Preschool and primary school markets: Evidence from India*

Petter Berg[†] Mauricio Romero[‡] Abhijeet Singh[§]

November 18, 2024

Abstract

We study education markets at preschool and primary school levels using panel data from 220 villages in Tamil Nadu, India. Private preschools show higher test score value-added in math and language ($\sim 0.57 - 0.73\sigma$) and outperform government providers in nearly all villages. This productivity difference explains 60% of the socioeconomic test score gap before school entry. Test score value-added is positively correlated between private and government options in a village, both at preschool and primary school levels. Our findings inform debates on achieving universal foundational skills and underscore the need to improve the quality of preschools available to poorer families.

Keywords: preschool markets, primary markets, education systems

JEL Codes: H44, H52, I21, I25, I28, L10, L33

*We are grateful for comments from Jishnu Das and Ajinkya Keskar. Ramamurthy Sripada, Dwithiya Raghavan, Akanksha Soni, and Indira Tirumala provided excellent field management and research assistance. We received IRB approval from the Institute for Financial Management and Research (#113401). Financial support from the Asian Development Bank and the RISE Programme, — funded by the UK's Foreign, Commonwealth and Development Office, Australia's Department of Foreign Affairs and Trade, and the Bill and Melinda Gates Foundation — is gratefully acknowledged. Romero gratefully acknowledges financial support from the Asociación Mexicana de Cultura, A.C. All errors are our own.

[†]Stockholm School of Economics; E-mail: petter.berg@phdstudent.hhs.se

[‡]ITAM; J-PAL; E-mail: mtromero@itam.mx

[§]Stockholm School of Economics; J-PAL; E-mail: abhijeet.singh@hhs.se

1 Introduction

Over 250 million children under 5, mostly in low- and middle-income countries (LMICs), do not fulfill their cognitive potential (Engle et al., 2007; Grantham-McGregor et al., 2007; Behrman et al., 2013; Black et al., 2017). Thereafter, a substantial fraction of children fail to acquire foundational literacy and numeracy in primary school (World Bank, 2017). However, although this "learning crisis" reflects the contributions of both preschools *and* schools, policy and research attention have typically been centered around primary schools — in particular, we know relatively little about how preschool markets are organized, the differential in productivity across private and public preschool providers, the heterogeneity across markets, and how these results compare with primary school markets (where there is a more substantial literature).[1] Such evidence is important for understanding where learning gaps emerge and how they may be rectified, and its lack is especially surprising given explicit international policy targets for providing quality preschool services (World Bank, 2017; Holla et al., 2021).[2]

We address this gap using individual-level panel data on a broad sample of children aged 3–10 in 220 rural villages in the Indian state of Tamil Nadu. Our panel includes multiple measures of socioeconomic status and educational inputs and, importantly, tests of student achievement in math and the local language (Tamil). These tests are age-appropriate and, using common items, can be used to generate vertically-equated test scores on a common scale both over time and across ages. Children in these villages can attend preschool at free public child development centers (*anganwadis*) or fee-charging private preschools, followed by enrollment in either government or private primary schools. Since children rarely attend preschools or schools outside their village, each village is a separate educational market.[3] Although preschool is not compulsory, nearly all children are enrolled at age 4. Primary school enrollment is compulsory by law and near-universal by age 6. Private providers account for roughly one-third of enrollment at preschool age and one-quarter in primary schooling.

We use these data to conduct four exercises. First, we compute test score value-added in private and government preschools and schools, focusing on the differential value-added of private providers over government options ("the private preschool/school premium").

[1] See, for instance, Allende (2019); Neilson (2021); Andrabi et al. (2022); Bau (2022) for recent analyses of market structure, productivity and competition for primary schools in LMICs.

[2] Globally, primary school education is almost universal (World Bank, 2019). However, as of 2020, pre-primary education enrollment rates were 60% worldwide and below 20% in low-income countries (World Bank, 2020). Privately provided education is also more common around the world at the pre-primary (36% in 2022) than at the primary level (19% in 2022) (UNESCO, 2022).

[3] Previous studies on primary school markets exploit this institutional feature in similar settings. One closely related context is the LEAPS program in Pakistan (Andrabi et al., 2017, 2022).

Compared to public options, the private preschool premium on test score value-added is substantial at 0.73 and 0.57 standard deviations (σ) of the test score distributions in math and Tamil. This premium is roughly double the (cross-sectional) achievement difference in language between 3-year-olds and 4-year-olds enrolled in public preschools, and roughly four times the difference in math. In contrast, we find no evidence of a private premium in primary schools on average: value-added in private schools is indistinguishable from that of government options in math and significantly lower in the local language (by $\sim 0.17\sigma$).[4]

Second, we compute *village-specific* estimates of the productivity of government and private preschools and schools (recognizing the potential for substantial heterogeneity). The estimated productivity of preschools and schools varies widely across villages but value-added in government preschools is dominated by that of private options in virtually *all* villages. Further, there is very limited common support in the distributions of value-added between government and private preschools in both language and math.[5] In contrast, there is near complete overlap in value-added in math and substantial overlap in Tamil at the primary school level.

Third, using village-specific estimates of the productivity of government (private) preschools and schools, we investigate market-level correlations between the productivity of different sectors. The average value-added of private preschools and schools is positively correlated with that of local public options: an increase in the value-added of public preschools in a village of 1σ of student achievement predicts 0.2–0.37σ higher private sector productivity; in primary schooling, this correlation is 0.49 and 0.59 respectively. Further, the value-added of both government and private providers is also correlated across levels — i.e., higher preschool value-added in the government (private) sector also predicts higher value-added in government (private) primary schooling.

Fourth, and finally, we estimate the contribution of private education in explaining socioeconomic gaps in student achievement. Specifically, we examine the cross-sectional gap in test scores between the top and bottom quartiles of socioeconomic status (SES) and quantify the extent to which differential private sector enrollment (which is significantly higher for students from top-quartile households) explains the gap. Reflecting our results above, differential private sector enrollment accounts for approximately 60% of the SES test score gap at preschool ages but does not explain SES gaps at later ages.

[4] These results for primary schools are similar to those reported by Muralidharan & Sundararaman (2015) and Singh (2015) in the neighboring states of Andhra Pradesh and Telangana, and likely represents a greater focus on English over the local language in private primary schools.

[5] This is consistent with prior evidence that government care centers typically provide little structured cognitive stimulation (Ganimian et al., 2024), whereas private preschools focus much more substantially on early childhood education (Singh, 2014).

The analyses above provide a *unified* treatment of preschool and primary school markets using comparable test scores and a large set of markets. Consequently, we contribute novel facts to multiple distinct strands of economic research on education systems.

First, we contribute to the literature on private schooling and school markets. Our empirical approach is closest to that of Andrabi et al. (2022), who study the heterogeneity in private and government primary school value-added across village markets in Pakistan.[6] We complement this literature, which has mostly focused on primary education, by studying preschools, which differ from primary schools in terms of organization, funding, and take-up, while providing comparable results from primary schools as a benchmark. In our setting, the private sector premium in productivity is much higher at the preschool level than in primary education (where we find no evidence of greater value-added). As a result, its contribution to socioeconomic inequality in achievement is also much greater.

Second, we complement the literature on improving children's cognitive skills before they start primary school. This literature has mostly focused on evaluating individual interventions — for example, programs to support parents (e.g., Heckman et al. (2010); Attanasio et al. (2014); Andrew et al. (2024); Attanasio et al. (2022)), to improve preschools (e.g., Ganimian et al. (2024)), or to send children to private preschools (e.g., Dean & Jayachandran (2019); Bjorvatn et al. (2024)). In contrast, we focus on the characteristics of preschool markets, thus highlighting *where* such interventions may be needed most. In particular, addressing the low productivity of government preschools appears central both for improving education quality and for addressing socioeconomic inequality in learning. This is especially relevant for global policy discussions on ensuring universal foundational skills in childhood (see, e.g., World Bank (2017); Muralidharan & Singh (2021)).

Finally, we contribute to the literature on socioeconomic inequality in early childhood cognitive skills (e.g., Engle et al. (2011); Fernald et al. (2012); Rubio-Codina et al. (2015); Schady et al. (2015); Elango et al. (2015); Reynolds et al. (2017)). This work has documented the existence and evolution of disparities in the cognitive skills of young children from more- and less-advantaged backgrounds in multiple settings.[7] We show the extent to which test score gaps at the time of primary school entry between students from more- and less-advantaged households result from differential exposure to private preschools.

[6]For other examples in South Asia, which focus only on an average private school effect, see e.g., Andrabi et al. (2011); Singh (2015); Muralidharan & Sundararaman (2015). A broader review of private schooling markets in developing countries is presented by Crawfurd et al. (2024).

[7]The analogous literature in the United States studies racial disparities from very young ages (Fryer Jr & Levitt, 2004, 2006, 2013), as well as the income-achievement gap (Reardon, 2011, 2021; Nielsen, 2023).

2 Context and data

2.1 Context

Primary school enrollment in India is mandatory and near-universal. The official school starting age is 6, although many students start at age 5. Government schools are free for students to attend and also offer in-kind benefits, including free cooked school meals, textbooks, and uniforms. Nearly all government schools use the local state language as the medium of instruction. Private schools charge tuition fees, and a substantial fraction of them use English as the medium of instruction. There is wide heterogeneity in the price, education quality, and amenities of private providers.

Children between 3 and 5 years of age are meant to be enrolled in preschools. However, preschool is not mandatory, and enrollment rates vary substantially across states. The principal public preschool option is *anganwadi* centers, which are part of the Integrated Child Development Services (ICDS) program — the largest early childhood program in the world. ICDS provides early childhood education, nutritional supplementation and child health services in 1.35 million *anganwadi* centers around the country that serve 36 million children between ages 3 to 6 (Ganimian et al., 2024). Services are provided free-of-charge with open enrollment. Private preschools charge fees and focus on pre-primary education in Nursery and kindergarten classes. They are often integrated with private primary schools.

Our study is based in rural areas of Tamil Nadu, a large south Indian state with an estimated population of 74 million and an education system that serves 13 million children annually (Government of India, 2019). In Tamil Nadu, preschool enrollment is near-universal: more than 96% of 4-year-olds in the state are enrolled in preschool (Pratham, 2022). About 61% of 4-year-olds are enrolled in *anganwadi* centers and about 33% in private preschools — the remaining 4-year olds are already enrolled in primary school education. Tamil Nadu's public early childhood education system is considered high-performing within India but remains under-resourced; on average, only about 20 minutes per day are spent on academic activities (Ganimian et al., 2024; Singh & Romero, 2022). School-readiness levels also remain low: fewer than 10% of children can read individual words when they enter primary school, and 60% cannot recognize individual letters (Pratham, 2022).

2.2 Data

2.2.1 Sample

Our data covers 220 villages in four districts of Tamil Nadu. In these communities, we administered comparable achievement tests to students aged 3–10 in early 2022 (baseline) and 2023 (endline). Our core analysis sample includes the set of children aged 4–10 in 2023 for which we have access to baseline assessments (N=19,021).[8] Although these villages were not randomly selected, our sample is mostly similar in observable characteristics to the state's rural population (see Table A.1).[9]

2.2.2 Assessments

The learning assessments were designed to capture student achievement across the preschool and primary populations. Children were tested in math and Tamil (the local language) using age-appropriate booklets and overlapping items. For preschoolers (ages 3–4), the tests captured oral comprehension, letter recognition, quantitative comparisons, number recognition, and counting; at ages 5 and 6, they also included word recognition, more complex counting, and basic addition; for children aged 6–10, the tests additionally included more complex arithmetic computation and word problems in math and passage comprehension and reading exercises in Tamil. Test booklets included common items across waves and ages, which allows us to link achievement on a common metric using Item Response Theory (IRT) models (Das & Zajonc, 2010). We estimate these scores by pooling all test observations across rounds, separately for math and Tamil. We standardize test scores to have mean zero and standard deviation of one in the sample of children aged 5 in the 2022 survey wave. The test scores display, as expected, a shifting of the distribution of achievement with age (indicating skill acquisition over time; see Figure E.3) and no evidence of Differential Item Functioning by age or by round (see Figures E.4–E.29). See the Online Appendix for more details.[10]

[8] Attrition between survey rounds was ∼25%, but does not vary by socioeconomic status (SES) or test scores (see Appendix B). The survey waves were administered slightly more than a year apart. Since we only have data on age in completed years (and not months), we measure attrition for children aged 3–9 in 2021/2022.

[9] This data was originally collected for an experimental evaluation of a government program to improve preschool education. The intervention and the evaluation were canceled due to the COVID-19 pandemic and subsequent preschool and school closures. See `https://doi.org/10.1257/rct.5599` for more details. We continued to collect data to study the learning loss during the pandemic and the pace of recovery afterward (Singh et al., 2024).

[10] In 2022, we divided the sample into two randomly-assigned groups within villages that were administered the tests in a staggered manner between December 2021 and April 2022 (Singh et al., 2024). We administered the tests with a similar staggering and the same assigned groups in 2023 to maintain a similar gap between assessments. In 2023, we see some signs of ceiling effects for children of school age in the first round of testing. We remedied this by adjusting test booklets for the second (randomly-assigned) testing round, keeping common items for linking. No results are sensitive to only using the second round.

2.2.3 Household survey

We collected extensive household data about their socioeconomic status and children's education in both survey waves. We use detailed information about household ownership of various assets in 2022 to construct a socioeconomic status index using principal component analysis (PCA). We use a household's percentile rank in this index as the main measure of SES (see Appendix C for details). We also use maternal and paternal education information as additional measures of SES and control for them in our regressions.

2.2.4 Strengths and limitations of the data

Our data have several important strengths. The most important of these is the availability of panel data on achievement for children over the full span of preschool and primary school ages across a large number of spatially disjoint education markets. A related strength is the comparability of measurement over time and across ages: datasets with vertically-linked IRT scores are uncommon in low- and middle-income countries but are crucial to our goals of expressing preschool and primary school productivity on a common scale. The final strength is the complete enumeration of households in sampled geographical areas: this prevents the attrition typical in many school-based surveys due to student absence on the day of testing.[11]

However, the expansiveness of the dataset also imposes some trade-offs that limit our analysis. Most importantly, we can only provide sector-specific estimates within a village rather than estimates for every facility separately. This issue arises for a combination of reasons. First, matching children to centers generates substantial measurement error since we only collected facility names (which are hard to map to individual facilities, especially government facilities that do not have distinctive names). Second, because the total number of children in each facility is often very small, any individual facility estimates would be very noisy even with complete matching. Finally, in large villages, we restricted our censuses to a radius of $\sim$2 kilometers from a reference point; this does not affect our interpretation of each village as a distinct market, but does affect our ability to interpret our survey as a complete enumeration of the full market.

Second, since we did not collect detailed facility surveys, we only have information about whether a child attends a private or public preschool but no other characteristics, such as staffing, fees, or instructional practices. Finally, since our data collection focused on foundational math and local language skills, we did not administer tests for English language skills (an important differentiator for private schools).

[11] For instance, ASER reports indicate that student absence ranges from 10% to 45% in different states of India (Pratham, 2022). This absence-induced attrition is non-random: it is typically higher in the public sector and for children with lower test scores and from poorer households.

3 Pre- and primary school choices and value-added

The first part of our analysis focuses on selection into and average productivity differences across different schooling options.

3.1 Selection and educational trajectories

We investigate enrollment patterns and child characteristics by age in Table 1. Virtually all children ($\sim$ 95%) are enrolled in private preschools or public care centers (*anganwadis*) at age 4. At age 5, roughly half of all children begin enrolling in primary school. Between ages 6 and 10, primary school enrollment becomes universal.

Private operators serve a large part of the market in both pre- and primary school. At age 4, a third of all children are enrolled in a private preschool. The market share of private providers reduces to one-quarter in primary school.

There is a significant SES gap in private enrollment. Private school children are about 25% more likely to have mothers with completed secondary education. The average child in private preschool ranks 18 percentiles higher in the socio-economic distribution compared to children in public preschools. In primary school, this gap increases to around 21 percentiles. Additionally, there is a clear gender gap in private enrollment at the primary school level, amounting to 7 percentage points in favor of boys.

We also document a substantial gap in test scores, as measured in our baseline assessments, between private and public school children. At age 4, this gap amounts to 0.18 standard deviations (σ) in the test score distribution in math, and 0.28σ in Tamil. At age 5, the gap increases to around 0.45σ in both subjects. During the main primary school ages (6–10), the gap in Tamil reduces to 0.13σ, but remains large in math (0.3σ).

In the remainder of this paper, we will conduct our analyses separately for children aged 4, 5, and 6–10. For children aged 4 and 6–10, there is an almost perfect overlap with the schooling levels of interest: pre- and primary school. Age 5, in turn, corresponds to the transition period between these two stages of the educational system.

Table 1: Child and household characteristics by age and enrollment status

	No	Preschool			Primary school		
	school	Public	Private	Diff.	Public	Private	Diff.
Panel A: Age 4							
Female	0.55	0.48	0.45	-0.03	0.43	0.65	0.21
Mother educ.: $<$ Gr. 9	0.26	0.24	0.16	-0.08***	0.27	0.12	-0.15
Mother educ.: $\geq$ Gr. 12	0.29	0.35	0.60	0.25***	0.46	0.53	0.07
SES percentile	49.11	47.20	65.60	18.40***	55.78	66.18	10.39
Math IRT score in 2022	-1.33	-1.27	-1.10	0.18***	-0.77	-0.65	0.11
Tamil IRT score in 2022	-1.80	-1.56	-1.28	0.28***	-0.60	-0.81	-0.21
Share of students	0.04	0.61	0.33		0.02	0.01	
Observations	87	1349	726		37	17	
Panel B: Age 5							
Female	0.46	0.51	0.46	-0.05	0.51	0.50	-0.01
Mother educ.: $<$ Gr. 9	0.28	0.23	0.16	-0.07***	0.25	0.15	-0.10***
Mother educ.: $\geq$ Gr. 12	0.40	0.34	0.57	0.24***	0.31	0.56	0.25***
SES percentile	43.98	45.25	67.74	22.49***	43.87	66.58	22.71***
Math IRT score in 2022	-0.98	-1.05	-0.65	0.40***	-0.82	-0.36	0.46***
Tamil IRT score in 2022	-1.19	-1.14	-0.71	0.43***	-0.86	-0.38	0.48***
Share of students	0.02	0.24	0.20		0.39	0.15	
Observations	50	683	592		1132	434	
Panel C: Ages 6–10							
Female	0.44	0.40	0.34	-0.06	0.51	0.44	-0.07***
Mother educ.: $<$ Gr. 9	0.26	0.00	0.12	0.12	0.31	0.15	-0.16***
Mother educ.: $\geq$ Gr. 12	0.21	0.20	0.52	0.32**	0.24	0.51	0.27***
SES percentile	46.87	44.47	68.45	23.98***	42.66	63.25	20.60***
Math IRT score in 2022	0.29	-0.29	-0.44	-0.15	0.61	0.91	0.30***
Tamil IRT score in 2022	0.15	-0.21	-0.50	-0.28	0.68	0.81	0.13***
Share of students	0.00	0.00	0.00		0.73	0.26	
Observations	39	15	67		10465	3797	

Notes: This table reports average differences in child and household characteristics by type of enrollment, separately by age, in three panels. The types of enrollment are no school and private/public pre-/primary school. Columns 4 and 7 show the difference between children in the private and public sectors, respectively. Virtually all children attend preschool at age 4. At age 5, children start transitioning into primary school. Between 6 and 10, virtually all are enrolled in primary school.

3.2 Estimating value-added by sector

We rely on conventional value-added models to measure test score improvements from attending a private pre- and primary school. Specifically, test scores are regressed on school characteristics (e.g., private/public indicators) while conditioning on lagged scores and student socio-demographic characteristics to account for student selection (see, e.g., Todd & Wolpin (2003, 2007)).

For each subject s (math and Tamil), we estimate the following equation:

$$y_{isv}^{2023} = \lambda y_{isv}^{2022} + \beta Private_s + \Gamma X_{isv} + \epsilon_{isv}, \tag{1}$$

where i denotes a child, s a school, and v a village. The variable y_{isv}^{t} denotes student i's test score in a particular subject in year t, λ captures the effect of lagged test scores, and $Private_i$ is an indicator for whether student i attended a private preschool or primary school between the assessment waves. X_{isv} is a vector of additional controls, including village fixed effects, deciles of the SES wealth index, paternal and maternal education levels, and the child's gender.[12] The coefficient of interest is β, which captures the effect of attending a private preschool/school on test scores. We estimate this equation separately for children aged 4, 5, and 6–10, but the results are very similar if we pool all children and allow β to vary by pre- or primary school enrollment.

This analysis generates unbiased value-added estimates as long as our controls are rich enough to account for the selection of children into private vs. public operators. Similar approaches to estimation appear to produce valid value-added measures when compared to estimates using identification from design-based experimental or quasi-experimental variation in both developed and developing country settings (J. D. Angrist et al., 2017; J. Angrist et al., 2023; Andrabi et al., 2011, 2022; Singh, 2015, 2020). Although we do not have similar design-based identification to validate our value-added estimates, in a similar approach to Chetty et al. (2014), we show that our estimates are invariant to including a large set of additional controls. Thus, although the nature of selection may differ across settings, we think it is plausible that the value-added estimates in our data reflect true productivity differences rather than selection effects.

3.3 The private premium in preschool and primary school

Table 2 reports the estimated private premium (β) from Equation (1), separately for children aged 4 (preschool-age), 5 (transition), and 6–10 (primary school). Columns 1, 3, and 5 report differences in test scores by private school attendance conditional only on village fixed effects; Columns 2, 4, and 6 further conditions on lagged scores and covariates as specified in Equation (1).

[12] We control for lagged achievement using a linear control with a common persistence parameter (λ) in our benchmark specification for ease of exposition. We show robustness exercises in Tables A.2 and A.3, which include quadratic polynomials in lagged scores in both subjects as well as a battery of additional controls.

Table 2: Private school value-added in preschool and primary school

	Age 4		Age 5		Ages 6–10	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Math						
Private school	0.803***	0.728***	0.322***	0.179***	0.127***	-0.002
	(0.048)	(0.052)	(0.045)	(0.049)	(0.021)	(0.019)
Math IRT score in 2022		0.173***		0.239***		0.303***
		(0.029)		(0.026)		(0.010)
Constant	-0.567***	-0.336***	0.335***	0.570***	1.426***	1.253***
	(0.017)	(0.039)	(0.016)	(0.029)	(0.006)	(0.008)
Panel B: Tamil						
Private school	0.635***	0.571***	0.167***	0.072	-0.090***	-0.166***
	(0.053)	(0.056)	(0.043)	(0.046)	(0.023)	(0.023)
Tamil IRT score in 2022		0.157***		0.191***		0.337***
		(0.022)		(0.022)		(0.011)
Constant	-0.536***	-0.292***	0.340***	0.531***	1.655***	1.438***
	(0.019)	(0.039)	(0.016)	(0.028)	(0.006)	(0.010)
Controls	Village FE	All	Village FE	All	Village FE	All
Observations	1,837	1,837	2,840	2,840	14,344	14,344

Notes: Robust standard errors, clustered at the village level, in parentheses. Columns 1, 3 and 5 show raw means by private school attendance within villages. Columns 2, 4 and 6 include lagged scores, village fixed effects, and controls for deciles of the SES wealth index, paternal and maternal education, and child gender. Test scores refer to equated IRT scores, standardized with respect to children aged 5 in the 2022 assessments. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Raw differences in academic achievement by private school attendance are substantial at the preschool level, amounting to 0.8 standard deviations of the baseline test score distribution (σ) in math and 0.64σ in Tamil (Column 1). Most of these gaps reflect stark productivity differences between sectors: conditional on lagged test scores and socioeconomic characteristics, the average private premium in the preschool market is 0.73σ in math and 0.57σ in Tamil (Column 2).[13] This is equivalent to almost *twice* the raw difference in Tamil achievement between children aged 3 and 4 in public preschools in our endline assessments, and *four times* the difference in math.[14]

[13]Since children aged 5 are a mix of pre- and primary school students, their private premium is positive but muted compared to their younger peers. If we focus on children aged 5 — of which around half will have started primary school — and allow private school effects to differ at the pre- and primary level, we obtain very similar estimates as for children aged 4 and 6–10 (see Table A.4).

[14]In Tables A.5 and A.6, we divide items by competencies being assessed. In math, private pre-schools

These patterns differ substantially at the primary school level (Columns 5 and 6). The private premium in math is virtually zero (Column 6), which is lower but largely in line with previous studies on private primary schools in India (Muralidharan & Sundararaman, 2015; Singh, 2015). In Tamil, the "premium" at the primary level is negative, which likely reflects a greater focus on English teaching in private schools (Muralidharan & Sundararaman, 2015; Singh, 2015).

This analysis identifies the difference in productivity between public and private options; determining their absolute levels would require comparing them to no enrollment. The latter margin is less relevant in Tamil Nadu, since preschool enrollment is near universal, but remains relevant in many Indian states (such as Uttar Pradesh, where 33% of children enroll in preschool). In our sample, only $\sim$4% of 4-year-olds (N=67) are not enrolled in any level. Estimating the value-added of both public and private preschool facilities at age 4, relative to the baseline of no enrollment, suggests attending public care centers leads to learning gains of 0.27σ and 0.31σ in math and Tamil, respectively — roughly a quarter to a third of private preschool value-added (Table A.7).

4 Understanding preschool and primary school markets

The previous section estimated the *average* productivity differential across sectors and their market shares. Yet, these averages likely conceal substantial heterogeneity across markets. Further, the productivity and market shares of public and government preschools/schools are likely to interact within markets; reducing productivity differentials to a sample-wide average restricts us from investigating such relationships. In this section, we advance our understanding of these market-level associations.

4.1 Estimating village-level value-added

Our empirical approach to estimating village- and sector-specific school productivity extends the value-added framework described in Section 3.2. We estimate value-added for each level-sector-village cell, where level refers to pre-/primary schooling and sector to public/private operators. We define θ_{slv} as a set of dummy variables that indicate attendance at a private or public option s at the pre- or primary level l in village v.

increase the proportion correct on test items by 18–33 percentage points, relative to public school averages of 19–58%. In Tamil, this figure is 9–22 percentage points, relative to public school averages of 36–73%. These effects are largest, both in absolute and relative terms, in competencies where public preschool children are particularly weak.

To improve precision, we pool all children aged 4–10 (adding subscript a for age) in the same regression and estimate the following equation:

$$y_{iaslv}^{2023} = \lambda_a y_{iaslv}^{2022} + \theta_{slv} + \Gamma X_{iaslv} + \epsilon_{iaslv}. \tag{2}$$

The coefficient on the lagged score λ_a is allowed to differ by child age a. The vector X_{iaslv} contains controls for deciles of the SES wealth index, paternal and maternal education, as well as the age and gender of the child. The public preschool sector of one of the villages is left out as the omitted category. This procedure gives us four measures of value-added for each village — the average value-added in private and public options at the preschool and primary levels — which serve as the basis for this analysis. We define the private premium at pre-/primary school level l in village v as

$$\Theta_{lv} = \theta_{1lv} - \theta_{0lv}, \tag{3}$$

which is simply the difference in value-added between private and public options at level l and village v.

By pooling children aged 4–10 in the same specification, we assume that (age-varying) lagged scores and background characteristics can address selection into early primary schooling: as previously shown, around half of all children are enrolled in primary school at age 5 (Table 1). The alternative strategy of omitting 5-year-olds, to estimate value-added at the pre- and primary level only for ages when there is no differential enrollment, yields similar results.

Finally, since value-added estimates in θ_{slv} will be measured with uncertainty, we shrink the estimates toward their level-sector averages using an Empirical Bayes approach (for details, see Appendix D).[15] Unless stated otherwise, we always report results using these Empirical Bayes estimates.

4.2 Value-added across markets, sectors and levels

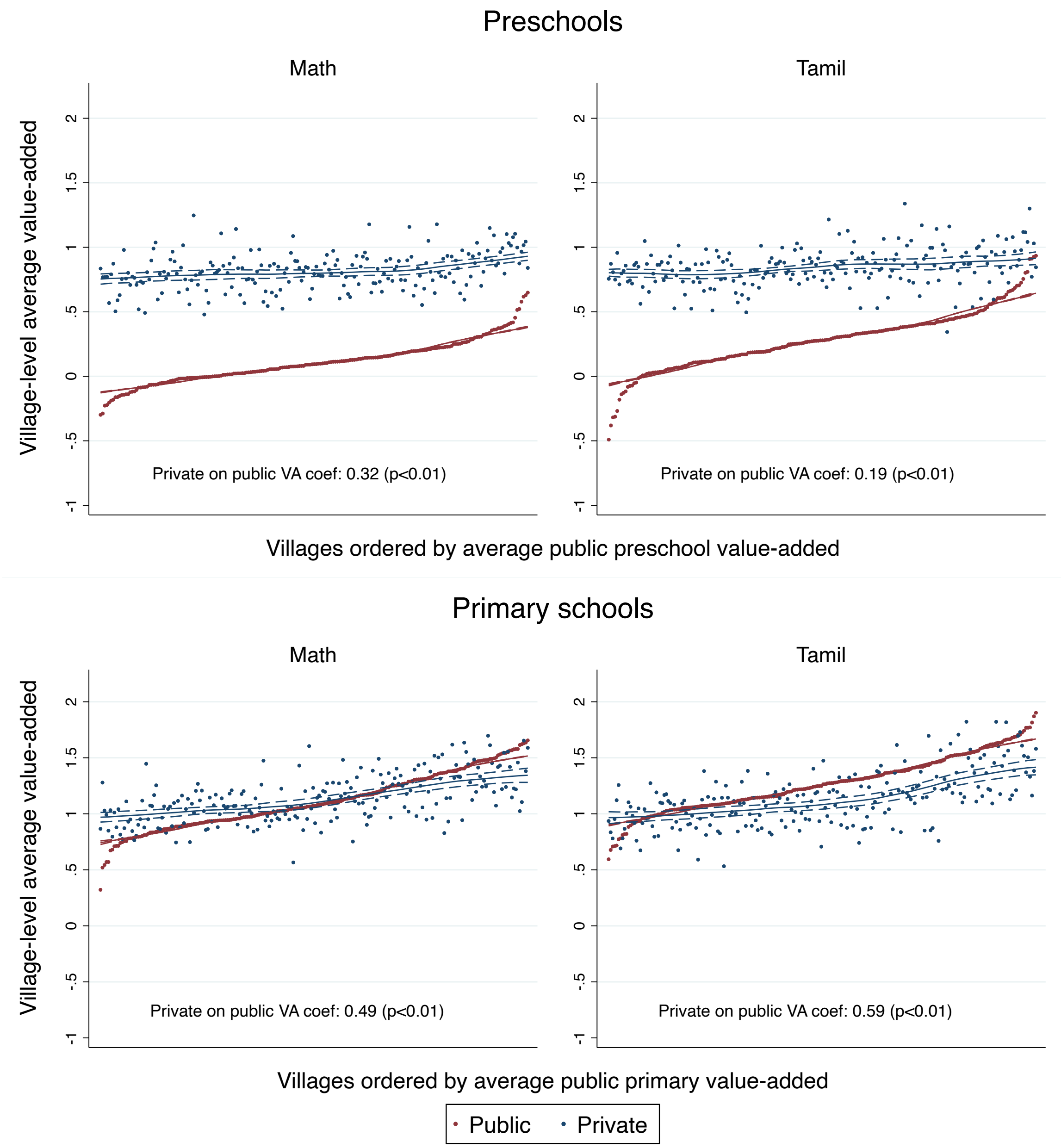
We now present the results of estimating the specification from Equation (2), which gives us four measures of value-added for each village — the average value-added in private and public options at the preschool and primary levels, shrunk by the Empirical Bayes approach.[16] We focus on correlations of and differences in value-added

[15]The Empirical Bayes estimates are linear combinations of raw value-added estimates and their sector-level averages. The weights in these linear combinations apply more 'shrinkage' towards the sector average in cases where value-added estimates are relatively more imprecise (J. Angrist et al., 2023).

[16]Figure A.1 presents density plots of these estimates. Village-sector clusters are generally smaller at the preschool level, leading the Empirical Bayes approach to shrink estimates more heavily toward the common

along three dimensions: across individual markets, comparing private and public sectors, and between pre- and primary school levels.

Figure 1: Village-level average value-added of private and public options



Notes: These figures show village-level average school value-added by sector (public/private) and level (preschool/primary) using Empirical Bayes measures as described in Appendix D. Villages are ordered along the x-axis by their average value-added in government schools. The regression specification generating these estimates is given by Equation 2.

Figure 1 orders villages by their average value-added in the public sector along the

mean. As a result, the dispersions of the pre- and primary school value-added distributions are not directly comparable.

horizontal axis. Red (blue) dots denote each village's average value-added in the public (private) sector. The top panels show preschool results, separately by subject, and the bottom panels show primary school results.

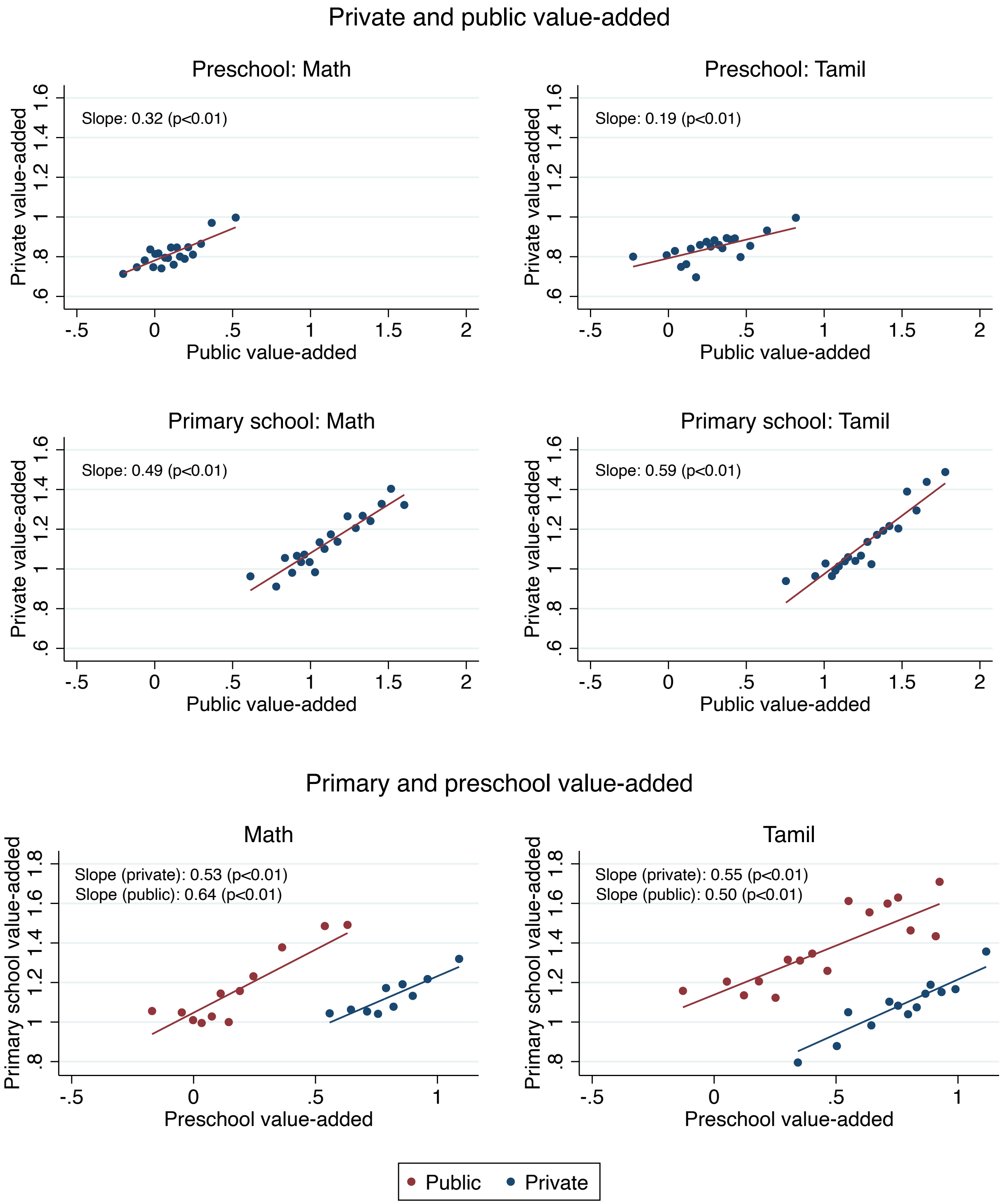
No village has a public preschool sector that, on average, performs better than its private sector in math — with few exceptions, the same is true for Tamil. The average private premium is 0.71σ in math and 0.56σ in Tamil, which is very similar to the results in Section 3.3 (see Table A.8 for further summary statistics). While productivity differences vary across markets, they are substantial almost everywhere. This pattern is very different at the primary school level. There is no gap between private and public sectors in average productivity in math, and a negative private premium in Tamil.

We identify a positive correlation between private and public sector value-added. Regressing private on public preschool value-added yields a coefficient of 0.32 in math and 0.19 in Tamil (the top panel of Figure 2 shows binned scatter plots of these correlations). The correlation between private and public sector value-added is substantially higher in primary school, at 0.49 in math and 0.59 in Tamil. The differences across pre- and primary school correlations are statistically significant ($p = 0.02$ in math, $p < 0.01$ in Tamil). This association in productivity across sectors is consistent with potential "multiplier effects" in which, due to market-level incentives and competition, an increase in public sector quality also leads to improvements in the private sector; Andrabi et al. (2024) present experimental evidence of such a mechanism at work in primary schooling markets in Pakistan, although we are not aware of similar evidence at the preschool level. In short, villages with particularly low-quality public schools also tend to have weaker-performing *private* schools than other villages, and this pattern is stronger in primary relative to preschool markets.

We also investigate the within-sector correlation of value-added between pre- and primary schools within the same markets (bottom panel of Figure 2). In both the public and private sectors, productivity is clearly correlated across levels of schooling. On average, an increase of one standard deviation of *pre-school* value-added predicts an increase of roughly half a standard deviation of *primary school* value-added. This correlation is not surprising in the private sector since private preschools are often vertically integrated with private schools. Government preschools, on the other hand, are managed by a parallel administrative set-up separate from the School Education Department but appear to display similar correlations.[17] As such, productivity differences across markets appear to be relatively persistent throughout early childhood and adolescence.

[17] Public preschools are run by the Ministry of Women and Child Welfare at the national level, not the Ministry of Education. Staffing, pay, management, and overall capacity all differ between these two structures.

Figure 2: Correlations of village value-added across sectors and levels



Notes: These figures show binned scatter plots of village-level Empirical Bayes value-added estimates. In the top panel, we show the correlation between private and public sector value-added, separately by subject and level (preschool/primary school). In the bottom panel, we show correlations between pre- and primary school value-added, separately by subject and sector. Slopes and p-values associated with a test of zero slopes are shown in the figures; robust standard errors are used for inference.

Turning to market shares at the preschool level, we find that private preschool enrollment does *not* increase with the size of the private premium (see Figure A.2). At the primary school level, in contrast, we *do* find that market shares reflect differences in the private premium. In math, a one standard deviation increase in the village-level private premium is associated with a 15 p.p. increase in private enrollment. In Tamil, the correlation is essentially flat, which is not unsurprising given that many households opt for private primary schools precisely because of their focus on English rather than the local language.

However, market shares are equilibrium outcomes. Thus, the lack of association between private preschool premia and market shares is consistent with distinct explanations. For example, it is possible that (i) households do not value cognitive skill production for very young children and/or, (ii), higher-quality preschools also charge higher prices (which we do not measure), or (iii) higher-quality preschools want to keep enrollment low to maintain quality. Since market shares do seem to respond at the primary level with value-added in math (which is emphasized in both public and government schools), explanations for this pattern are likely to be specific to preschools.

We provide further analyses of the correlates of village-level value-added in the Appendix, which reveal three additional findings. First, value-added is highly correlated across subjects (Figure A.3): within a village, private/public preschools providing high value-added in math also tend to do so in Tamil. Second, private premia across markets are largely uncorrelated with market size, measured as the number of children aged 4–6 in each village, and village-level SES at any level of schooling (Table A.9). If anything, villages with weaker socio-economic composition tend to have larger private premia compared to those with stronger composition.

5 Socioeconomic learning gaps

There are substantial productivity differences across private and public providers of pre-primary education, but not at the primary school level. Further, there is a large SES gap in both (i) take-up of private education and (ii) test scores (Section 3.1). In this section, we investigate the extent to which differential enrollment in (effective) private options may explain these test score gaps across the SES distribution.

We do this by regressing the 2023 test scores on a dummy indicating whether the child belongs to the top vs. bottom 25% of the SES distribution (omitting those in between), successively conditioning on lagged scores as well as an indicator for attending a private option. As in our previous analyses, we estimate this separately for 4-, 5-, and 6 to 10-year-olds.

Table 3: Decomposition of SES gap (top/bottom 25%) in 2023 test scores, preschool and primary level

	Age 4			Age 5			Ages 6–10		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Math									
Top 25% SES	0.439***	0.415***	0.178*	0.304***	0.253***	0.144*	0.235***	0.162***	0.153***
	(0.072)	(0.072)	(0.082)	(0.056)	(0.055)	(0.065)	(0.029)	(0.026)	(0.025)
Math IRT score in 2022		0.172***	0.155***		0.248***	0.226***		0.312***	0.311***
		(0.044)	(0.042)		(0.033)	(0.034)		(0.013)	(0.013)
Private school			0.658***			0.251**			0.024
			(0.083)			(0.077)			(0.027)
Constant	-0.457***	-0.247***	-0.398***	0.320***	0.526***	0.466***	1.344***	1.165***	1.162***
	(0.040)	(0.071)	(0.072)	(0.030)	(0.040)	(0.044)	(0.012)	(0.014)	(0.014)
Panel B: Tamil									
Top 25% SES	0.346***	0.301***	0.111	0.132*	0.077	0.037	0.105***	0.056*	0.106***
	(0.071)	(0.070)	(0.074)	(0.059)	(0.059)	(0.065)	(0.031)	(0.028)	(0.030)
Tamil IRT score in 2022		0.161***	0.153***		0.196***	0.190***		0.350***	0.352***
		(0.034)	(0.033)		(0.032)	(0.034)		(0.015)	(0.015)
Private school			0.529***			0.092			-0.139***
			(0.084)			(0.071)			(0.031)
Constant	-0.458***	-0.211**	-0.329***	0.361***	0.541***	0.520***	1.559***	1.341***	1.358***
	(0.040)	(0.063)	(0.069)	(0.031)	(0.044)	(0.050)	(0.013)	(0.014)	(0.015)
Observations	912	912	912	1,360	1,360	1,360	6,872	6,872	6,872

Notes: Robust standard errors, clustered at the village level, in parentheses. Village fixed effects and controls for child gender are included in all regressions. Test scores refer to the IRT EAP scores, standardized with respect to children aged 5 in the 2022 assessments. The SES index is based on questions regarding the availability of household amenities and computed with PCA. The omitted category contains students in households with an SES index below the 25th percentile. Households with an SES index between the 25th and 75th percentiles are excluded from the regressions. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table 3 presents the results. At age 4, there is a noticeable test score gap between children from high- and low-SES backgrounds in math and Tamil (0.44σ and 0.35σ, respectively). Differences in baseline ability explain only around 10% of this gap (Columns 1 and 2), but approximately 60% of it can be attributed to private preschool attendance (Columns 2 and 3). When conditioning on private preschool enrollment and lagged test scores, the remainder of the SES gap in test scores is relatively stable across age groups.

The patterns are different for primary school children. At ages 6–10, private enrollment explains little of the SES gap in math and widens it in Tamil, which aligns with our previous findings on the primary private premium (Columns 8 and 9). Differences in baseline ability — i.e., a child's academic achievement when entering primary school — account for around 40% of the test score gap in both subjects (Columns 7 and 8).

Overall, these findings suggest that high-SES children enter primary school with substantially stronger academic achievement compared to their low-SES peers, which then persist at later stages. These early gaps are primarily driven by differences in private preschool enrollment.

6 Conclusion

We provide four new facts to the study of education systems in LMICs. In our setting, (i) private institutions have a substantial premium in test score value-added over government options at the preschool stage but not in primary schooling, (ii) this premium at preschool level exists in nearly all villages in our sample, (iii) it accounts for nearly two-thirds of the socioeconomic test score gap at school entry age and (iv) within-village correlations between the productivity of sectors, both at preschool and primary levels, is suggestive of a multiplier effect, wherein higher public sector productivity also spills over to the private sector.

These results connect to important policy goals. Providing *quality* preschool to all children is enshrined in both international and national policy goals — such as the Sustainable Development Goal 4.2 and India's new National Education Policy 2020, which makes quality preschool an essential component for achieving universal foundational skills in literacy and numeracy by the end of Grade 3 (Muralidharan & Singh, 2021). Yet, there are few estimates of the productivity of preschools using broad samples in low- and middle-income countries, or attempts to quantify the heterogeneity in these effects. Our results, providing new facts in this area, suggest that improving the productivity of government preschools could be substantially productive both in raising skill levels and reducing socioeconomic inequality in learning. Investments to spur such improvements could be substantially productive (see, e.g., Ganimian et al. (2024)).

Our results also highlight the importance of considering preschool markets in both research and policy formulation. Educational markets for children between 3–5 years appear as diverse in the range of providers, their productivity, the dispersion in productivity across markets, and the potential links between sectors, as education markets at primary or secondary schooling age. Yet, in contrast to the substantial literature on similar themes in primary and secondary school, studies of parental information, provider competition, product differentiation, or policy tools such as vouchers or admissions reforms are nascent at best for the preschool stage.[18] Further research on these themes is likely to be substantially rewarding.

[18] For example, understanding market structures is key to designing effective voucher policies. As Dean & Jayachandran (2019) note, the average effect of voucher schemes will depend substantially on whether they induce substitution from home care, from public preschools, or within the private sector. See Dean & Jayachandran (2019) and Bjorvatn et al. (2024) for two recent voucher schemes that subsidize preschool attendance with positive treatment effects.

References

Allende, C. (2019). Competition under social interactions and the design of education policies. *Working Paper*.

Andrabi, T., Bau, N., Das, J., Karachiwalla, N., & Ijaz Khwaja, A. (2024, 05). Crowding in Private Quality: The Equilibrium Effects of Public Spending in Education. *The Quarterly Journal of Economics*, qjae014. Retrieved from https://doi.org/10.1093/qje/qjae014 doi: 10.1093/qje/qjae014

Andrabi, T., Bau, N., Das, J., & Khwaja, A. I. (2022). *Heterogeneity in school value-added and the private premium* (Working Paper No. 30627). National Bureau of Economic Research.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, *107*(6), 1535-63.

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, *3*(3), 29-54.

Andrew, A., Attanasio, O. P., Bernal, R., Sosa, L. C., Krutikova, S., & Rubio-Codina, M. (2024). Preschool quality and child development. *Journal of Political Economy*, *132*(7), 2304-2345. Retrieved from https://doi.org/10.1086/728744 doi: 10.1086/728744

Angrist, J., Hull, P., & Walters, C. (2023). Chapter 1 – Methods for measuring school effectiveness. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 7, p. 1-60). Elsevier.

Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, *132*(2), 871–919.

Attanasio, O., Fernández, C., Fitzsimons, E. O. A., Grantham-McGregor, S. M., Meghir, C., & Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *BMJ*, *349*.

Attanasio, O., Paes de Barros, R., Carneiro, P., Evans, D. K., Lima, L., Olinto, P., & Schady, N. (2022). *Public childcare, labor market outcomes of caregivers, and child development: Experimental evidence from Brazil* (Working Paper No. 30653). National Bureau of Economic Research.

Bau, N. (2022). Estimating an equilibrium model of horizontal competition in education. *Journal of Political Economy*, *130*(7), 1717–1764.

Behrman, J. R., Fernald, L., & Engle, P. (2013). Preschool programs in developing countries. *Education policy in developing countries*, 65–105.

Bjorvatn, K., Ferris, D., Gulesci, S., Nasgowitz, A., Somville, V., & Vandewalle, L. (2024). Childcare, labor supply, and business development: Experimental evidence from Uganda. *American Economic Journal: Applied Economics*.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... others (2017). Early childhood development coming of age: science through the life course. *The Lancet*, *389*(10064), 77–90.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American economic review*, *104*(9), 2593–2632.

Crawfurd, L., Hares, S., & Todd, R. (2024). The impact of private schools, school chains and PPPs in developing countries. *The World Bank Research Observer*, *39*(1), 97–123.

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*, *2*(92), 175–187.

Dean, J. T., & Jayachandran, S. (2019). *Attending kindergarten improves cognitive but not socioemotional development in India* (Working Paper).

Elango, S., Garcia, J. L., Heckman, J., & Hojman, A. (2015). Early childhood education. In *Economics of means-tested transfer programs in the United States, Volume 2* (p. 235-297). National Bureau of Economic Research.

Engle, P. L., Black, M. M., Behrman, J. R., De Mello, M. C., 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.

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., ... others (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353.

Fernald, L. C., Kariger, P., Hidrobo, M., & Gertler, P. J. (2012). Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences*, *109*, 17273–17280.

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 Jr, 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 Jr, R. G., & Levitt, S. D. (2013). Testing for racial differences in the mental ability of young children. *American Economic Review*, *103*(2), 981–1005.

Ganimian, A. J., Muralidharan, K., & Walters, C. R. (2024). Augmenting state capacity for child development: Experimental evidence from India. *Journal of Political Economy*. doi: 10.1086/728109

Government of India. (2019). *Unified district information system for education plus (UDISE+)*. Ministry of Education, Government of India, New Delhi.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*(1), 114-128.

Holla, A., Bendini, M., Dinarte, L., & Trako, I. (2021). Is investment in preprimary education too low? Lessons from (quasi) experimental evidence across countries. *World Bank Working Paper 9723*.

Muralidharan, K., & Singh, A. (2021). India's new national education policy: Evidence and challenges. *Science*, *372*(6537), 36–38.

Muralidharan, K., & 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. A. (2021). Targeted vouchers, competition among schools, and the academic achievement of poor students. *Working Paper*.

Nielsen, E. (2023). The income-achievement gap and adult outcome inequality. *Journal of Human Resources*.

Pratham. (2022). *Annual status of education report 2022*. Pratham, New Delhi.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: new evidence and possible explanations. *Whither opportunity*, *1*(1), 91–116.

Reardon, S. F. (2021). The economic achievement gap in the US, 1960-2020: Reconciling recent empirical findings. CEPA Working Paper No. 21-09. *Stanford Center for Education Policy Analysis*.

Reynolds, A. J., Hayakawa, M., Ou, S.-R., Mondi, C. F., Englund, M. M., Candee, A. J., & Smerillo, N. E. (2017). Scaling and sustaining effective early childhood programs through school–family–university collaboration. *Child development*, *88*(5), 1453–1465.

Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., & Grantham-McGregor, S. (2015). The socioeconomic gradient of child development: Cross-sectional evidence from children 6–42 months in Bogota. *The Journal of Human Resources*, *50*(2), 464–483.

Schady, N., Behrman, J., Araujo, M. C., Azuero, R., Bernal, R., Bravo, D., ... Vakis, R. (2015). Wealth gradients in early childhood cognitive development in five Latin American countries. *Journal of Human Resources*, *50*(2), 446–463.

Singh, A. (2014). Test score gaps between private and government sector students at school entry age in India. *Oxford Review of Education*, *40*(1), 30–49.

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

Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, *18*(4), 1770–1813.

Singh, A., & Romero, M. (2022). *Preschool services in Tamil Nadu.* RISE Programme Blog. RISE. Retrieved 2023-03-03, from https://riseprogramme.org/blog/preschool_services_Tamil_Nadu

Singh, A., Romero, M., & Muralidharan, K. (2024). Covid-19 learning loss and recovery: Panel data evidence from India. *Journal of Human Resources*, *Forthcoming*.

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

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.

UNESCO. (2022). *Distribution of enrollment by type of institution.* (data retrieved from UNESCO Institute for Statistics on March 18, 2024, http://data.uis.unesco.org/)

World Bank. (2017). *World development report 2018: Learning to realize education's promise (overview)*. The World Bank, Washington D.C.

World Bank. (2019). *School enrollment, primary (% net).* (data retrieved from World Development Indicators on April 23, 2023, https://data.worldbank.org/indicator/SE.PRM.NENR)

World Bank. (2020). *School enrollment, preprimary (% gross).* (data retrieved from World Development Indicators on March 18, 2024, https://data.worldbank.org/indicator/SE.PRE.ENRR)

A Further tables and figures

Table A.1: Comparing baseline sample to NFHS

	NFHS-V	Baseline	Difference
Panel A: Assets and household characteristics			
Internet	0.59	0.48	-0.11***
	(0.49)	(0.50)	p=0.00
Washing machine	0.14	0.09	-0.06***
	(0.35)	(0.28)	p=0.00
Fridge	0.56	0.46	-0.10***
	(0.50)	(0.50)	p=0.00
Computer	0.09	0.08	-0.01
	(0.28)	(0.27)	p=0.20
Television	0.94	0.93	-0.02***
	(0.23)	(0.26)	p=0.01
Fan	0.97	0.97	-0.00
	(0.16)	(0.17)	p=0.83
Electricity	0.99	0.94	-0.05***
	(0.08)	(0.23)	p=0.00
Car	0.05	0.05	-0.01
	(0.22)	(0.21)	p=0.34
Tractor	0.02	0.03	0.00
	(0.15)	(0.16)	p=0.35
Bike	0.77	0.75	-0.02
	(0.42)	(0.43)	p=0.14
Bicycle	0.46	0.36	-0.10***
	(0.50)	(0.48)	p=0.00
Number of children (3-10 yrs old)	1.62	1.55	-0.07***
	(0.68)	(0.62)	p=0.00
Scheduled caste	0.36	0.33	-0.03
	(0.48)	(0.47)	p=0.16
Owns land	0.31	0.25	-0.06***
	(0.46)	(0.43)	p=0.00
Observations	2,561	17,486	
Panel B: Maternal education			
Mother education: at least some primary	0.96	0.96	-0.00
	(0.20)	(0.21)	p=0.38
Mother education: at least some secondary	0.87	0.93	0.06***
	(0.33)	(0.25)	p=0.00
Observations	2,542	16,280	

Notes: The table presents means and standard deviations for households in Tamil Nadu with children aged 3–10 in the NFHS-V survey (Column 1) and households in our baseline sample in 2022 (Column 2). Column 3 shows differences and statistical significance (clustering standard errors at the sampling cluster level for NFHS-V and the village level in our sample). $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.2: Robustness of the private school premium: children aged 4

	(1)	(2)	(3)	(4)
Panel A: Math				
Private school	0.728***	0.725***	0.713***	0.712***
	(0.052)	(0.052)	(0.051)	(0.052)
Core specification	Yes			
Lagged, squared Tamil & Math scores		Yes		
Survey month 2022×2023 FEs			Yes	
Caste & home inputs in 2022				Yes
Observations	1,837	1,837	1,837	1,800
Panel B: Tamil				
Private school	0.579***	0.571***	0.550***	0.552***
	(0.056)	(0.056)	(0.055)	(0.056)
Core specification	Yes			
Lagged, squared Tamil & Math scores		Yes		
Survey month 2022×2023 FEs			Yes	
Caste & home inputs in 2022				Yes
Observations	1,837	1,837	1,837	1,800

Notes: Robust standard errors, clustered at the village level, in parentheses. This table shows regression of IRT scores in 2023 on having attended a private rather than public preschool or primary school during previous school year. Column 1 reports estimates of this private premium in our core specification (Equation 1). Column 2 adds controls for baseline (2022) scores quadratically in both math and Tamil simultaneously. Column 3 further includes fixed effects for the month in which the baseline and endline surveys took place, and their interactions. Finally, Column 4 adds controls for caste and several educational inputs measured at baseline: whether the child had recently received educational content via 1) internet, 2) TV or 3) books at home. These measures are missing 37 children. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.3: Robustness of the private school premium: children aged 6–10

	(1)	(2)	(3)	(4)
Panel A: Math				
Private school	-0.002	0.014	0.014	0.018
	(0.019)	(0.019)	(0.019)	(0.020)
Core specification	Yes			
Lagged, squared Tamil & Math scores		Yes		
Survey month 2022×2023 FEs			Yes	
Caste & home inputs in 2022				Yes
Observations	14,344	14,344	14,344	13,932
Panel B: Tamil				
Private school	-0.206***	-0.188***	-0.188***	-0.173***
	(0.023)	(0.022)	(0.022)	(0.023)
Core specification	Yes			
Lagged, squared Tamil & Math scores		Yes		
Survey month 2022×2023 FEs			Yes	
Caste & home inputs in 2022				Yes
Observations	14,344	14,344	14,344	13,932

Notes: Robust standard errors, clustered at the village level, in parentheses. This table shows regression of IRT scores in 2023 on having attended a private option rather than public preschool or primary school during previous school year. Column 1 reports estimates of this private premium in our core specification (Equation 1). Column 2 adds controls for baseline (2022) scores quadratically in both math and Tamil. Column 3 further includes fixed effects for the month in which both the baseline and endline surveys took place. Finally, Column 4 adds controls for caste and several educational inputs measured at baseline: whether the child had recently received educational content via 1) internet, 2) TV or 3) books at home. These measures are missing for 412 children. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.4: Private school value-added in preschool and primary school for children aged 5

	Math (1)	Tamil (2)
Private school	0.706***	0.578***
	(0.0592)	(0.0553)
Primary school	1.035***	1.011***
	(0.0488)	(0.0462)
Private school × Primary school	-0.686***	-0.660***
	(0.0680)	(0.0659)
Math IRT score in 2022	0.171***	
	(0.0240)	
Tamil IRT score in 2022		0.134***
		(0.0204)
Constant	-0.139***	-0.155***
	(0.0404)	(0.0393)
Controls	All	All
Observations	2,840	2,840

Notes: Robust standard errors, clustered at the village level, in parentheses. The regressions include children aged 5 only, around half of which are already enrolled in primary school. The coefficient on the private school dummy captures the private premium in preschools. The sum of this coefficient with that of the interaction between private and primary school captures the private premium in primary school. Lagged scores, village fixed effects, and controls for deciles of the SES wealth index, paternal and maternal education, as well as child gender are included in both regressions. Test scores refer to the IRT Expected A Posteriori (EAP) scores, standardized with respect to children aged 5 in the 2022 assessments. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.5: Private school value-added in math competencies

	Dep. var.: Proportion correct on math items							
Panel A: Age 4	*Addition*	*Subtraction*	*Number identification*	*Quantitative comparison*				
Private school	0.248***	0.246***	0.325***	0.186***				
	(0.026)	(0.027)	(0.020)	(0.020)				
Math IRT score in 2022	0.071***	0.078***	0.057***	0.054***				
	(0.013)	(0.015)	(0.011)	(0.010)				
Constant	0.517***	0.411***	0.269***	0.652***				
	(0.019)	(0.021)	(0.016)	(0.015)				
Observations	1,837	1,837	1,837	1,837				
Number of items	2	1	7	5				
Public sector avg.	0.414	0.317	0.188	0.577				
Panel B: Ages 6–10	*Addition*	*Subtraction*	*Number identification*	*Quantitative comparison*	*Multiplic. & division*	*Applied problems*	*Geometry*	*Measurement*
Private school	0.012*	0.010	-0.011	0.021*	0.014	0.052***	-0.028	0.005
	(0.006)	(0.006)	(0.009)	(0.010)	(0.013)	(0.009)	(0.031)	(0.011)
Math IRT score in 2022	0.048***	0.044***	0.057***	-0.042***	0.060***	0.058***	0.050**	0.076***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	(0.015)	(0.005)
Constant	0.756***	0.661***	0.893***	0.698***	0.402***	0.710***	0.311***	0.631***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.007)	(0.003)	(0.022)	(0.006)
Observations	14,344	14,344	5,200	14,344	9,142	14,344	804	9,142
Number of items	12	6	4	6	4	5	3	2
Public sector avg.	0.782	0.684	0.890	0.665	0.457	0.740	0.377	0.700

Notes: Robust standard errors, clustered at the village level, in parentheses. Lagged scores, village fixed effects, and controls for deciles of the SES wealth index, paternal and maternal education, as well as child gender are included in all regressions. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.6: Private school value-added in Tamil competencies

	Dep. var.: Proportion correct on Tamil items					
Panel A: Age 4	Object *recognition*	Letter & word *recognition*	*Oral comprehension*			
Private school	0.106***	0.218***	0.091***			
	(0.013)	(0.020)	(0.018)			
Tamil IRT score in 2022	0.028***	0.043***	0.051***			
	(0.006)	(0.007)	(0.007)			
Constant	0.775***	0.437***	0.584***			
	(0.009)	(0.013)	(0.012)			
Observations	1,837	1,837	1,837			
Number of items	12	7	5			
Public sector avg.	0.728	0.364	0.504			
Panel B: Ages 6–10	Object *recognition*	Letter & word *recognition*	*Oral comprehension*	*Word & sentence comprehension*	*Sentence & story comprehension*	*Spelling*
Private school	0.006	-0.058***	-0.026	-0.003	-0.024*	-0.001
	(0.007)	(0.006)	(0.014)	(0.004)	(0.010)	(0.034)
Tamil IRT score in 2022	0.015***	0.075***	0.047***	0.034***	0.085***	0.049**
	(0.004)	(0.003)	(0.007)	(0.002)	(0.005)	(0.017)
Constant	0.963***	0.800***	0.820***	0.873***	0.416***	0.626***
	(0.002)	(0.002)	(0.005)	(0.002)	(0.006)	(0.030)
Observations	1,836	13,525	1,836	14,344	9,142	804
Number of items	4	14	5	12	16	2
Public sector avg.	0.958	0.846	0.800	0.894	0.504	0.713

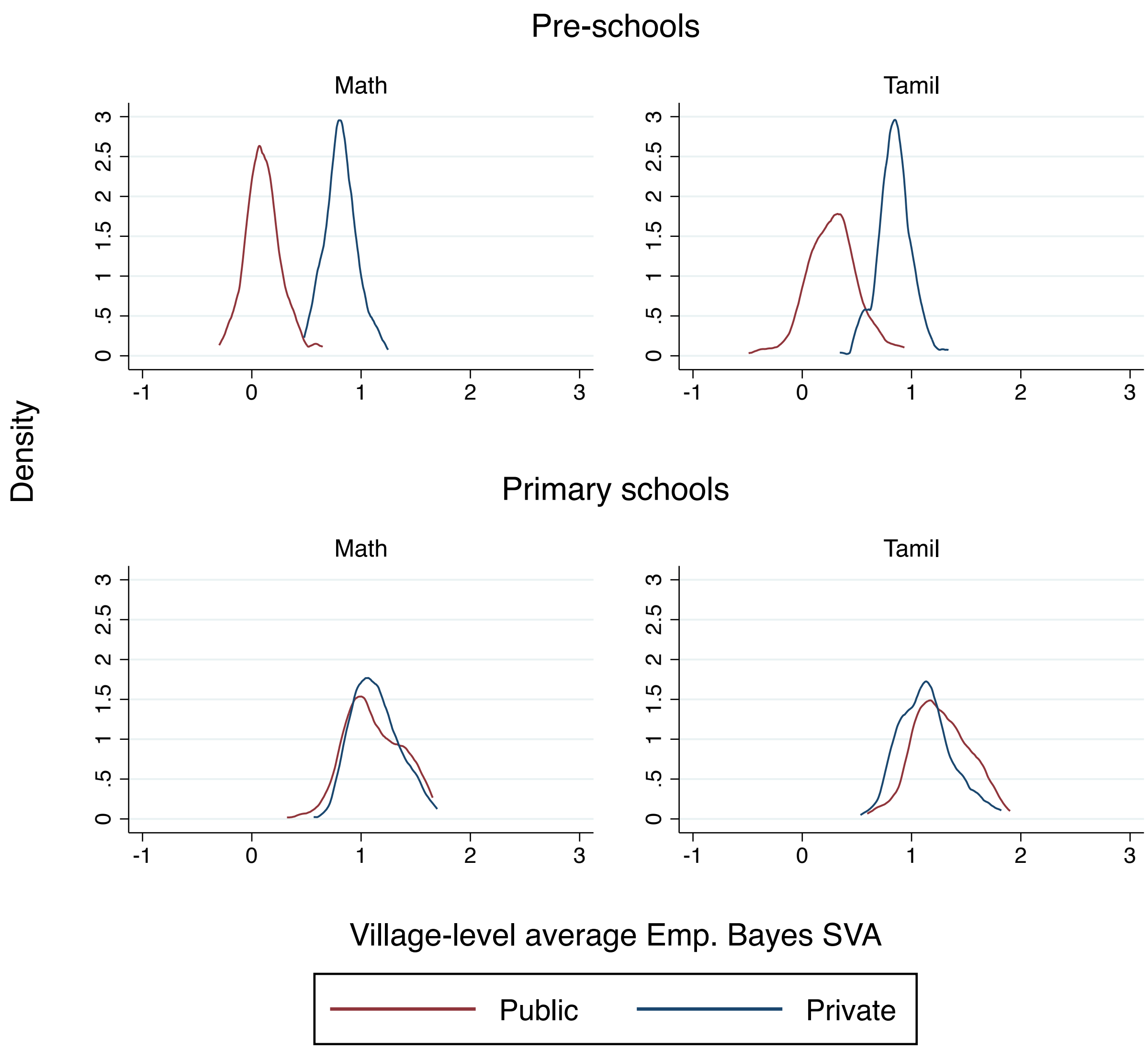
Notes: Robust standard errors, clustered at the village level, in parentheses. Lagged scores, village fixed effects, and controls for deciles of the SES wealth index, paternal and maternal education, as well as child gender are included in all regressions. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Table A.7: Value-added of public/private preschools relative to no enrollment, age 4

	Math		Tamil	
	(1)	(2)	(3)	(4)
Public	0.254*	0.272*	0.312*	0.310*
	(0.124)	(0.123)	(0.153)	(0.148)
Private	1.059***	0.999***	0.949***	0.883***
	(0.127)	(0.128)	(0.153)	(0.150)
Math IRT score in 2022		0.171***		
		(0.0280)		
Tamil IRT score in 2022				0.161***
				(0.0217)
Constant	-0.822***	-0.609***	-0.846***	-0.595***
	(0.119)	(0.127)	(0.146)	(0.147)
Controls	Village FE	All	Village FE	All
Observations	1,904	1,904	1,904	1,904

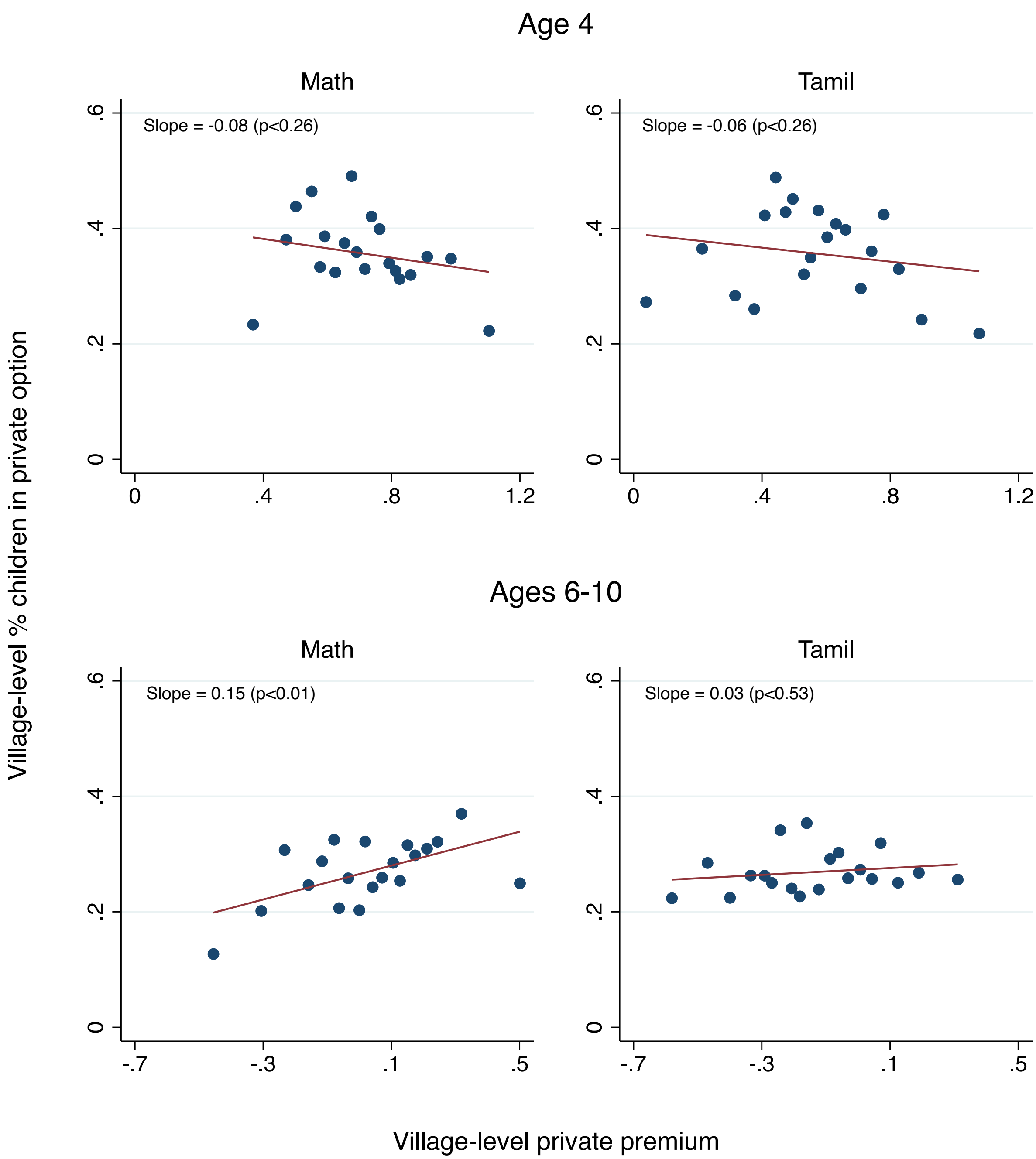
Notes: Robust standard errors, clustered at the village level, in parentheses. The omitted category is children not enrolled in any type of education (67 children). Columns 1 and 3 report raw test score differences by type of school attended, within villages. Columns 2 and 4 include village fixed effects and controls for lagged scores, deciles of the SES wealth index, paternal and maternal education, as well as child gender. Test scores refer to the IRT Expected A Posteriori (EAP) scores, standardized with respect to children aged 5 in the 2022 assessments. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

Figure A.1: Distributions of village-level average value-added



Notes: These figures show kernel density plots of village-level average school value-added by sector (public/private) and level (preschool/primary) using Empirical Bayes measures as described in Section D. These are generated in a regression that includes children aged 4–10, as described in Section 4.1.

Figure A.2: Village-level private premium and share of private enrollment



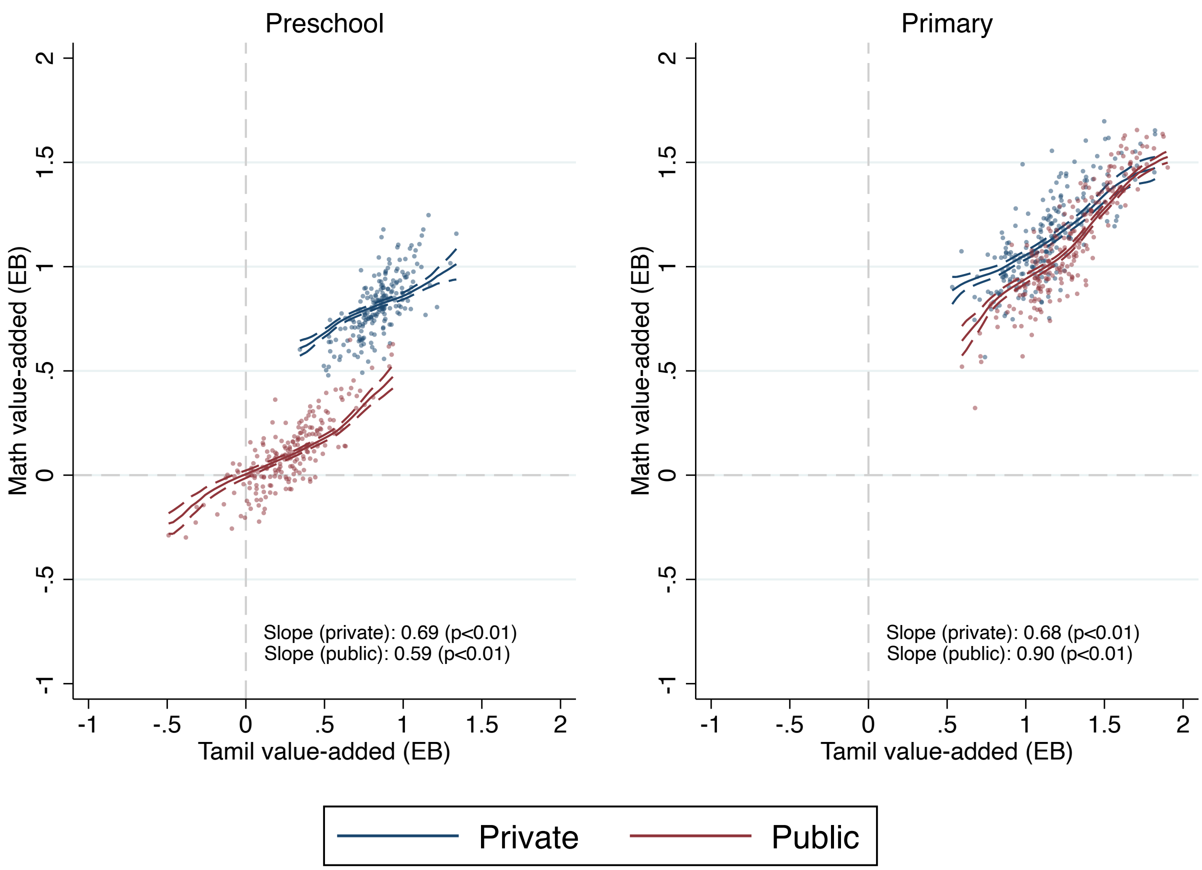
Notes: These figures depict the correlation between (1) the village-level share of children in a private school and (2) the difference between the average *private* and *public* school value-added in the same village (i.e., the private premium). These value-added measures are Empirical Bayes estimates, estimated as described in Appendix D.

Table A.8: Summary statistics of village-level private premia

	Mean	SD	Min	Max	Villages
Private premium: math, pre-school	0.71	0.18	0.19	1.26	203
Private premium: Tamil, pre-school	0.56	0.24	-0.14	1.25	203
Private premium: math, primary	0.02	0.22	-0.59	0.76	216
Private premium: Tamil, primary	-0.14	0.22	-0.71	0.48	216

Notes: This table presents summary statistics of village-level differences in average private vs. public school (Empirical Bayes) value-added measures, separately by subject and preschool/primary level. For details on the estimation of these measures, see Section 4.1.

Figure A.3: Cross-subject correlations of village-level value-added



Notes: These figures show local polynomial fits of village-level value-added in math vs. Tamil at the preschool/primary school level, respectively. These value-added measures are Empirical Bayes estimates, estimated as described in Appendix D.

Table A.9: Regressions of private premia on market size and village average SES

	Preschool private premium		Primary school private premium	
	Math	Tamil	Math	Tamil
Number of children in market (std)	0.001	0.002	-0.008	-0.006
	(0.013)	(0.018)	(0.016)	(0.014)
Village average SES (std)	-0.023	-0.047*	0.015	-0.023
	(0.014)	(0.020)	(0.015)	(0.018)
Constant	0.709***	0.565***	0.024	-0.136***
	(0.012)	(0.017)	(0.015)	(0.015)
Observations	203	203	216	216

Notes: Robust standard errors in parentheses. This table shows village-level regressions of private premia, across sectors and levels, on the standardized number of children aged 6–10 and the standardized average SES percentile in the village. One standard deviation of number of children corresponds to 45 children; for village-level average SES, it corresponds to 9.7 percentiles of the SES distribution. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

B Attrition between 2022 and 2023 survey waves

This section investigates whether attrition between the 2022 and 2023 survey waves correlates with socioeconomic status, age, or baseline ability.

B.1 Age distribution in the main sample

In 2022, children from age 3 were surveyed and assessed. The 2023 survey was administered just over a year after the 2022 wave, which means that children aged slightly more than one year between the waves (1.2 years, on average, in our main sample). Hence, some children who of age 3 at baseline had turned 5 by the endline. Likewise, children of age 2 at baseline – some of which would be 4 by endline – were not assessed in 2022 and are therefore excluded. This gives a slightly skewed age distribution in the main sample, such that children aged 4 at endline are underrepresented. The main sample contains 1,837 children with endline age 4, 2,840 with endline age 5, and 14,344 with endline age 6–10.

B.2 Attrition between 2022 and 2023

Since the main sample focuses on children aged 4–10 at the time of the endline survey, most of these children were aged 3–9 at baseline. However, some children who were 9 years old at baseline will have aged out of the sample frame (e.g., turned 11) by the endline. This makes it non-trivial to specify exactly which children at baseline should be considered for possible attrition.

To address this issue, we look at attrition for *all* children aged 3–9 at baseline. While a subset of these children will not be included in the main sample (i.e., those who turned 11 by the endline), this gives a fair representation of attrition in the relevant age span. Table B.1 shows the results of this analysis.

Table B.1: Attrition of children aged 3–9 in the 2022 survey wave

	Re-surveyed	Attrited	Difference (village FE)
Child age (years)	5.79	6.04	0.30***
	(1.82)	(2.08)	p=0.00
SES percentile	49.24	51.86	0.28
	(28.42)	(29.79)	p=0.61
Mother Edu: $<$ Gr.9	0.25	0.24	0.00
	(0.43)	(0.43)	p=0.64
Mother Edu: Gr. 9-11	0.41	0.40	-0.01
	(0.49)	(0.49)	p=0.13
Mother Edu: Gr. 12+	0.34	0.36	0.01
	(0.47)	(0.48)	p=0.26
Math (2022)†	-0.01	0.03	0.02
	(0.90)	(0.90)	p=0.15
Tamil (2022)†	-0.00	0.00	-0.00
	(0.98)	(1.01)	p=1.00
Observations	19,200	6,161	

Notes: This table presents means and standard deviations (in parentheses) for children aged 3–9 in the 2022 survey wave along a number of characteristics measured in 2022. The first column displays this information for children who were successfully re-surveyed in 2023, and the second for those who were not. The third column shows differences between these groups along with the p-value of the difference, conditioning on village fixed effects. Standard errors of differences are clustered at the village level. † Math and Tamil (2022) baseline scores correspond to the residuals after regressing the original IRT scores on age brackets in years. $p < 0.01$ = ***, $p < 0.05$ = **, $p < 0.1$ = *.

C Measuring socioeconomic status

We construct a household-level socioeconomic status (SES) index based on ownership of a set of assets in the 2022 survey round. Households were asked whether they own a washing machine, refrigerator, grinder, mixer, computer, TV, fan, electric lights, car, tractor, motorbike/scooter, bicycle, and a telephone. Further, we recorded whether the household owns agricultural land and the house they live in, and if they have access to running water. These responses are coded as binary variables, and combined into a single index using principal component analysis (PCA): the first eigenvector constitutes our SES index. Finally, this index is transformed into percentiles. Table C.1 reports descriptive statistics of household asset ownership and maternal education by quartiles of our constructed SES index, as well as for the full sample.

Table C.1: Household characteristics by quartiles of the constructed SES index

	SES quartiles				
	First	Second	Third	Fourth	All
Panel A: Household assets					
Internet	0.20	0.21	0.60	0.87	0.46
	(0.40)	(0.41)	(0.49)	(0.34)	(0.50)
Washing machine	0.01	0.01	0.02	0.28	0.07
	(0.09)	(0.09)	(0.15)	(0.45)	(0.26)
Refrigerator	0.09	0.22	0.58	0.91	0.44
	(0.29)	(0.41)	(0.49)	(0.28)	(0.50)
Grinder	0.56	0.97	1.00	1.00	0.88
	(0.50)	(0.16)	(0.07)	(0.03)	(0.33)
Mixer	0.62	0.99	1.00	1.00	0.90
	(0.49)	(0.10)	(0.05)	(0.00)	(0.30)
Computer	0.02	0.03	0.05	0.21	0.07
	(0.14)	(0.18)	(0.22)	(0.41)	(0.26)
TV	0.75	0.97	1.00	1.00	0.93
	(0.43)	(0.16)	(0.07)	(0.03)	(0.26)
Fan	0.89	1.00	1.00	1.00	0.97
	(0.31)	(0.04)	(0.02)	(0.02)	(0.17)
Electric lights	0.90	0.96	0.97	0.98	0.95
	(0.30)	(0.20)	(0.17)	(0.14)	(0.22)
Car	0.01	0.01	0.03	0.14	0.04
	(0.08)	(0.11)	(0.16)	(0.35)	(0.20)
Tractor	0.01	0.01	0.01	0.08	0.03
	(0.10)	(0.11)	(0.12)	(0.27)	(0.16)
Motorbike	0.38	0.71	0.94	0.99	0.75
	(0.49)	(0.45)	(0.24)	(0.12)	(0.43)
Bicycle	0.27	0.39	0.30	0.48	0.36
	(0.44)	(0.49)	(0.46)	(0.50)	(0.48)
Telephone	0.92	0.99	1.00	1.00	0.98
	(0.27)	(0.10)	(0.01)	(0.02)	(0.15)
Owns land	0.16	0.23	0.24	0.43	0.26
	(0.36)	(0.42)	(0.43)	(0.50)	(0.44)
Owns house	0.83	0.86	0.93	0.96	0.89
	(0.38)	(0.35)	(0.25)	(0.19)	(0.31)
Running water	0.16	0.22	0.27	0.48	0.27
	(0.36)	(0.42)	(0.44)	(0.50)	(0.45)
Panel B: Maternal education					
Mother's education: $<$ Grade 9	0.35	0.28	0.20	0.16	0.25
	(0.48)	(0.45)	(0.40)	(0.37)	(0.43)
Mother's education: $\geq$ Grade 12	0.21	0.28	0.36	0.52	0.34
	(0.41)	(0.45)	(0.48)	(0.50)	(0.47)
Observations	5,013	4,781	5,216	4,304	19,314

Notes: This table shows means and standard deviations (in parentheses) of household asset ownership that forms the basis of our SES index (Panel A) and maternal education (Panel B). The sample is split along quartiles of the SES index in Columns 2–5, and Column 6 shows descriptives for the full sample.

D Empirical Bayes estimates of village-level value-added measures

This section details the construction of Empirical Bayes estimates used throughout the paper. We follow a simplified version of the approach used by Andrabi et al. (2022). Let

$$y_{islv} = \theta_{slv} + \Gamma X_{islv} + \epsilon_{islv} \tag{4}$$

where y_{islv} is the test score of child i in the private/public sector s and private/primary level l, in village v. θ_{slv} is the village-level average value-added in a given sector, at a given level. X_{islv} is a vector of controls (lagged test scores, SES index deciles, maternal and paternal education, gender, and age), and ϵ_{islv} is an idiosyncratic error term. The variance of value-added, denoted σ^2_{sl}, is common across villages but allowed to differ between sectors. The variance of the error term is denoted by σ^2_{ϵ}. Both are assumed to be independent and homoskedastic. We denote the number of children in a given village-level-sector cell as N_{slv}.

Our estimate of θ_{slv} — i.e. the village-level-sector fixed effect — is

$$\hat{\theta}_{slv} = \theta_{slv} + \frac{1}{N_{slv}} \sum_{i \in slv} \epsilon_{islv} \tag{5}$$

The variance of this estimate is equal to

$$Var(\hat{\theta}_{slv}) = E\Big[\Big(\theta_{slv} + \frac{1}{N_{slv}} \sum_{i \in slv} \epsilon_{islv}\Big)^2\Big] \tag{6}$$

$$= E(\theta^2_{slv}) + E\Big(\frac{1}{N^2_{slv}} \sum_{i \in slv} \epsilon^2_{islv}\Big) \tag{7}$$

$$= \sigma^2_{sl} + E\Big(\frac{1}{N_{slv}} \sigma^2_{\epsilon}\Big) \tag{8}$$

The second equality follows from the assumption that ϵ_{islv} is independent and identically distributed at the child level. Rearranging terms, the variance of value-added purged of estimation error is equal to

$$\sigma^2_{sl} = Var(\hat{\theta}_{slv}) - E\Big(\frac{1}{N_{slv}} \sigma^2_{\epsilon}\Big) \tag{9}$$

We can obtain an estimator of the left-hand side by plugging in moment estimators on the right-hand side. $Var(\hat{\theta}_{slv})$ is estimated as the sample variance of the fixed effects.[19] The variance of the error term, σ^2_{ϵ}, is estimated using residuals from Equation (4). An estimate of σ^2_{sl} in Equation (9) is obtained by taking the average of the right-hand side.

Given a standard hierarchical model with normal priors, the Empirical Bayes scaling term is

[19] V denotes the number of villages: $Var(\hat{\theta}_{slv}) = \frac{1}{V} \sum_{v=1}^{V} (\hat{\theta}_{slv} - \hat{\mu}_{sl})^2$, where $\hat{\mu}_{sl}$ is equal to $\frac{1}{V} \sum_{v=1}^{V} \hat{\theta}_{slv}$.

then given by

$$h_{slv} = \frac{\sigma_{sl}^2}{\sigma_{sl}^2 + \frac{1}{N_{slv}}\sigma_\epsilon^2} \tag{10}$$

We shrink each fixed effect toward its level-sector mean $\hat{\mu}_{sl} = \frac{1}{V}\sum_{v=1}^{V}\hat{\theta}_{slv}$, where V is the number of villages. The Empirical Bayes estimate of average value-added in a village-level-sector cell is therefore given by:

$$\hat{\theta}_{slv}^{EB} = h_{slv} \cdot \hat{\theta}_{slv} + (1 - h_{slv}) \cdot \hat{\mu}_{sl} \tag{11}$$

Intuitively, as the sample size of a given cell (N_{slv}) approaches infinity, h_{slv} tends to 1 such that the Empirical Bayes estimate is simply equal to the fixed effect. At the other extreme, the Empirical Bayes estimate shrinks the coefficient completely to the level-sector mean $\hat{\mu}_{sl}$.