institution-type the child is enrolled in, I control for the trajectory that he/she has taken thus far i.e. a dummy variable for each of the categories presented in Table 1 with the exception of children who have never been enrolled in either preschool or in school (who form the base category). These results are presented in Table 4 with relevant F-tests as in the previous table.

Two of the key patterns that were first distinguished in Table 3 also hold true for this set of results: first, all trajectories presented in the table out-perform the base category (never having attended preschool or school) significantly, both with and without controls, and for both CDA and PPVT z-scores; second, for the four ‘exclusive’ trajectories i.e. where the child has attended only a preschool or a school but not both types of institutions, private sector institutions report significantly higher coefficients than public sector counterparts. However the table also allows us to investigate further patterns. The first is whether having attended a preschool previously is associated with students currently enrolled in the same type of school performing differently. Specifically, I conduct F-tests to evaluate whether the coefficient on having only gone to public primary school differs from the coefficient for having gone to both public preschool and school and a similar test for the private sector¹⁶. The picture here is rather more mixed: among children currently enrolled in public school, having attended a public preschool is associated with significantly higher CDA test scores than having just attended public primary school but a similar association is not perceivable in PPVT scores or in the private sector.

Results presented here need to be interpreted with caution and it is worthwhile considering what they do and don’t show. The most important caveat to note is that the results cannot be interpreted causally – in particular, the threat of omitted variables remains important and may be responsible for the difference in the average test performance of children across institution types rather than the effectiveness of the institutions themselves¹⁷. What can however be said with some confidence on the basis of the results is that, even if public institutions do not themselves *cause* students enrolled therein to do worse than children in private preschools and schools, they certainly do not succeed in levelling out social disadvantage: children in private educational institutions, who come from relatively better-off socio-economic backgrounds, continue to do better in test scores by a large magnitude. The second pattern that is also clearly evident throughout the analysis, and of central importance to the questions raised in the introduction, is that the gaps between children in private and state sectors in education predate their entry into school: this is important because even descriptive and associational evidence in this regard has been scarce.

5. Conclusions

These results tell a clear story. The differences in the learning outcomes between children in public and private sectors in education start early and are, even at that early stage, substantial.

¹⁶ The number of children who cross over between sectors when going from preschool to school is very limited in the sample, thus precluding sufficiently precise statistical inference.

¹⁷ Note that all the aspects of enrolment as observed in the data, including whether a child is enrolled, the type of institution enrolled in and the age of enrolment into both preschool and school are choices made by parents, conditional on endowments. It is plausible that the determinants of these choices (e.g. parental preferences over education), some of which might be unobserved in the data, also independently affect learning and/or they influence not just choices around enrolment but also other investments (such as time spent reading to children) which then affect test scores.

Two policy issues are currently prominent in discussion of educational policy, one particularly relevant for India and the other of wider application. The first is the role of the private relative to the public sector in education¹⁸. The second is a literature referred to in the introduction, mostly from OECD countries but also with increasing evidence from developing countries, arguing that many of the crucial differences that affect young adults in both their learning and labour market outcomes are due to their early learning history.

This paper supplies evidence relevant to both these strands of research. Differences in the learning levels of children in private and state schools in India have, as noted above, been extensively studied. This paper suggests that these differences are also large at the point of school entry. The issue for public policy is how much of this difference can be imputed to the preschool and how much to the differences in the children who attend the preschool. I have not sought to address that issue directly as I have not attempted to estimate a treatment effect for being in a private or public preschools; results presented here cannot be interpreted as causal ‘effects’ of attending particular types of educational institutions and remain vulnerable to the (non-trivial) possibility of omitted variable bias.

The combined effects from school and parental background produce large differences in learning very early in the child’s life. If the recent literature on skill formation is correct in arguing that these early differences in learning outcomes can have very long run implications for the life chances of these children (see for example Cunha and Heckman 2008, Cunha, Heckman and Schennach 2010) then these differences and their causes matter considerably for policy. The argument in these papers is that skill formation is a sequential process and if the early building blocks are not in place then there is a much weaker foundation for education further down the chain to build on. The window of opportunity for effective interventions lies earlier in childhood and that ‘the later remediation is given to a child, the less effective it is’ (Cunha and Heckman, 2008). If this is the case then the sizeable gap between outcomes in the private and the public sector which I have identified will have very long run implications for the life chances of these children. That this is the case, at least until the age of 10 years of age, has been documented in detail by a complementary study by Singh (2013)¹⁹.

As is shown here, test score gaps emerge early and are systematically related to previous preschool history, although not necessarily caused by it, even conditional on a set of background characteristics; as Singh (2013) has shown, these gaps then determine cognitive gaps measured later in life. More broadly, the results also indicate four important questions which have not yet been answered conclusively in the Indian context. *First*, they highlight that the question of what is the causal effect of attending a private or public preschool is of central importance if we are interested in understanding the divergence between the cognitive skill acquisition of children from different backgrounds. While our data are insufficient to answer this question robustly, such investigations may be possible, for example through field experiments (e.g. lottery-based vouchers). *Second*, we need a more detailed

¹⁸ In addition to the studies cited in Section 2, see also the article by James and Woodhead (2013) in this issue and the recent special issue of this journal on private and public education (Vol. 39, No. 4).

¹⁹ Singh (2013) uses the third round of the Young Lives survey in 2009/10, along with data collected from schools in 2011, to document that gaps between students enrolled in private and government schools in Maths and Telugu, which are large cross-sectionally, are significantly reduced upon controlling for the gaps in PPVT and CDA that are documented in this paper along with controls for the differing socio-economic background of children; controlling further for differences in the time-use patterns across children in private and government schools, the gap is closed entirely. Clearly, the gaps documented in this paper are central in explaining the differences in human capital acquisition of children from different backgrounds educated in the private and state sectors.

understanding of what drives enrolment choices surrounding whether children are enrolled, where, and at what age, in order to identify policy levers that may be utilized to encourage preschool attendance (assuming a positive causal effect) at the correct age. *Third*, if future research does find that the out-performance of the private sector at the preschool level is a causal effect, it would be very important to identify why public preschools underperform preschools in the state sector. *Finally*, even if part of the test score gap between private and state preschool students documented here is causal, it is likely that preschools themselves do not account for the whole gap. From a policy perspective it is important to identify what other levers (for instance early childhood nutrition, broader public health investments targeting disease environments and early remedial classes) might ameliorate the emergence of this gap at all, especially if these other levers are more easily manipulable and/or cost-effective.

References

- Atherton, P., & Kingdon, G. (2010). The relative effectiveness and costs of contract and regular teachers in India. *Centre for the Study of African Economies (CSAE) Working Paper Series*, 15.
- Banerjee, A., S. Cole, E. Duflo and L. Linden (2007): Remedying Education: Evidence from Two Randomized Experiments in India, *Quarterly Journal of Economics*, August 2007, Vol. 122, No. 3, Pages 1235-1264
- Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1), 219-234.
- Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review*, 31(4), 376-390.
- Citizen's Initiative for the Rights of Children Under Six (2006) Focus on Children Under Six Abridged Report, Secretariat of the Right to Food Campaign, New Delhi
- Cueto, S., J.Leon, G. Guerrero, and I. Munoz (2009): Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives, Young Lives Technical Note 15, available at www.younglives.org.uk
- Cunha, F., Heckman, J., Lochner, L. and Masterov, D.V. (2005) Interpreting the Evidence on Life Cycle Skill Formation, IZA Discussion Paper No. 1675 available at SSRN: <http://ssrn.com/abstract=766744>
- Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-782.
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883-931.
- Das, J., & Zajonc, T. (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, 92(2), 175-187.
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2), 29-57.
- Desai, S., Dubey, A., Vanneman, R., & Banerji, R. (2009). Private schooling in india: A new educational landscape. In *India Policy Forum* (Vol. 5, pp. 1-58).
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4), 1241-78.
- Dunn, L. and L. Dunn (1997) Examiner's Manual for the PPVT-III. Form IIIA and IIIB, Minnesota: AGS.

- Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P. J., Kapiriri, L., ... & Young, M. E. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, 369(9557), 229-242.
- French, R., & Kingdon, G. (2010). The relative effectiveness of private and government schools in Rural India: Evidence from ASER data. *Department of Quantitative Social Science Working Paper10-03, Institute of Education, University of London, London.*
- Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2), 447-464.
- Fryer, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, 8(2), 249-281.
- Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240.
- James, Z. and Woodhead, M. (2013) Choosing and changing schools in India's private and government sectors, *Oxford Review of Education*
- Kingdon, G. G. (1996): The Quality and Efficiency of Public and Private Schools: A Case Study of Urban India, *Oxford Bulletin of Economics and Statistics*, 58, No.1: 55-80, February 1996
- Kingdon, G. G. (2007). "The progress of school education in India." *Oxford Review of Economic Policy* 23.2 (2007): 168-195.
- Kingdon, G. G. (2008): Private and public schooling: The Indian experience, in Chakrabarti, Rajashri, and Paul E. Peterson (Eds.) *School Choice International: Exploring Public-Private Partnerships*. MIT Press. Cambridge, MA, 2008.
- Kremer, M., K. Muralidharan, N. Chaudhury, J. Hammer, and F. H. Rogers (2005): "Teacher Absence in India: A Snapshot" *Journal of the European Economic Association*, Volume 3, pp 658-667.
- Muralidharan, K. and Kremer, M. (2008) : Public and Private Schools in Rural India in Chakrabarti, Rajashri, and Paul E. Peterson (Eds.) *School Choice International: Exploring Public-Private Partnerships*. MIT Press. Cambridge, MA, 2008.
- Muralidharan, K., & Sundararaman, V. (2010). The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India*. *The Economic Journal*, 120(546), F187-F203.
- Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1), 39-77.
- Muralidharan, K., & Sundararaman, V. (2012). Contract Teachers: Experimental Evidence from India. Unpublished *UC San Diego*.

Muralidharan, K. & Sundararaman, V. (2013). The Aggregate Effect of School Choice - Evidence from a two-stage experiment in India. *NBER Working Paper 19441*

Pratham (2013). Annual Status of Education Report 2012. Technical report, Pratham, New Delhi.

PROBE Team (1999): *Public Report on Basic Education in India*. New Delhi: Oxford University Press.

Schweinhart, L. J. (1993). *Significant Benefits: The High/Scope Perry Preschool Study through Age 27. Monographs of the High/Scope Educational Research Foundation, No. Ten*. High/Scope Educational Research Foundation, 600 North River Street, Ypsilanti, MI 48198-2898.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: the High/Scope Perry Preschool study through age 40. Ypsilanti:High/Scope Press

Singh, A. (2013). Size and Sources of the Private School Premium in India, Young Lives Working Paper 98, University of Oxford

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The effective provision of pre-school education (EPPE) project: Final Report: A longitudinal study funded by the DfES 1997-2004*. Institute of Education, University of London/Department for Education and Skills/Sure Start.

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (eds) (2010) *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education Project*. London: Routledge.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*,1(1), 91-136.

Tooley, J., & Dixon, P. (2006). 'De facto'privatisation of education and the poor: implications of a study from sub-Saharan Africa and India. *Compare*, 36(4), 443-462.

Tooley, J., Dixon, P., & Gomathi, S. V. (2007). Private schools and the millennium development goal of universal primary education: a census and comparative survey in Hyderabad, India. *Oxford Review of Education*, 33(5), 539-560.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. Springer.

Walker, S.P., Wachs, T.D., Grantham-McGregor, S., M. M. Black, C. A. Nelson, Sandra L Huffman, Helen Baker-Henningham, Susan M Chang, Jena D Hamadani, Betsy Lozoff, Julie M Meeks Gardner, Christine A Powell, Atif Rahman, Linda Richter, Inequality in early childhood: risk and protective factors for early child development, *The Lancet*, Volume 378, Issue 9799, 8–14 October 2011, Pages 1325-1338.

Woodhead, M., Ames, P., Vennam, U., Abebe, W., & Streuli, N. (2009). Access, Equity and Quality in Early Education and Transitions to Primary School: Evidence from Young Lives Research in Ethiopia, India And Peru, *Young Lives Working Paper 55*.

Figure 1

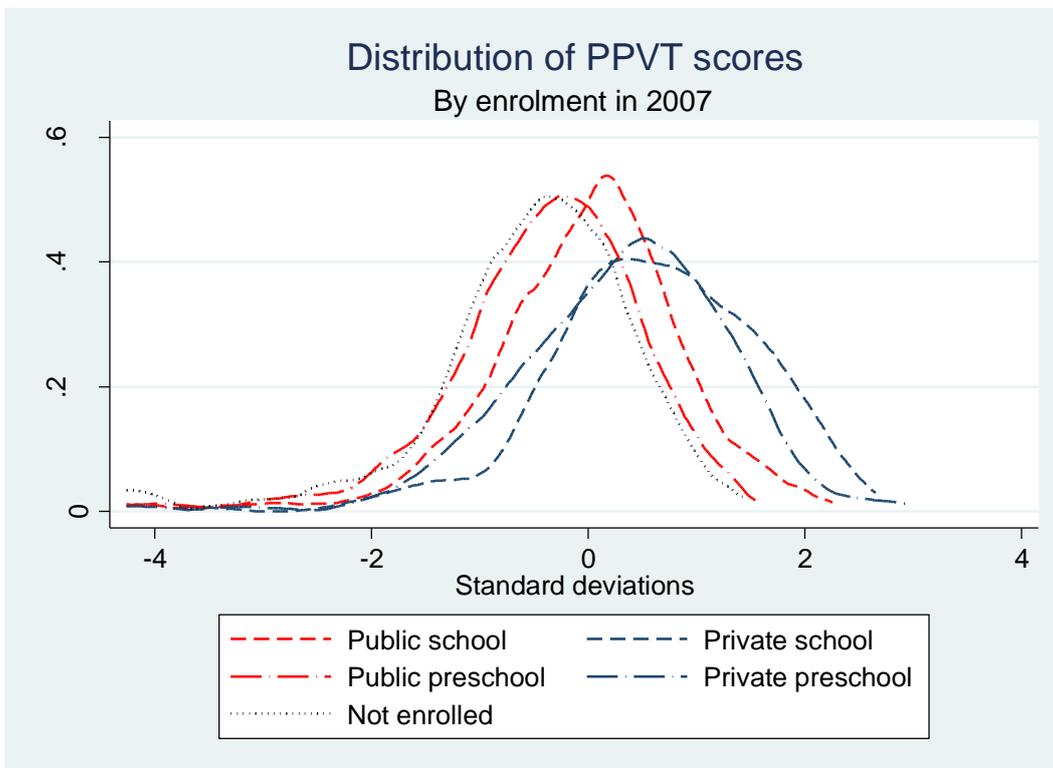


Figure 2

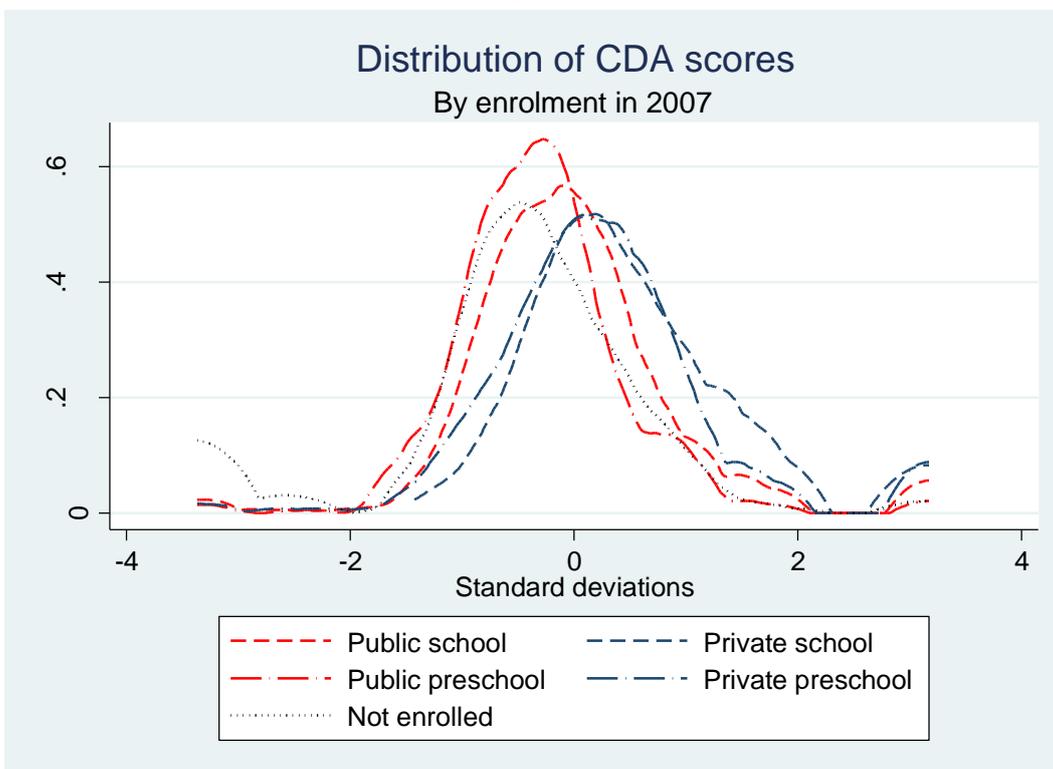


Table 1 Descriptive statistics of CDA and PPVT test scores, by trajectory

Trajectory	Statistics	CDA	PPVT
No preschool or school	Mean	-0.87	-0.72
	Median	-0.64	-0.57
	Standard deviation	1.2	1.1
	N	86	76
Private Preschool only	Mean	0.28	0.38
	Median	0.21	0.44
	Standard deviation	1	0.95
	N	499	447
Public Preschool only	Mean	-0.26	-0.38
	Median	-0.3	-0.29
	Standard deviation	0.83	0.9
	N	496	475
Public Primary school only	Mean	-0.21	-0.15
	Median	-0.27	0.015
	Standard deviation	0.87	0.79
	N	126	119
Private Primary school only	Mean	0.64	0.74
	Median	0.32	0.57
	Standard deviation	1.1	0.95
	N	42	41
Private Preschool to Public school	Mean	0.13	0.37
	Median	-0.0039	0.39
	Standard deviation	0.87	0.84
	N	18	17
Private Preschool to Private school	Mean	0.48	0.66
	Median	0.3	0.88
	Standard deviation	0.99	0.92
	N	61	54
Public Preschool to Public school	Mean	0.0063	-0.023
	Median	-0.052	0.063
	Standard deviation	0.96	0.96
	N	549	538
Public Preschool to Private school	Mean	0.44	0.4
	Median	0.35	0.43
	Standard deviation	0.86	0.98
	N	64	62
Total	Mean	0	0
	Median	-0.08	0.058
	Standard deviation	1	1
	N	1941	1829

Table 2 Descriptive statistics of socio-economic background, by enrolment

variable	Public school		Private school		Public preschool		Private preschool		Not enrolled		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	mean	sd	mean	sd
Male	0.52	0.50	0.56	0.50	0.50	0.50	0.57	0.49	0.52	0.50	0.53	0.50
Urban	0.06	0.23	0.54	0.50	0.09	0.28	0.58	0.49	0.13	0.34	0.26	0.44
Female-headed household	0.05	0.22	0.04	0.19	0.05	0.21	0.06	0.24	0.06	0.24	0.05	0.22
First-born child	0.36	0.48	0.49	0.50	0.33	0.47	0.46	0.50	0.28	0.45	0.39	0.49
Father's education (in years)	4.17	4.37	8.36	4.58	3.54	4.05	8.48	4.89	4.19	4.47	5.60	4.98
Mothers Education (in years)	2.18	3.30	6.32	4.87	1.93	3.05	6.61	4.80	1.85	3.13	3.70	4.43
Wealth index	0.37	0.15	0.62	0.16	0.36	0.15	0.61	0.17	0.38	0.17	0.46	0.20
Age (in months)	66.01	3.27	65.88	3.71	63.52	3.82	64.28	3.98	63.30	4.07	64.80	3.85
Scheduled Caste	0.24	0.43	0.13	0.34	0.20	0.40	0.09	0.29	0.26	0.44	0.18	0.39
Scheduled Tribes	0.19	0.39	0.08	0.27	0.13	0.33	0.06	0.23	0.16	0.37	0.13	0.33
Other Backward Classes	0.45	0.50	0.41	0.49	0.57	0.50	0.48	0.50	0.41	0.49	0.48	0.50
Other Castes	0.12	0.32	0.38	0.49	0.10	0.31	0.37	0.48	0.17	0.38	0.21	0.41
Household Size	5.51	2.27	5.31	1.93	5.82	2.37	5.34	2.14	5.62	2.13	5.52	2.23
N	689		160		413		549		130		1941	

N.B. The sample used in this table is the full estimation sample used for CDA analysis.

Table 3 Association of test scores with current enrolment and background characteristics

VARIABLES	(1)	(2)	(3)	(4)
	CDA z-score		PPVT z-score	
Public school	0.56*** (0.099)	0.46*** (0.098)	0.46*** (0.10)	0.38*** (0.11)
Private school	1.03*** (0.14)	0.72*** (0.13)	0.93*** (0.15)	0.57*** (0.13)
Public preschool	0.25** (0.095)	0.28*** (0.090)	0.13 (0.090)	0.16 (0.11)
Private preschool	0.78*** (0.11)	0.53*** (0.090)	0.74*** (0.15)	0.43*** (0.12)
Household size		-0.0031 (0.0099)		-0.0046 (0.0090)
Mother's education level		0.034*** (0.0073)		0.043*** (0.0061)
Father's education level		0.023*** (0.0048)		0.025*** (0.0071)
Age in months		0.035*** (0.0065)		0.031*** (0.0061)
Female-headed household		-0.29*** (0.088)		-0.28** (0.12)
First-born child		0.024 (0.044)		-0.0098 (0.047)
Constant	-0.56*** (0.087)	-2.98*** (0.43)	-0.47*** (0.098)	-2.69*** (0.36)
Observations	1,941	1,941	1,829	1,829
R-squared	0.156	0.216	0.216	0.291
F-tests for equality of coefficients (p values)				
Private preschool = Public preschool	.00	.00	.00	.00
Private school = Public school	.00	.02	.00	.04
Private preschool = Private school	.00	.02	.02	.03
Public preschool = Public school	.00	.00	.00	.00

Robust standard errors in parentheses. Standard errors clustered at site level.
Regressions also include dummies for castes, sites, terciles of wealth index and being male for which coefficients are not reported.

*** p<0.01, ** p<0.05, * p<0.1 testing that coefficients are significantly different from zero.

Table 4 Association of test scores with prior and current educational path

VARIABLES	(1)	(2)	(3)	(4)
		CDA		PPVT
Private Preschool only	1.15*** (0.14)	0.85*** (0.12)	1.01*** (0.21)	0.65*** (0.16)
Public Preschool only	0.62*** (0.11)	0.61*** (0.12)	0.32* (0.16)	0.30* (0.17)
Public Primary school only	0.71*** (0.12)	0.61*** (0.13)	0.55*** (0.15)	0.47*** (0.16)
Private Primary school only	1.33*** (0.19)	1.06*** (0.15)	1.11*** (0.22)	0.79*** (0.20)
Private Preschool to Public school	0.80*** (0.28)	0.65** (0.29)	0.75** (0.26)	0.62** (0.22)
Private Preschool to Private School	1.26*** (0.22)	0.85*** (0.17)	1.15*** (0.21)	0.69*** (0.15)
Public Preschool to Public School	0.90*** (0.13)	0.78*** (0.13)	0.64*** (0.19)	0.53*** (0.18)
Public Preschool to Private School	1.38*** (0.18)	1.12*** (0.18)	1.09*** (0.24)	0.80*** (0.21)
Household size		-0.0024 (0.0098)		-0.0030 (0.0084)
Mother's education level		0.034*** (0.0073)		0.043*** (0.0059)
Father's education level		0.023*** (0.0046)		0.025*** (0.0069)
Age in months		0.035*** (0.0062)		0.030*** (0.0058)
Female-headed household		-0.29*** (0.084)		-0.28** (0.13)
First-born child		0.017 (0.047)		-0.019 (0.048)
Constant	-0.87*** (0.11)	-3.27*** (0.39)	-0.66*** (0.17)	-2.76*** (0.32)
Observations	1,941	1,941	1,829	1,829
R-squared	0.171	0.228	0.231	0.298

F-tests of equality of coefficients

Pvt. preschool = Public preschool	.00	.012	.00	.00
Pvt. school = Public school	.00	.00	.00	.05
Public Primary = Pub. preschool to Pub. school	.06	.09	.46	.56
Pvt. Primary school = Pvt. preschool to Pvt. school	.69	.07	.61	.47

Robust standard errors in parentheses. Standard errors clustered at site level.

Regressions also include dummies for castes, sites, terciles of wealth index and being male for which coefficients are not reported.

*** p<0.01, ** p<0.05, * p<0.1 testing that coefficients are significantly different from zero.