Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets
Tahir Andrabi, Jishnu Das and Asim Ijaz Khwaja

ONLINE APPENDIX

Appendix I: Sampling, Survey, Data and Experimental Protocol

A. Sampling Design

Our sample includes 112 villages across three districts in Punjab, Pakistan, with the districts chosen to represent the different socioeconomic zones within the province (one each from the North, Center and South of the province). Villages were randomly chosen from among those with at least one private school, where the list frame is the 2000 private school census of Pakistan. Across a broad range of socioeconomic characteristics, the LEAPS villages are identical to villages with at least one private school in the three sampled districts (Andrabi et al., 2007).

The random selection of villages into the experimental frame helps alleviate "partner and location selection" concerns of the sort raised by Allcott (2015). Figure 1 below shows a typical village in our sample, overlaying GPS coordinates for public and private schools on the Google-Earth™ image of the village. The village is two square kilometers in area but has seven different schools, public and private.

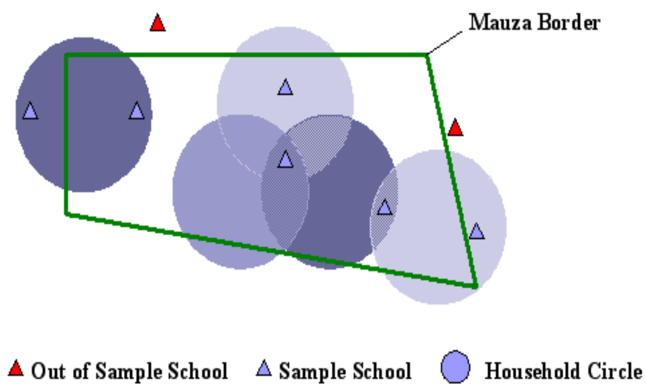
Figure 1: Village Aerial View

This figure presents a Google Earth™ image of a typical village in our sample. The school type icons are labeled (government/public or private) and the shopping cart icon designates

To evaluate the market-level impact of the intervention, we required our sample villages to be "closed markets." To be able to define a complete and closed marketplace can be challenging. For example, a closed market for college education in the U.S. would likely cover all U.S. colleges. More generally, differential travel costs for different types of households can create problems in environments with significant travel options—such as the primary schooling market that includes private schools in high-income countries. In the absence of administrative attendance data, identifying closed markets requires that an educational market be effectively defined within narrow geographical boundaries. Our geographical setting of rural Punjab allows us to define such markets since villages are often separated by natural boundaries, and distance to school largely determines schooling choices.

In particular, we constructed boundaries around the sampled villages that were within a fifteen-minute walking distance from *any* house in the village. All institutions offering formal primary education within this boundary were covered by our study and are considered to be the "village" schools. Figure 2 illustrates this using a *hypothetical* example. The darker/red schools in the diagram would not be in our sample (they are more than 15 minutes from any household). The polygon represents the village ("Mauza") border. As we note in the paper, this process in practice lead us to effectively cover relevant school in a village and achieve our objective of capturing the impact at the market level. Specially, our schooling census confirmed that 92 percent of children in the sampled villages attended the schools in our study. Of the 8 percent that didn't, 6 percent were older children who had graduated to middle school and 2 percent were studying in locations farther away, including outside the country likely due to family members who had migrated for work outside the village.

Figure 2: School Sample Selection



B. Survey Instruments and Timeline

We use data from a range of surveys over a 2-year period of the LEAPS project. In year 2003 before the start of the project, a household and school census were conducted in the sample villages to construct the final sample for the project. In the first year of the project (2004), we administer two sets of surveys, school-based and household-based surveys, which act as the baseline for the study. School-based surveys are administered on the school premises, and include general school surveys, teacher surveys, child tests and child surveys. Household-based surveys are administered to a randomly selected set of households in the sample villages. These surveys are repeated around a year later. In addition, prior to the baseline, we conducted household and school censuses in all of our villages to establish our sampling frame from the study. The following table provides a timeline of these surveys and the intervention.

Timeline for Leaps Surveys		
Activity	Survey Type	Date
Sampling Frame	Household Census	August, 2003
	School Census	August, 2003
Year 1 Surveys	Household-based Surveys	February – April, 2004
	School-based Surveys	February – March, 2004
Intervention	Report Card Delivery	September, 2004
Year 2 Surveys	Household-based Surveys	November –December, 2004
	School-based Surveys	January – February, 2005

School-based surveys

i. Census

This survey collected information on all schools present in the village including the grades taught at the school and distance to school from the center of the settlement. This information was then used to construct the final sample for the LEAPS project.

ii. General School surveys

This survey is answered either by the owner or by the head teacher and collects information on fees, enrollment, infrastructure, funding sources, expenditures, school time-roster. Over 800 schools across 112 villages in our sample receive this survey.

iii. Teacher surveys

There are three components to the teacher survey: (i) a teacher roster survey that collects basic demographic information on all teachers in the school; (ii) a head teacher survey that collects basic demographic information for the head teacher in the school, and information regarding their contracts and tenure at the school; and (iii) a teacher survey administered to Grade 3 teachers, which collects information about the teacher's personal and educational background, their professional history, and other details about their work environment. Through these surveys, we collect information on over 6,000 teachers across our sample. The roster exercise gives us information on around 4,900 teachers and the other two more detailed surveys provide information on another 1,600 teachers.

iv. Child Tests

We administered tests of English, Mathematics and Urdu (the national language) as part of the LEAPS survey. This exercise was bundled with the other school surveying activities and took two and a half hours to complete in each school. These norm-referenced tests were designed and validated by our team. Norm, rather than criterion-referenced testing was chosen since the former allows us to measure learning with higher levels of precision at all levels of knowledge.[1] All children in Grade 3 in the sample schools who were not absent on the day of the test were tested in the three subjects at baseline; in the follow-up round, we tracked all children in the roster from the previous year and tested them if present at any school in the village in addition to any new kids enrolled in the tested grade (see Appendix I.E. for further details). The same tests were administered by our team across all schools and test materials were not shared after testing to ensure impartial and comparable test circumstances. In order to facilitate comparisons of the test over time and to better relate the test to underlying student knowledge, we use item response theory to score and equate the test appropriately adjusting for the difficulty of each question.[2] The scores thus derived are standardized for the first year (with mean 0 and standard deviation 1), but the test-equating methodology imposes no further restrictions across years, so that the 2nd year scores are standard deviation changes normed by the first year distribution.

Like in other low-income countries, average learning levels are low. By the end of Grade 3, most children have only mastered the subject curricula for Grade 1. They can add and subtract single but not double-digit numbers, cannot tell the time, and only top performers can complete simple multiplication and division problems. In Urdu, they cannot form a sentence with the words "school" or "beautiful," and less than 20 percent can comprehend a simple paragraph. In English, most children cannot recognize simple three-letter words such as "bat."

v. Child Surveys

This survey is administered to randomly selected children in the tested grade and gathers information on child educational history, household composition, household assets, and child anthropometrics. In every school, we survey 10 children from the tested grade; in schools with less than 10 children in the tested grade, all children are surveyed. This exercise gives us information on over 6,000 children.

[1] Criterion-referenced tests help identify whether students meet a criterion but can be less informative for those below or above the critical level. Norm-referenced tests seek to maximize variation by estimating the learning level of a particular student. To design the test, an extensive pilot was carried out to identify lower and upper limits of learning in the population and analyze the validity and reliability of the instrument used. The data from this phase was used to refine the final test. As a result, all three tests (English, mathematics, and Urdu) start from simple problems and gradually increase in difficulty. Andrabi et al. (2002) detail the psychometric properties of the test, including tests of reliability.

[2] In addition to the appropriate weighting of test score items by their difficulty, item response allows us to determine the precision of the test instrument at different points along the children's knowledge distribution. Our analysis shows that scores around the middle of the distribution are more precisely estimated than at the ends of the distribution; this is a standard issue with all tests, since items designed to provide information at the extremes of the distribution also add to information for the mean, but not necessarily the other way around (Andrabi et al., 2002). We equate test scores across years using item response scaled scores, where identification is based on a set of rotating questions that were repeated across years (Das and Zajonc 2010). This ensures that all test-scores are reported on the same scale and comparable over time.

Household-based Surveys

i. Census

This survey collected information on all households present in the sample villages; this information included a household roster, other household demographics, and schooling details for children aged 5-15 in the household. In total, over 80,000 households across 112 villages were surveyed. The data was used to stratify the household sample by child enrollment status and to over-sample households with children who might be eligible by age for (tested) Grade 3.

ii. Household surveys

This survey is administered to over 1,800 households. It collects information on a range of variables: household roster, household expenditures, educational history, health, child and parental time use, child care, child ability, school information and ranking, teacher information, mobility, household learning environment, and more. In the first year, the household survey is administered to both one male and one female member of the household. In the second year, only one member (either male or female) responds to this questionnaire.

C. Variable Descriptions

Variable	Description	Survey Source
	A. Primary Outcomes	
Perception of School Quality	School quality is ranked on a Likert Scale ranging from 1 to 5, where 1 is very poor and 5 is excellent. Respondents are only asked to rank familiar schools. Specifically, perception observations are recorded at household (HH)*School level in the survey. *Note:* In the baseline survey (year 1) perception questions were asked from both female and male respondents while in the post survey (year 2), we had an abridged survey and these questions were asked only from one respondent in the household.	Household
Fees	Annual measure of fee charged by the school, which combines admission fee and tuition fee. For construction at the village level, we simply average the school fee variable. *Note:* We also collect a household report on fee paid by the household for child, and this is used in robustness checks in some tables.	School
Child Average Test Score	Average of test score across three subjects – Math, English, and Urdu. For construction at school or village level, we average the child average test score.	Child Tests

	Note: We use Item Response Theory to score the tests, and we use linking items across the years to ensure that the test-scores are graded on a common scale. This implies all scores are normed to the same distribution.	
Primary Enrollment	Sum of enrollment across grades 1 to 5 for a given school. For construction at the village level, we simply aggregate school enrollment.	School
	B. Secondary Outcomes	
Number of children going into school	This variable is generated using tracking data from the child roster for the tested cohort (grade 3 at baseline) and is available for year 2 only; it cannot be constructed for any other cohort in the school. It includes all children who were confirmed to have switched into a school or who were newly enrolled in a school in year 2.	Child Roster
Number of children going out of school	This variable is generated using tracking data from the child roster for the tested cohort (grade 3 at baseline) and is available for year 2 only; it cannot be constructed for any other cohort in the school. It includes children who were confirmed to have switched out or dropped out of school of their baseline school, and untracked children from schools that closed down in year 2.	Child Roster
Private school closure	Dummy for whether a private school is open in year 2	School
Parental time spent on teaching child	Sum of parental time spent reading and helping child with homework, measured in hours per week	Household
Parental non-fee spending on education	Sum of spending on non-fee educational items in rupees per year including transport, private tuition and pocket money. This is collected for each child in the household.	Household
Parental-Teacher Interaction Variables	These variables include: (i) Whether a parent has ever met their child's teacher; (ii) If they are able to recall the teachers name; and (iii) Their assessment of the class teacher's involvement. *Note:* We do not verify whether the name the parent recalls is correct, merely whether they can recall.	Household
Break time	Minutes of playtime students have per day at school recorded as part of a school time-roster.	School
Basic Infrastructure Variables	Basic infrastructure refers to whether school has desks/chairs as sitting arrangement (relative to floor, mats, etc.), blackboards per child, toilets per child and classrooms per child.	School
Extra Infrastructure Variables	Extra infrastructure includes dummies for the presence of a library, computer, sports facility, fans, electricity and wall/fence at a school	School
Percent Teacher	Percentage of teachers at a school with at least 12 years	Teacher Roster

with at least Higher Secondary Degree	of schooling. In Pakistan, this is known as Higher Secondary degree, whereas in the U.S., this would typically be referred to as a High School degree.	
	C. Primary Village Level Control Variables	
Village wealth	Median monthly expenditure incurred by the household	Household Census
Village size	Number of households in village	Household Census
Herfindahl index of school	Measure of competitiveness in the village with a range between 0 and 1. More competitive settings have lower with lower Herfindahl values.	School Census
Village literacy	Percent of adults over the age of 24 in the village who are literate	Household Census

D. Report Card Design and Delivery

The content, design, and delivery of the report cards were based on focus group discussions with parents and schools, where the consensus was that parents wanted both the (absolute and relative) scores of their child and of schools in their village. The two main design challenges we had to address were that (a) school-level test scores potentially reflect a combination of school performance and child selection and (b) test information may be insufficiently granular to allow households to distinguish between different kinds of schools.

1) Content and Design

To assess whether raw test scores, which we eventually used, were preferable to value-added test scores (where we parse out the contribution of observed characteristics of households and children), we ran a set of regressions with the school test scores as the dependent variable and varying sets of parental and child characteristics as the explanatory variables. In each of these regressions, the joint contribution of parental and child characteristics was small and there was little difference between the value-added and the raw scores (see Das et al., 2012). We traded off the small difference in school rankings with the fact that explaining value-added scores to both parents and schools was harder, lacked transparency, and ignored the possibility that parents themselves had information that was not available to us. For instance, in focus group meetings, parents sometimes raised issues about teachers in certain schools that could explain performance—information that remained unobserved to us as researchers.

We were also concerned about measurement error in test scores and the ability of our tests to distinguish among different types of schools. Using the estimated errors from the item response methods, we computed for every school the measurement error in the tests. Figure 3, for instance, shows box plots of the mean score and the measurement error for all schools in a single village: the high variance in test scores *across* schools implies that the reliability of our school rankings was fairly high (the ratio of the variance of the measurement error to the variance of test scores was low).[3] In most villages, as in the figure, three very different groups of schools emerged with large differences in test scores across them.

Figure 3: Mean Scores and Measurement Error for all Schools in a Single Village

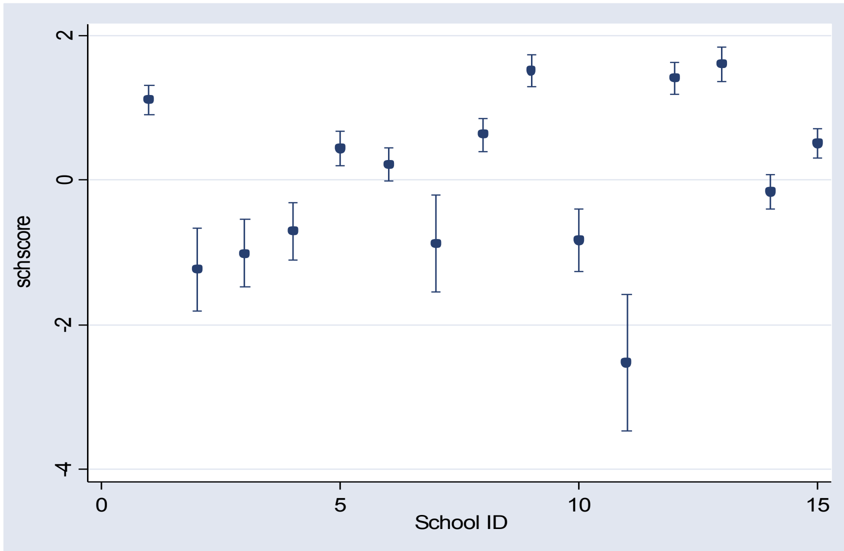


Figure 4 on the next page shows a template of the report cards used. We have filled it in using fake names to protect privacy. Card 1 (the top image) reports the score of the child in English, Mathematics, and Urdu, as well as her quintile rank across all tested children. Quintile rank is described as: 1 – "Very Good"; 2 – "Good", 3- "Satisfactory", 4 – "Needs Improvement" and 5 – "Needs Significant Improvement". The three rows display information for the child, the average child in her school, and in her village.

The lower image shows Card 2 of the Report Card (school information) and gives information on the village schools, one on each row. The columns display the school name, number of tested children, and school scores and quintiles for each of the three subjects. Each card also had detailed instructions (on the reverse side) on how to read the card and what the rankings meant. A school version of the report card included the breakdown by sub-categories of the subject scores for each child and every school also received a bound booklet that contained the report cards for all children to be used by both the teacher and head teacher and to serve as an extra copy in case parents lost theirs.

[3] In the U.S., the precision of tests is high, but the variation across schools is very small; in contrast, in our environment the variation across schools is very high leading to reliable rankings of schools. See Kane and Staiger (2002) and Rogosa (2005).

Figure 4: Report Cards

Card 1

Learning and Educational Achievement in Punjab Schools

رپورٹ کارڈ برائے تعلیمی کارکردگی

New Day School **School Name**

3 **Grade**

Fatima Malik **Child Name**

Rafiq Malik **Father Name**

Math Rank	Math Obtained Marks (Total Marks 100)	English Rank	English Obtained Marks (Total Marks 100)	Urdu Rank	Urdu Obtained Marks (Total Marks 100)	
1st	89	1st	77	1st	85	Child Performance
1st	80	1st	78	1st	67	Average School Score
1st	57	2nd	43	2nd	46	Average Village Score

Ranking
1st: Very Good
2nd: Good
3rd: Satisfactory
4th: Needs Improvement
5th: Needs Significant Improvement

Tahir Andrabi
پروفیسر ڈاکٹر طاہر اندرابی
پراجیکٹ ڈائریکٹر

ستاروں سے آگے جہاں اور بھی ہیں

Chak 2004 **Mauza**

February 23, 2004 **Exam Date**

Card 2

Learning and Educational Achievement in Punjab Schools

کے تمام سکولوں کے بچوں کی اوسط کارکردگی

Math Rank	Math Obtained Marks (Total Marks 100)	English Rank	English Obtained Marks (Total Marks 100)	Urdu Rank	Urdu Obtained Marks (Total Marks 100)	Number of Tested Students	School Name
1st	80	1st	78	1st	67	23	New Day School
1st	51	2nd	43	3rd	33	34	Government Boys
2nd	47	2nd	42	1st	45	30	Government Girls
2nd	50	1st	55	2nd	41	18	Bright Day School

LEAPS
Learning and Educational Achievement in Punjab Schools

Math	English	Urdu	Marks Scale
<27 = Needs significant improvement	<20 = Needs significant improvement	<18 = Needs significant improvement	
27-34 = Needs improvement	20-27 = Needs improvement	18-24 = Needs improvement	
35-42 = Satisfactory	28-34 = Satisfactory	25-33 = Satisfactory	
43-50 = Good	35-43 = Good	34-42 = Good	
>50 = Very Good	>43 = Very Good	>42 = Very Good	

2) Delivery

Given that many (illiterate) parents needed to have the cards explained to them, cards were delivered in person in discussion groups rather than sent by mail. In each discussion group, parents were given a sealed envelope with their child's report card, which they could open and discuss with others, or with members of the LEAPS team. Every group started with a 30-minute open discussion on what influences test score results (teacher, home environment, school environment, and the child), followed by the card distribution. At every discussion group, the team focused on the positive aspects of the card rather than using the card to assign blame. We were concerned about the risk that a poor result would lead to blaming the child. The team was careful to not offer any advice to parents or schools. The goal of the meetings was to provide the report cards and explain what the information meant but not to advocate or discuss any particular plan of action.

The report card intervention thus bundles information provision with discussion of the card during delivery. While these discussions may have had their own independent impact, the additional step arguably helped parents better comprehend the information presented. From the point of view of feasible interventions, information interventions in settings with low parental education may have to be undertaken through such school-level discussions. It is hard to expect people to respond to information unless they also are able to comprehend it or to expect schools to react to information unless they are convinced that parents will react to it.

The total cost of the report card intervention was $1 per child, which includes the child level testing, and the production and delivery of report cards. The policy actionable costs of a scaled-up program are likely to be lower given scale economies (for instance, delivering report cards in a highly geographically dispersed sample added significantly to our costs).

E. Balance and Attrition

I. Baseline Balance

We confirm that baseline values of outcomes and control variables are balanced between treatment and control villages (Appendix III, Table I). Apart from one variable, father's education (slightly lower in treatment villages), expected given random chance, all of the other variables are balanced. Column 1 shows the control means and Column 2 shows the difference between treatment and control villages after accounting for the district-level stratification in the randomization. We are unable to find any significant differences at the usual levels of confidence for a large number of village-level, child-level and school-level attributes in the control and treatment villages. The p-value for the joint test of significance at the village level is 0.56.

II. Child Tracking and Attrition

We track and attempted to test all children in our Grade 3 roster in Year 1 across the follow-up rounds and also newly enrolled children in the grades being tested each round. The initial roster in Year 1 included 13,735 children across all 804 schools, public and private, that offered Grade

3 instruction. Of the total children, we tested 12,110 children in 2004; those (12 percent) not tested was because they were absent the day of the test.[4] All children (tested and non-tested) were tracked and retested in 2005 in whatever grade they were enrolled in at the time. All children were tracked between surveys since children could: (a) drop out; (b) remain in the same school and be promoted; (c) remain in the same school and not be promoted; (d) switch schools within the village and be promoted/not promoted (in which case they would be tested in another school); or (e) switch to non-sample schools (usually due to household migration). Although close to 1,750 children out of the tested 12,110 children were no longer in the same class-school combination that they would have been in if they did not switch schools and followed the natural grade progression, we were able to determine the status of all but 530, giving us a tracking rate of over 96 percent throughout the LEAPS project survey period.

This tracking exercise was undertaken to understand the types and level of attrition that occur in educational interventions. Attrition can often be a serious issue in low-income countries both because children may drop out between testing rounds and because of student absence on a day-to-day basis. Of the total 13,735 children, we have test scores in both years for 9,888 children (72 percent). Of the 12,110 children tested in the first year, we were able to retest 82 percent in the second year. Absenteeism, rather than dropouts, was the main reason for children not being tested. In both rounds, the rate of absenteeism is 12 percent, which is reassuring since it suggests that neither round is an outlier. Since 9 percent of the children drop out or are "lost" (i.e. we are unable to track them) between the two years (and therefore with probability 1 cannot be retested), the expected fraction of (first round class roster) children for whom we have two test scores is (0.88*0.88*0.91=) 71 percent, which is very similar to the 72 percent actually obtained.

Furthermore, we are unable to detect any difference in the attrition rate across treatment and control villages (Appendix III, Table II, Panel A), and are also unable to detect any differences in child or parental characteristics and child test scores for attriters in the treatment and control groups (Appendix III, Table II, Panel B, Columns 1-2). In addition, separating attriters into those who were absent on the day of the test (Panel B, Columns 3-4), those who definitely dropped out (Panel B, Columns 5-6), and those who left the village or whose status we were unable to confirm (Panel B, Columns 7-8) shows that within each of these three types of attriters, the children who were not tested look similar on a host of baseline characteristics (including test scores) in the treatment and control groups. Similar attrition rates and the similarity in baseline characteristics of attriters across treatment and control groups suggest that attrition in our sample is uncorrelated to the intervention, and, therefore, is unlikely to pose a salient concern.

F. Further Measurement and Robustness

I. Measurement of Perceptions

In the text, the perception measure is constructed for each school by averaging over all parents who ranked the school. Since households were not expected to offer their views on schools they were not familiar with (they could respond with "don't know"), we do not always have perceptions for each school. Moreover, in order to compare these measures consistently over

[4] This is quite reasonable since nationwide surveys in Pakistan and India (ASER Pakistan, 2012; ASER India, 2012) show student attendance ranging from 70-79 percent on any given day.

time, we restrict to those household-school observations where we have data in Years 1 and 2. Thus, the total number of schools is lower than our full sample.

We obtain similar results in Appendix III, Table III if we: (i) also include data/schools where a different set of households provided perceptions across the two time periods (or restrict the data further to the same respondent in both periods as opposed to the same household); or (ii) run these regressions at the household X school level rather than aggregating at the school. We prefer our approach as it both avoids inter-household perception comparisons and is not as restrictive as insisting that the same respondent reports in both periods. We prefer aggregation at the school-level because it reduces noise, assigns equal weight to all schools, and also reveals the market-level perception of each school's quality, which is our primary object of interest.

II. Bounding Estimates from Switching

In the main text, we show that the impact of report cards on test scores is identical for children who did not switch schools as for the full sample. But if switching responds to the treatment, estimates restricted to non-switchers will be biased. As a bounding exercise, we compute the gain required by switchers to drive the observed overall treatment effect (OTE) and show this to be too large to be plausible (2.25 standard deviations). To see this, note that the OTE in Table III, Column 6 can be mechanically decomposed into OTE = $= \delta + \gamma(R_S^T - R_S^C) + \beta R_S^T$. The first term, δ, is treatment gain (between treatment and control) for children who do not switch; γ from the second term gives the general gain (common to treatment and control) that switchers experience, which is multiplied by the change in the number of switchers (between treatment and control) where R_S^T and R_S^C are the fraction of switchers in the treatment and control villages; and β is the additional gain for switchers (R_S^T) in the treatment villages. This calculation gives us OTE = 0.113+0.046*(0.009)–0.104*(0.049)=0.113+0.0004–0.005 = 0.108. We caution that this decomposition should not be interpreted as presenting causal effects for each category, given that switching may be an outcome of the treatment. Moreover, it assumes equal number of children in treatment and control villages; in practice, the two numbers differ by 1 percent. The decomposition also provides our bounding number. Assuming no treatment effect for non-switchers (δ =0) and no differential switching between treatment and control ($R_S^T = R_S^C$), switchers in treatment relative to control villages would need to experience a gain (β) of 0.108/0.048 = 2.25 standard deviations. While this may lower if switchers experience general gains, with only 0.9 percent more switchers in treatment villages (not significantly different from control villages), the learning gains from switching would still need to be implausibly large in order to generate the observed overall treatment effect.

III. Thresholds for heterogeneity results

Our preferred binary classification sets the 60th (rather than 50th) percentile as the threshold for classifying a school as high scoring and uses the full sample of schools (as opposed to within village) to define the percentile rank for the threshold. Because private schools generally score higher than public, this leads to a more even division of private schools classified as initially low achieving. Using a within-village cutoff may force schools with similar test scores to be classified as initially high or low achieving depending on the village in which they are located. Moreover, it is quite likely that the differential cost of upgrading quality (relevant for when we show impact heterogeneity on test scores) relies on an absolute notion of quality (rather than just within village quality). In practice, the within sample or within village classifications are not that different given that there is more variation in school quality within than across villages.

Nevertheless, Appendix III Table VI shows that our results are similar regardless of which classification is used: Initially high scoring private schools always show a greater price decline (with similar magnitudes) and the effect is statistically significant at conventional levels in all specifications except when we split at sample median and introduce baseline controls (the p-value in that case becomes 0.13).

Appendix II: Theory

1 Introduction

The model of Wolinksy (1983) presented in the text outlines the characteristics of the separating equilibrium and the price markup required to sustain separation. The aim of this accompanying Appendix is twofold. First, we present closed form solutions for the price markup. The closed form solution highlights the role of the cost functions and the information structure and helps map the model to the empirical predictions. Second, we examine the comparative statics of optimal quality choice when information improves. This second part demonstrates the link between the structure of demand and quality choices. It derives the necessary conditions for our empirical predictions to hold.

The Appendix is structured as follows. We first describe the setup and define the utility function of consumers and the profits of schools. We then show the price conditions required for incentive compatibility and derive a closed-form solution for the price markup. Using the close-form solution, we study the comparative statics of an improvement in information from the report cards on the price markup. Finally, we use this price markup to study the comparative statics of quality choice under asymmetric information. Turning from asymmetric information to other candidate models, we briefly discuss the predictions from a model of symmetric information where the report cards provide feedback to schools and parents who were unaware of their own performance.

The Appendix clarifies the intuition and derivation with market clearing conditions. As such, rather than a theoretical contribution, it should be regarded as a (more detailed) map to the empirical findings in the main text. There are key technical issues discussed in Wolinsky (1983) that we point the reader to, as required, but do not elaborate in this Appendix.

2 Setup

- **Consumers**: There is a continuum of consumers indexed by j. Every consumers purchases one unit of schooling and consumers differ in their tastes, θ_j where $\theta_j \sim F(\theta)$ and $\theta_j \geq 0 \; \forall \; j$. Consumers receive utility, $U_j = \theta_j v_i - P_i$, where v_i is the quality of the school they choose and P_i is the price (school fees) that they pay. Higher θ_j consumers 'value' schooling more, and are therefore willing to pay a higher price for the same quality.

- **Schools:** The profits of a private school i, are $\Pi_i = (P_i - c(v_i))q_i(P_i, v_i)$, where P_i is the price charged by school i, $c(v_i)$ is the cost of producing quality v_i and q_i is the mass of consumers who will purchase school $i's$ product at P_i, v_i. However, consumers need not purchase the private school product; they can always choose an outside option whose price and quality we normalize to zero. In our model of how quality is produced we maintain the direct correspondence with Wolinsky (1983), assuming that an increase in the quality increases the cost of production for every unit sold. That is quality enters as variable rather than fixed cost.[1]

3 Model with Asymmetric Information (Wolinsky 1983)

3.1 Structure of Information

The fundamental building block of the model is the presence of asymmetric information. In the model, schools always know what quality they have produced, but consumers receive a noisy signal. This noisy signal of quality can be modelled in a number of ways as long as it satisfies the property of the $D(.)$ function described in the text and Wolinsky (1983); we specify consumer information as follows.

Consider two potential quality choices, v_i and $\widetilde{v_i}$ and introduce the notion of an announcement, whereby schools can produce v_i, but announce that it has produced $\widetilde{v_i}$ instead. Define $x(v_i, \widetilde{v_i})$ as the fraction of consumers who receive a signal consistent with the school having produced $\widetilde{v_i}$ when in fact, it has not (if

[1] We do not include a fixed-cost of entry, which is the technical condition required in the free-entry model in Wolinksy (1983). The key intuitions and mapping to the empirics do not rely on the free entry condition.

$\widetilde{v_i} \neq v_i$). In deriving the conditions for separation, these consumers will purchase from the school believing that its quality is $\widetilde{v_i}$. Then, $1-x(v_i, \widetilde{v_i})$ is the fraction of consumers who receive a signal that a school *cannot* have produced $\widetilde{v_i}$ and will therefore not purchase from the school. For algebraic simplicity in this Appendix, we have redefined the D(.) function in the text, so that $x(v_i, \widetilde{v_i})$ in the Appendix is $1 - D(.)$ in the text. The function $x(v_i, \widetilde{v_i})$ satisfies the following two assumptions:

- [A1]: $x(v_i, \widetilde{v_i}) = 1$ when $v_i = \widetilde{v_i}$. We normalize the signal such that whenever a school announces and produces v_i all consumers receives a signal consistent with v_i.
- [A2]: $\frac{\partial x(.)}{\partial(\widetilde{v_i} - v_i)} < 0$. That is, the likelihood that consumers will receive the signal that it has produced $\widetilde{v_i}$ when it has actually produced v_i decreases with the difference $\widetilde{v_i} - v_i$. In essence, it is easier to "fool" consumers when the school produces a quality just below what it announces.

This fraction $x(.)$ can be micro-founded from the $D(.)$ function as described in the text. Wolinsky (1983) specifies the required out-of-equilibrium actions–that is, what happens when consumers try to purchase from a school that they believe to be $\widetilde{v_i}$ based on the announcement but receive a signal at the time of purchase that is inconsistent with $\widetilde{v_i}$.

3.2 Price Conditions for Separation

3.2.1 Price and Incentive Compatibility

We are interested in the price required for incentive compatibility (IC)–that is, the price under which a school that announces v_i indeed produces v_i. Consider first the simpler case where schools can produce only one of two qualities: v_h or $v_l \leq v_h$. The timeline is that schools first produce $v_i \in \{v_h, v_l\}$ and then announce that they have produced $\widetilde{v_i} \in \{v_h, v_l\}$. We are interested in the price that guarantees $\widetilde{v_i} = v_i$, which we define as $P^{IC}(v_i, \widetilde{v_i})$.

A school will never produce v_h and announce that it has produced v_l because it will receive a lower price from doing so. Therefore, we need to derive the price such that a school that has produced v_l will not announce that it has produced v_h.Formally we wish to prevent an equilibrium where $v_i = v_l$ and $\widetilde{v_i} = v_h$. In order to derive the price required for separation, we compare the profit from producing v_h and announcing v_h (that is, $\widetilde{v_i} = v_i = v_h$) versus the profit from producing v_l but announcing v_h (that is, $\widetilde{v_i} = v_h$ and $v_i = v_l$). From condition [A1], when $\widetilde{v_i} = v_i = v_h$, $x(v_h, v_l) = 1$. Therefore, the profit from announcing v_h and producing v_h is

$$(P^{IC}(v_h, v_h) - c(v_h))q_h(P^{IC}(v_h, v_h), v_h) \tag{1}$$

Alternatively, the profit from announcing v_h but producing v_l is

$$(P^{IC}(v_h, v_l) - c(v_l))q_h(P^{IC}(v_h, v_l), v_h)x(.) \tag{2}$$

As the price depends only on the announcement and not the actual quality produced, $P^{IC}(v_h, v_h) = P^{IC}(v_h, v_l)$. Therefore, Equation (2) shows that when the school announces v_h but actually produces v_l, it makes an additional markup on every unit sold because it produces at the lower marginal cost $c(v_l)$, but continues to receive the price $P^{IC}(v_h, v_l)$. However, when it does so, it retains $x(.) < 1$ consumers, as $1-x(.)$ consumers will receive a signal inconsistent with v_h and purchase schooling elsewhere as in Wolinsky (1983). Equating the two yields the lowest possible incentive compatible price, $P^{IC}(v_h, v_l)$. Specifically:

$$P^{IC}(v_h, v_l) = \frac{c(v_h) - x(v_h, v_l)c(v_l)}{1 - x(v_h, v_l)} \tag{3}$$

$$\Rightarrow \quad P^{IC}(v_h, v_l) = c(v_l) + \frac{c(v_h) - c(v_l)}{1 - x(v_h, v_l)} \tag{4}$$

Noting that we have defined $x(.)$ as the equivalent of $1 - D(.)$ in the text, Equation 3 is the incentive compatibility condition described in the text. To see the markup required for incentive compatibility to hold, we rewrite

$$P^{IC}(v_h, v_l) = c(v_h) + \frac{x(.)}{1 - x(.)}[c(v_h) - c(v_l)] \tag{5}$$

The markup required to satisfy incentive compatibility therefore depends (a) on the cost difference between the two quality levels over which the comparison is conducted and (b) the precision of information given by the term $x(.)/1-x(.)$, which is the ratio of consumers who receive an incorrect signal to those that receive a correct signal. Note that in this particular formulation of the profit function, the structure of demand does not affect the price markup required for incentive compatibility.

3.2.2 Comparative Statics: Markup and Information

Equation 5 immediately leads to the first result on markup and information. Since a higher $x(.)$ implies less precise information and $\frac{\partial x(.)/(1-x(.))}{\partial x} > 0$, we make the following observation also discussed in the text.

Claim 1 *In a separating equilibrium, incentive compatibility implies that schools earn a markup over and above their production cost and this markup increases as the precision of the information signal declines.*

Equation 5 was derived for only two quality choices v_h, v_l, but in general, there could be a continuum of potential v_i that the school could choose to produce and a continuum of $\widetilde{v_i}$ it could choose to announce. We now generalize Equation 5 by first deriving the conditions so that the IC condition holds for every $\widetilde{v_i}$ that the school may wish to announce once it has produced a specific v_i. This generalization will then allow us to derive a specific closed-form for the markup and provide further results on the evolution of the markup for different initial school qualities.

Define $P^{IC}(v_i, \widetilde{v_i})$ as the price required for separation so that the announcement of v_i also implies that the school has produced v_i and as before let the announcement of quality be $\widetilde{v_i}$ and the true quality produced v_i. The profit function can then be defined as

$$\pi(v_i, \widetilde{v_i}) = (P^{IC}(v_i, \widetilde{v_i}) - c(v_i))x(v_i, \widetilde{v_i})q(P^{IC}(v_i, \widetilde{v_i}), v_i) \tag{6}$$

Then, for $\widetilde{v_i} = v_i$, that is the announcement satisfies truth-telling, we require that

$$P^{IC}(v_i) : v_i = \arg\max_{\widetilde{v_i}} \pi(v_i, \widetilde{v_i}) \forall v_i. \tag{7}$$

For a tractable solution, we specify functional forms for the information and cost function as follows.

- $x(v_i, \widetilde{v_i}) = 1 - \alpha(\widetilde{v_i} - v_i)$, where $\alpha \in (0, 1)$. That is, the fraction who receive a signal consistent with $\widetilde{v_i}$ when the school announced $\widetilde{v_i}$ but produces v_i decays in the difference between the announced value, $\widetilde{v_i}$ and the true value, v_i at the rate α.When $\alpha = 0$, information is very poor and every consumer receives a signal consistent with $\widetilde{v_i}$ regardless of true production quality. This proportion decreases as α increases. Note that when $\alpha = 1$, information is more precise, but it is not perfect. We choose this formulation because we do not wish to claim that our report card intervention led to perfect information; only that it increased precision.

- The cost function is quadratic so that $c(v_i) = \frac{v_i^2}{K}$.

Then, for any comparison $v_i, \widetilde{v_i}$, use Equation 3 to derive:

$$\begin{aligned} P^{IC}(v_i, \widetilde{v_i}) &= \frac{v_i^2}{K} + \frac{(\widetilde{v_i^2} - v_i^2)}{\alpha K(\widetilde{v_i} - v_i)} \\ &= \frac{v_i^2}{K} + \frac{(v_i + \widetilde{v_i})}{\alpha K} \end{aligned} \tag{8}$$

To ensure that $\widetilde{v_i} = v_i$, we now only need to pick the highest $P^{IC}(v_i, \widetilde{v_i})$ for all potential $\widetilde{v_i}$.Theoretically, the specific $\widetilde{v_i}$ that generates the highest $P^{IC}(v_i)$ can be at any point on the potential distribution of v_i. Note however, the familiar result from the asymmetric information literature that only constraints in the *neighborhood* of v_i need to be checked when the IC inequalities are well behaved. As seen from Equation 8, this holds in this case as well because $P^{IC}(v_i, \widetilde{v_i})$ is *strictly increasing in* $\widetilde{v_i}$. Consequently, the constraint where maximal incentives need to be provided will be as $\widetilde{v_i} \to v_i$, and we are able to derive the closed-form solution:

$$P^{IC}(v_i) = \frac{v_i^2}{K} + \frac{2v_i}{\alpha K} \tag{9}$$

The first term is the cost of producing v_i and the second is the markup over and above this cost. The markup is (a) increasing for higher values of v_i; (b) increasing as information becomes less precise (α declines) and; (c) decreasing as the cost of production declines (K increases).

Equation 9 shows the price link between our empirical results and the theory. Specifically, the markup is consistent with our empirical result that the decline in prices should be higher for schools with higher baseline quality. To see this, note that $\frac{\partial P^{IC}(v_i)}{\partial \alpha} = -\frac{2v_i}{\alpha^2 K}$, which implies that when information improves with a higher value of α, the decline in the IC price is higher at higher values of v_i. This leads to the following observation:

Claim 2 *Information markups are positive for all quality levels, but higher for higher quality schools. Therefore, when the precision of information improves, prices will decline for all schools and more for higher quality schools.*

3.3 Market Clearing Optimal Quality Choice

We are now in position to examine the second part of our empirical predictions, which relate to quality. Thus far, we have derived the IC price as a function of quality such that schools will always produce what they announce in equilibrium. Once this price function has been derived, the school will optimize its choice of v_i subject to the $P^{IC}(v_i)$ function and the demand function $q(P^{IC}(v_i), v_i)$. The intuition for how schools will respond in their quality choices when the precision of information improves can be understood from the decision of a single school, which we turn to next.

Given the $P^{IC}(v_i)$ strictly binding at all values of v_i, which we can verify in our case, we can now derive the optimal v_i^* as follows. For any $P^{IC}(v_i)$, v_i, consumers will purchase from the school as long as $U_j = \theta_j v_i - P^{IC}(v_i) \geq 0$, or $\theta_j \geq \frac{P^{IC}(v_i)}{v_i}$. Since $\theta \sim F(\theta)$, the fraction of consumers who will purchase from the school at price $P^{IC}(v_i)$ and quality v_i is $1 - F(\frac{P^{IC}(v_i)}{v_i})$ Therefore, profits are given by

$$\Pi_i(v_i) = \frac{2v_i}{\alpha K}(1 - F(\frac{P^{IC}(v_i)}{v_i})) \tag{10}$$

where the first term outside the brackets is $P^{IC}(v_i) - c(v_i)$ and $P^{IC}(v_i)$ is given by Equation 9. Note first that $\frac{P^{IC}(v_i)}{v_i} = \frac{v_i}{K} + \frac{2}{\alpha K}$ and as α becomes small (information worsens), profits will become zero or negative at higher values of v_i. Therefore, for every α, there will be a range of v_i that result in zero or negative profits, where schools will never produce. Although the specific values will depend on the $F(.)$ function, this implies that poor information will lower equilibrium quality.

Now assume that α is sufficiently large for an interior solution and differentiating with respect to v_i yields the first order condition for the optimal choice of v_i^*:

$$1 - F(\frac{v_i^*}{K} + \frac{2}{\alpha K}) - \frac{v_i^*}{K} f(\frac{v_i^*}{K} + \frac{2}{\alpha K}) = 0 \tag{11}$$

It is useful to rewrite this as $\frac{f(.)}{1-F(.)} = \frac{K}{v_i}$ to highlight the role of demand in the determination of optimal quality. Given the information markup, the school must compare the loss in profits from decreased demand at the higher price in the same location to an increase in demand from distorting its quality choice.

3.3.1 Comparative Statics: Optimal Quality Choice and Information

We are now in position to examine differential responses among schools to a change in the information environment. Let Equation 11 define an implicit function in α and v_i^*, $I(\alpha; v_i^*)$. Using the implicit function theorem, $\frac{\partial v_i^*}{\partial \alpha} = -\frac{\partial I/\partial \alpha}{\partial I/\partial v_i^*}$. The term in the denominator is the second order condition of the maximization and we assume the regularity condition $-2Kf(.) < v_i^* f'(.)$ ensuring that $\partial I/\partial v_i^*$ is negative for an interior solution; since the term on the LHS is always negative, this rules out sudden large declines in the valuation function.

How the optimal choice of quality varies with α then depends on the behavior of $\partial I/\partial\alpha$. Differentiating Equation 11 with respect to α yields $\frac{2}{\alpha^2 K}f(.) + \frac{v_i^*}{k}\frac{2}{\alpha^2 K}f'(.)$. Whether this is positive or negative depends on whether

$$f(.) + \frac{v_i^*}{K}f'(.) \lesseqgtr 0 \tag{12}$$

For any point on the valuation distribution that has positive mass, $f(.)$ is always positive. Therefore, the behavior of v_i^* with respect to a change in α is pinned down by the behavior of $f'(.)$. First, if $f'(.)$ is (weakly) positive at all v_i, an improvement in information will lead to an increase in the quality of all schools, regardless of where they are in the quality distribution. This will be the case for instance, for a uniform distribution of valuation where $f(.)$ is constant and therefore $f'(.) = 0$.

Second, when $f(.)$ is increasing so that $f'(.) > 0$, again v_i^* will increase regardless of the value of v_i^*. This is for instance, the case for any point to the left of the mean/mode in normal or lognormal distribution. However, when (a) v_i^* is large and (b) $f'(.) < 0$, so that we are at a part of the distribution where mass is declining and the school is at high quality, the second term in the equation becomes negative, and more so when v_i^* is large. In this case, we may find either no change in quality or even a decline in quality as information becomes more precise. Therefore weak declines in quality are possible in this model only when the mass is declining (in a regular unimodal distribution, this will be to the right of the mean) and is more likely when the school quality is high. To think of distributions where this is likely, rewrite Equation 12 as $1+\frac{v_i^*}{K}\frac{f'(.)}{f(.)} \lesseqgtr 0$. In the family of distributions where $\frac{f'(.)}{f(.)}$ is monotonically decreasing, smaller or even negative quality responses are more likely at higher v_i; this is the family of log-concave distributions frequently used in the literature on asymmetric information. This leads us to our second observation.

Claim 3 *When the probability distribution function is increasing, quality will always improve when information becomes more precise. When the probability distribution function is decreasing, quality may remain the same or even decline. For log concave distributions, quality will improve with information at low initial quality but may not improve at high initial quality.*

This is the central dynamic highlighted in our text: Improving information leads to a decline in price and increase in quality for schools at low quality. At high quality, $f'(.)$ may be negative and v_i^* is high. In this case quality changes are muted as the marginal increase in demand from increasing quality is smaller.[2]

4 Feedback

We make one final observation to help differentiate the model's predictions relative to an alternate class of models where the report card provides feedback to schools on their own performance. To model the report cards as a feedback mechanism where schools are unsure of their own performance, we assume that information was poor, but fully symmetric between schools and parents. Therefore, the arrival of report cards provides information to schools and parents about the schools performance and is new information for both.

Assume that schools produce quality v_i, but now, v_i depends on a fixed input Z_i and effort, e_i. The fixed input Z_i is costless and fixed for every school and captures the innate productivity of the school staff. Schools exert costly effort e_i to obtain quality $v_i = g(Z_i, e_i)$, with an associated cost function $c(e_i)$. For concreteness, assume that the function, $g(Z_i, e_i)$ is such that the two inputs are complements: $v_i = Z_i e_i$ so that the cost of producing v_i is $c(\frac{v_i}{Z_i})$. Z_i is unknown to both schools and consumers, so that there is symmetric lack of information. However, both schools and consumers know the distribution of Z_i, which we call $Q(Z_i)$. For concreteness, let $Z_i \sim Q(Z_i) : E(Z_i) = 1$. Schools and parents do not observe v_i, so that the final quality produced is unknown to both.

[2] Our tractable closed-form solution with market clearing highlights the key intuitions but can be complicated substantially. For instance, we have not introduced the additional strategic considerations that will come with multiple schools in the same market. When the market is fully competitive as in Wolinksy (1983), it is straightforward to see that in the full information case, each firm will price at marginal cost and the $P^{IC}(v_i)$ schedule will remain unchanged. Depending on specific parameter values, some schools may not exist in equilibrium as the price required to sustain these schools may be too high.

There are now two cases. First, if e_i is also unobserved to both, there is never any incentive to exert effort. All consumers will assume that schools exert zero effort and there is a single pooling equilibrium. This does not match the baseline results, where higher quality schools charge higher prices. The more interesting case occurs when e_i is observed both by schools and consumers, retaining the assumption of symmetric information. When only effort is observed, both schools and parents operate under the assumption that $Z_i = E(Z_i) = 1$. Therefore, the true Z_i cannot affect the choice of effort and consumers and schools will price based on *expected* rather than true quality. There is variation in quality, v_i, but because this variation arising from Z_i is unknown to both schools and parents, the equilibrium price depends only on effort and not Z_i. This is fully consistent with the baseline price-quality regression, since parents observe higher e_i and will reward such schools more. Since Z_i is unknown to both parents and schools, expected quality and price will be positively correlated. This is also why we now need two factors in the production of quality–to ensure a baseline correlation between price and quality, it must be that one factor is observed and priced, but another is not.

When v_i is revealed, there are two effects. Since Z_i and v_i are related by an onto mapping, revelation of v_i reveals Z_i perfectly. Therefore, a high Z_i school now realizes (a) that they were producing higher quality than expected *and* (b) their marginal cost of production is now lower. The first effect will reduce the equilibrium quality of the school, but the second will increase it; the ultimate quality and price will depend on the relative strength of these effects. For schools who are revealed to have low quality, the effects will be exactly the opposite. Critically, schools that see a quality increase should also see a price increase and those that see their quality decline should see a corresponding decline in price.

The fundamental difference with the asymmetric information model, where price and quality move in opposite directions, is that if information is symmetric, price and quality should always move together. Intuitively, the symmetric lack of information guarantees a tight correspondence between price and quality as there are no 'informational rents' required for incentive compatibility. These results are not consistent with our experimental results where we find an aggregate decline in prices and an aggregate increase in quality, pointing to a market failure arising from the lack of information in equilibrium at baseline.

APPENDIX III

Appendix III, Table I: Randomization Balance

	Control	Difference (Treatment - Control)
	(1)	(2)
Panel A: Village Level Variables		
Village Wealth (Median Monthly Expenditure)	4585.375	87.661
		(203.377)
Number of Households in Village	626.5	9.349
		(73.067)
Village Inequality (Gini Index)	0.533	-0.019
		(0.038)
Number of Government Schools in Village	4.125	0.425
		(0.372)
Number of Private Schools in Village	2.643	0.131
		(0.441)
Village enrollment % (All)	70.617	0.400
		(2.289)
Village enrollment % (Boys)	76.464	-0.455
		(2.005)
Village enrollment % (Girls)	64.106	1.389
		(2.820)
Level of Competition between Schools in Village (Herfindahl Index)	0.197	-0.005
		(0.014)
No. of Grade 3 Children Tested in Village	103.321	9.881
		(12.815)
Village Adult (>24 yrs) Literacy (%)	38.472	-2.441
		(1.910)
Panel B: School Level Variables		
School Average Test Score	0.028	0.001
		(0.062)
School Fees	510.934	-108.992
		(69.986)
Number of Students (Grades 1 to 5) Enrolled at School	91.613	-5.113
		(6.248)
Panel C: Child Level Variables		
Average Test Score	-0.013	-0.016
		(0.061)
Female Child	0.439	0.001
		(0.018)
Child Age	9.680	0.003
		(0.082)
Father's Education	2.206	-0.081
		(0.045)
Mother's Education	1.565	-0.002
		(0.045)
Wealth (Child Asset Index)	0.074	-0.173
		(0.130)

Notes:
This table presents balance checks on the village level randomization. We note that our sample is balanced everywhere except for one variable (Father's Education), which can be expected by random chance. Column 1 shows the raw mean of the variables for the control group. Column 2 tests the difference between treatment and control villages and controls for district stratification. Panel A considers village level variables; Panel B considers school level variables; and Panel C considers child level variables. Regressions for column 2 display robust standard errors for Panel A and clustered standard errors at the village level for Panels B and C. The p-value for the joint test of significance at the village level is 0.56.

APPENDIX III

Appendix III, Table II: Attrition Checks

	Attriters		Absentees		Dropouts		Untracked or Left Village	
	Control	**Difference**	**Control**	**Difference**	**Control**	**Difference**	**Control**	**Difference**
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Attrition Check by Treatment Status								
Report Card	0.181	0.008	0.09	-0.001	0.04	0.006	0.052	0.002
		(0.014)		(0.011)		(0.004)		(0.007)
Panel B: Differential Attrition across Baseline Characteristics								
Child Average Score	-0.116	-0.050	-0.190	-0.065	-0.150	-0.048	0.038	0.005
		(0.085)		(0.097)		(0.119)		(0.138)
English Score	-0.078	0.003	-0.165	0.033	-0.095	-0.071	0.089	0.052
		(0.094)		(0.097)		(0.111)		(0.171)
Math Score	-0.153	-0.104	-0.226	-0.162	-0.200	-0.004	0.009	-0.062
		(0.089)		(0.115)		(0.141)		(0.121)
Urdu Score	-0.118	-0.048	-0.178	-0.067	-0.155	-0.068	0.016	0.024
		(0.093)		(0.101)		(0.137)		(0.153)
Female	0.398	0.015	0.388	0.014	0.431	0.019	0.389	0.014
		(0.026)		(0.039)		(0.052)		(0.038)
Child Age	9.741	0.075	9.732	0.049	10.022	0.025	9.537	0.138
		(0.110)		(0.124)		(0.168)		(0.187)
Mother's Education	1.532	-0.009	1.518	-0.012	1.375	-0.003	1.651	0.020
		(0.066)		(0.085)		(0.106)		(0.130)
Father's Education	2.189	-0.119	2.176	-0.102	2.075	-0.172	2.281	-0.072
		(0.074)		(0.095)		(0.117)		(0.131)
Wealth (Child Asset Index)	-0.027	-0.174	-0.098	-0.115	-0.363	-0.241	0.283	-0.154
		(0.150)		(0.214)		(0.223)		(0.217)

Notes:

This table first checks whether child attrition is differential by treatment status (Panel A), and then checks for differences across baseline child and parental characteristics for the attriter group (Panel B). Note that a child attriter is defined as a child who was tested in the first year but not the second year. Columns 1 and 2 combine all attriters, while columns 3-8 divide attriters into different subgroups such as absentees, dropouts, and those who were untracked or left the village altogether. The odd number columns show the control group mean, whereas the even columns present coefficients from a simple difference (treatment-control) regression which includes district fixed effects and cluster standard errors at the village level. Panel A shows there is no differential attrition by treatment status, and Panel B shows that attriters are not different by treatment status across a range of child and household characteristics.

APPENDIX III

Appendix III, Table III: Robustness for Main Perceptions (Year 2) Result

	School Level			HH x School Level
	Preferred	**Mixed HH composition**	**Same Respondent**	**Preferred**
	(1)	(2)	(3)	(4)
Baseline Perception	0.228	0.231	0.120	0.127
	(0.0365)	(0.0382)	(0.0400)	(0.0170)
Baseline Fee	0.000129	0.000112	0.000154	0.000176
	(0.0000226)	(0.0000184)	(0.0000252)	(0.0000258)
Reportcard (RC)	0.00798	-0.0156	0.0182	-0.0459
	(0.0364)	(0.0317)	(0.0371)	(0.0371)
RC * Score	0.114	0.0784	0.101	0.0788
	(0.0438)	(0.0375)	(0.0514)	(0.0415)
Baseline School Score	-0.0279	-0.00614	0.0206	0.0346
	(0.0347)	(0.0256)	(0.0339)	(0.0292)
R-Squared	0.315	0.334	0.289	0.180
Observations	588	639	504	5441
Baseline Percceptions (mean)	3.28	3.25	3.26	3.26

Notes:

This table shows robustness of our main perceptions result from Table II, Column 4 to different samples and specification choices. Columns 1-3 are run at the school level, whereas Column 4 is run at the HH x School level. Column 1 simply replicates the Table II, Column 4 result for ease of comparison, where the average perception of a school is calculated using only those household-school combinations where we have perceptions data for both rounds. Column 2 perceptions are calculated by averaging across all households reporting on a school in a given round; this maximizes the sample utilized but gives us a mixed composition of households across years. Column 3 calculates average perceptions for a school only using those households where the same respondent was interviewed in both years. Column 4 simply runs the regression on the matched HH x School data across rounds without averaging across households for a school. All regressions include district fixed effects and village level controls, and cluster standard errors at the village level. Baseline Depvar (mean) displays the baseline mean of the dependent variable for the sample in these regressions.

APPENDIX III

Appendix III, Table IV: Test Scores across Subjects - Impact on Market Outcomes

	Test Scores (Year 2)		
	English	**Urdu**	**Math**
	(1)	(2)	(3)
Panel A: No Controls			
Report Card	0.129	0.105	0.149
	(0.0648)	(0.0578)	(0.0765)
Observations	112	112	112
R-Squared	0.338	0.369	0.377
Panel B: Baseline Control Only			
Report Card	0.0952	0.0931	0.138
	(0.0467)	(0.0404)	(0.0585)
Baseline	0.630	0.677	0.679
	(0.0617)	(0.0575)	(0.0752)
Observations	112	112	112
R-Squared	0.660	0.698	0.64
Panel C: Baseline and Village Controls			
Report Card	0.100	0.101	0.147
	(0.0482)	(0.0406)	(0.0588)
Baseline	0.624	0.675	0.678
	(0.0640)	(0.0590)	(0.0733)
Observations	112	112	112
R-Squared	0.662	0.707	0.644
Baseline Test Score (mean)	-0.044	-0.015	-0.037

Notes:

This table presents impact of the intervention on village average test scores for particular subjects (English, Urdu and Math). We observe gains between 0.09 - 0.15 standard deviations in treatment villages relative to control. The outcome variables are: Year 2 village level test scores in English (column 1), in Urdu (column 2), in Math (column 3). All regressions include district-fixed effects and robust standard errors. Panel A has no controls; Panel B includes baseline test scores as a control; and Panel C includes baseline test score as well as the same village controls as previous tables. Baseline Test Score (mean) displays the baseline test score mean for the sample for all outcome variables. Columns 1-3 are run on all 112 sample villages.

APPENDIX III

Appendix III, Table V: Village Enrollment Rate by Grades (Year 2)

	Preparatory	Class 1	Class 2	Class 3	Class 4	Class 5
	(1)	(2)	(3)	(4)	(5)	(6)
Report Card	0.178	0.0986	0.0161	0.0293	0.0678	0.0564
	(0.111)	(0.0407)	(0.0327)	(0.0374)	(0.0349)	(0.0518)
R-Squared	0.233	0.408	0.349	0.361	0.458	0.337
Observations	112	112	112	112	112	112

Notes:

This table disaggregates the enrollment result from Table IV to look at increase in enrollment rate by grade. The outcome variables are: Year 2 enrollment rate by each primary grade calculated at the village level. All regressions show robust standard errors.

APPENDIX III

Appendix III, Table VI: Fees - Impact by Baseline Test Score Dummy using Different Thresholds

	Fees (Year 2)								
	Village, p60			Village Median			Sample Median		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Report Card (RC)	-164.1	-92.24	-78.79	-60.56	-49.93	-40.89	-124.5	-78.54	-64.98
	(106.8)	(77.87)	(74.62)	(113.5)	(79.74)	(81.92)	(133.4)	(112.9)	(117.0)
High Scoring Private School	371.0	142.8	129.4	489.3	221.2	212.4	479.7	179.1	175.5
	(127.2)	(109.7)	(103.6)	(123.2)	(103.6)	(97.93)	(119.6)	(141.0)	(140.9)
RC* High Scoring Private School	-344.0	-277.7	-276.2	-426.9	-282.7	-277.5	-313.3	-224.6	-226.1
	(143.9)	(129.0)	(125.4)	(157.4)	(119.7)	(117.3)	(174.4)	(145.8)	(147.5)
Baseline Fee		0.685	0.699		0.674	0.685		0.679	0.690
		(0.111)	(0.113)		(0.113)	(0.116)		(0.117)	(0.120)
Controls	None	Baseline	Village	None	Baseline	Village	None	Baseline	Village
Observations	276	274	274	276	274	274	276	274	274
R-Squared	0.254	0.581	0.583	0.265	0.581	0.583	0.261	0.578	0.581
SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS									
Low Scoring Private School	-164.1	-92.24	-78.79	-60.56	-49.93	-40.89	-124.5	-78.54	-64.98
	[0.128]	[0.239]	[0.294]	[0.595]	[0.533]	[0.619]	[0.353]	[0.488]	[0.580]
High Scoring Private School	-508.1	-369.9	-355.0	-487.5	-332.6	-318.4	-437.8	-303.2	-291.1
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Baseline Fee (mean)	1188.5	1188.5	1188.5	1188.5	1188.5	1188.5	1188.5	1188.5	1188.5

Notes:

This table runs robustness checks on the heterogenous fee results in Table V by considering different definitions of high and low scoring schools and a series of controls. Columns 1 to 3 define schools as high scoring above the 60th percentile within the village. Columns 4 to 6 define high scoring as schools above the village median score. Columns 7 to 9 define high scoring as those schools above the median test score of the sample. Columns 1, 4 and 7 include no controls; Columns 2, 5 and 8 include baseline fee as a control; and Columns 3, 6 and 8 includes the same village controls as in previous tables in addition to baseline fee control. Regressions include district fixed effects and cluster standard errors at the village level. We have fewer observations than the total number of private schools in our sample due to private school closure and missing data in years 1 and 2. The lower panel displays the estimated coefficients and p-values [in square brackets] for relevant subgroups obtained from the coefficients estimated in the top panel. Baseline Fee (mean) displays the baseline mean fee for the sample in these regressions.

APPENDIX III

Appendix III, Table VII: Impact on Test Scores by Baseline Child Type and Provider Quality and Type

	Child Average Test Scores (Year 2)	
	Include Child Performance (below/above median)	Include Child Performance within School (below/above median)
	(1)	(2)
Report Card (RC)	0.383	0.336
	(0.173)	(0.177)
RC * Government School (Gov)	-0.290	-0.203
	(0.171)	(0.176)
RC * High Scoring Private School	-0.372	-0.375
	(0.194)	(0.186)
RC * High Achieving (Ach) Student in Sample	-0.180	
	(0.129)	
RC * Gov * High Ach Student in Sample	0.167	
	(0.160)	
RC * High Scoring Private * High Ach Student in Sample	0.110	
	(0.153)	
RC * High Ach Student in School		-0.0685
		(0.120)
RC * High Ach Student in School * High Scoring Private Schl		0.0366
		(0.131)
RC * High Ach Student in School * Gov		-0.0122
		(0.122)
Government School	-0.188	-0.261
	(0.0642)	(0.0672)
High Scoring Private School	0.001	0.140
	(0.0706)	(0.0747)
Baseline	0.622	0.606
	(0.0439)	(0.0447)
Controls	Village	Village
Observations	9888	9888
R-Squared	0.531	0.535

SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS				
	Low scoring private school, low ach child in sample	0.383	Low scoring private school, low ach child in school	0.336
		[0.0286]		[0.060]
	Low scoring private school, high ach child in sample	0.204	Low scoring private school, high ach child in school	0.267
		[0.005]		[0.001]
	High scoring private school, low ach child in sample	0.0113	High scoring private school, low ach child in school	-0.0393
		[0.871]		[0.460]
	High scoring private school, high ach child in sample	-0.0588	High scoring private school, high ach child in school	-0.0711
		[0.268]		[0.232]
	Government school, low ach child in sample	0.0933	Government school, low ach child in school	0.133
		[0.052]		[0.007]
	Government school, high ach child in sample	0.0802	Government school, high ach child in school	0.0523
		[0.335]		[0.297]
Baseline Test Score (mean)		0.009		0.009

Notes:

This table examines differential test score impact across types of school and child ability to understand whether initially low or high achieving children gain in initially high or low scoring schools. The outcome variable for all regressions is Year 2 child level average test score. Column 1 interacts the treatment status with school type and whether the child is above or below the sample median score. Column 2 repeats Column 1 with students split relative to the median score within the school they attend. High achieving student is a student whose scores are above median. All of the above regressions have standard errors clustered at the village level and include district fixed effects. In addition, they include baseline control of outcome variable and other village level controls. All columns also include interaction terms with NGO, as well as other interactions and level terms that are necessary given the interaction terms included. The lower panel displays the estimated coefficients and p-values [in square brackets] for relevant subgroups obtained from the coefficients estimated in the top panel. Baseline Test Score (mean) displays the baseline test score mean for the sample in these regressions.

APPENDIX III

Appendix III, Table VIII.A: School Inputs Detailed

	Basic Infrastructure Variables (Year 2)				Extra Infrastucture Variables (Year 2)					
	Has Desk/Chair	Class rooms per student	Toilets per student	Boards per student	Has a Library	Has a Computer	Has a Sports Facility	Has a Wall/Fence	Has Fans	Has Electricity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Report Card (RC)	0.015	-0.005	-0.002	0.005	-0.012	0.047	-0.022	-0.084	0.043	-0.058
	(0.0767)	(0.0049)	(0.0029)	(0.0065)	(0.1133)	(0.0614)	(0.0868)	(0.0401)	(0.0540)	(0.0445)
Gov	-0.293	-0.015	-0.004	-0.010	-0.065	-0.008	-0.119	-0.141	-0.169	-0.091
	(0.0631)	(0.0037)	(0.0021)	(0.0040)	(0.0909)	(0.0488)	(0.0609)	(0.0365)	(0.0428)	(0.0350)
High Scoring Private Schl	0.009	-0.003	-0.002	0.001	0.154	0.116	0.094	-0.049	-0.014	-0.029
	(0.0688)	(0.0042)	(0.0023)	(0.0050)	(0.1048)	(0.0621)	(0.0605)	(0.0314)	(0.0391)	(0.0281)
RC*Gov	-0.002	0.005	0.001	-0.003	0.046	-0.046	0.008	0.084	-0.004	0.065
	(0.0765)	(0.0049)	(0.0031)	(0.0067)	(0.1211)	(0.0619)	(0.0920)	(0.0560)	(0.0601)	(0.0473)
RC*High Scoring Private Schl	0.078	0.003	0.003	-0.011	-0.127	-0.037	-0.039	0.077	-0.013	0.065
	(0.0894)	(0.0060)	(0.0033)	(0.0077)	(0.1326)	(0.0754)	(0.1057)	(0.0513)	(0.0601)	(0.0483)
Baseline	0.510	0.703	0.572	0.662	0.370	0.628	0.468	0.552	0.697	0.803
	(0.0357)	(0.0594)	(0.0453)	(0.0673)	(0.0438)	(0.0501)	(0.0484)	(0.0362)	(0.0304)	(0.0269)
R-Squared	0.527	0.550	0.307	0.417	0.145	0.491	0.286	0.402	0.650	0.765
Observations	783	781	781	781	783	783	783	783	775	782
SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS										
Low Scoring Private School	0.015	-0.005	-0.002	0.005	-0.012	0.047	-0.022	-0.084	0.043	-0.058
	[0.846]	[0.348]	[0.540]	[0.400]	[0.919]	[0.450]	[0.804]	[0.039]	[0.429]	[0.194]
High Scoring Private School	0.093	-0.001	0.001	-0.006	-0.139	0.009	-0.061	-0.007	0.030	0.007
	[0.054]	[0.719]	[0.416]	[0.204]	[0.047]	[0.839]	[0.298]	[0.796]	[0.252]	[0.702]
Government School	0.013	0.000	-0.001	0.002	0.034	0.001	-0.013	0.000	0.039	0.006
	[0.703]	[0.808]	[0.413]	[0.198]	[0.403]	[0.938]	[0.660]	[0.988]	[0.127]	[0.788]
Baseline Depvar (mean)	0.484	0.033	0.013	0.0414	0.188	0.088	0.158	0.782	0.623	0.676

Notes:

This table shows impacts on school input variables by school type and baseline performance. All outcome variables are measured in year 2. Basic and extra infrastructure variables are the individual components of the average effect size (AES) regressions in Table VIII, Panel A. Basic infrastructure includes: A dummy for whether school has desk/chair as sitting arrangement relative to floor or mats (column 1); Number of classrooms per student (column 2), toilets per student (column 3), Black boards per student (column 4). Extra infrastructure variables are dummies for whether school has: library (column 5); computer facility: (column 6); sport facility (column 7); wall/fence (column 8); fans (column 9); or electricity (column 10). All regressions include district fixed effects and cluster standard errors at the village level. We have fewer observations than the total number of schools in our sample due to school closure and some missing data. The lower panel displays the estimated coefficients and p-values [in square brackets] for relevant subgroups obtained from the coefficients estimated in the top panel. Baseline Depvar (mean) displays the meanfor the baseline of the dependent variable for the sample in these regressions.

APPENDIX III

Appendix III, Table VIII.B: Household Inputs Detailed (Parent-Interaction Variables)

	Whether Parent Has Ever Met Class Teacher (Year 2)	Parental Knowledge of Class Teacher Name (Year 2)	Parental Assessment of Class Teacher Involvement (Year 2)
	(1)	(2)	(3)
Report Card (RC)	0.0763	0.126	0.209
	(0.105)	(0.128)	(0.0758)
Gov	-0.150	-0.0759	0.164
	(0.0648)	(0.102)	(0.0660)
High Scoring Private Schl	0.123	0.116	0.0184
	(0.0797)	(0.110)	(0.0714)
RC*Gov	-0.0149	-0.0433	-0.175
	(0.113)	(0.133)	(0.0876)
RC*High Scoring Private Schl	-0.154	-0.0906	-0.236
	(0.130)	(0.153)	(0.105)
Baseline	0.0887	0.166	0.0237
	(0.0379)	(0.0341)	(0.0303)
R-Squared	0.0806	0.0988	0.0708
Observations	954	958	690
SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS			
Low Scoring Private School	0.0763	0.126	0.209
	[0.468]	[0.327]	[0.0069]
High Scoring Private School	-0.0781	0.0353	-0.0275
	[0.295]	[0.661]	[0.720]
Government School	0.0614	0.0826	0.0340
	[0.156]	[0.052]	[0.361]
Baseline Mean (Depvar)	0.650	0.498	1.796

Notes:

This table shows impacts on additional household input variables by school type and baseline performance, specifically those variable that are used in the average effect size regression in Table VIII, Panel B. All outcome variables are measured in year 2 and include: Parental knowledge/view of the class teacher's involvement (column 1); Whether a parent has ever met their child's teacher (column 2); and Parental knowledge of teachers name (column 3). These data come from the household survey for children who were matched to the school testing roster, and is at the household X school level. All regressions include district fixed effects and cluster standard errors at the village level. The lower panel displays the estimated coefficients and p-values [in square brackets] for relevant subgroups obtained from the coefficients estimated in the top panel. Baseline Depvar (mean) displays the meanfor the baseline of the dependent variable for the sample in these regressions.

APPENDIX III

Appendix III, Table IX: Test Score and Fee Impact by Heterogeneity in Market Competition

	Child Test Score (Year 2)	Private School Fees (Year 2)
	(1)	(2)
Report Card (RC)	0.402	-58.59
	(0.175)	(109.3)
Government School	-0.230	
	(0.0847)	
High Scoring Private School	0.203	214.1
	(0.0891)	(153.8)
RC * Government School (Gov)	-0.310	
	(0.176)	
RC * High Scoring Private School	-0.461	-331.5
	(0.188)	(175.0)
RC * Low Competition	-0.256	12.03
	(0.203)	(164.0)
RC * Gov * Low Competition	0.253	
	(0.212)	
RC * High Scoring Priv * Low Competition	0.301	93.80
	(0.221)	(287.6)
Baseline	0.766	0.678
	(0.0173)	(0.120)
R-Squared	0.541	0.588
Observations	9887	274
SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS		
Low scoring private school, high competition	0.402	-58.59
	[0.0233]	[0.593]
Low scoring private school, low competition	0.147	-46.56
	[0.154]	[0.714]
High scoring private school, high competition	-0.0590	-390.1
	[0.442]	[0.001]
High scoring private school, low competition	-0.0138	-284.3
	[0.839]	[0.0387]
Government school, high competition	0.0926	
	[0.160]	
Government school, low competition	0.0902	
	[0.171]	
Baseline Depvar (mean)	0.009	1188.5

Notes:

This table examines score and fee impact across school types and market competitiveness to understand how schools in different competitive environments responded to the intervention. The outcome variables are: Year 2 child level average test score (column 1); and private school fees in levels (column 2). Both columns interact treatment status with school type and our measure of competition, a dummy for whether the Herfindahl index in the village is above the sample village median. Column 1 is at the child level, and Column 2 at the school level restricting to private schools. All regressions cluster standard errors at the village level, include district fixed effects, and baseline control of outcome variable and other village controls. All columns also include interaction terms with NGO, as well as other requisite interactions and level terms. The lower panel displays the estimated coefficients and p-values [in square brackets] for relevant subgroups obtained from the coefficients estimated in the top panel. Baseline Depvar (mean) displays the baseline mean for the dependent variable for the sample in these regressions.