Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets
Faculty Research Working Paper Series

Tahir Andrabi

Pomona College

Jishnu Das

World Bank Development Research Group

Asim Ijaz Khwaja

Harvard Kennedy School

Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets

By *Tahir Andrabi, Jishnu Das and Asim Ijaz Khwaja*[*]

June 2014

We study the impact of providing school and child test scores on subsequent test scores, prices, and enrollment in markets with multiple public and private providers. A randomly selected half of our sample villages (markets) received report cards. This increased test scores by 0.11 standard deviations, decreased private school fees by 17 percent and increased primary enrollment by 4.5 percent. Heterogeneity in the treatment impact by initial school quality is consistent with canonical models of asymmetric information. Information provision facilitates better comparisons across providers, improves market efficiency and raises child welfare through higher test scores, higher enrollment and lower fees.

JEL Codes: O12, I25, L15, L22, D22, D82
Keywords: Educational Markets, Information Provision, Private Schools, Market-level Experiments

It is a widely held belief that providing information to citizens is a powerful tool for improving public services. This is particularly prevalent in the education sector, where advocates claim that informing parents about school performance is key to improving school quality (World Bank, 2004; Hoxby, 2002). These beliefs underlie policies such as "No Child Left Behind" in the

---

United States, which requires states to regularly report and disseminate information on educational achievement.

The empirical evidence on the impact of information provision regarding provider quality, however, is mixed. Depending on the setting, the extent to which the information was bundled with other accountability measures, and the type of response that was studied, effect sizes on the impact of information range from zero to highly positive.[1] Worryingly, high-stakes information can also create incentives for manipulation through the selection of more "desirable" consumers (Dranove et al., 2003) or through cheating and direct manipulation (Jacob & Levitt, 2003; Figlio & Getzler, 2006). While these studies vary in terms of context, a common feature of the environments studied is that there are either no prices or prices are highly regulated.

This paper contributes to the literature by studying the impact of providing information in the presence of both a public sector *and* a (competitive) private market for schooling. This is important for two reasons. First, such market settings are becoming increasingly common in the education as well as health and financial sectors of many developing countries.[2] Second, in such settings the impact of information provision depends critically on the pre-existing equilibrium. Canonical models of asymmetric information recognize that prices adjust endogenously to mitigate the adverse impacts of poor information

---

[1] Information on school-level test scores did not have an impact in two studies in India and Chile (Banerjee et. al., 2010; Mizala & Urquiola, 2013), even though information given to households on the returns to education had a large effect on enrollment (Jensen, 2010). In the United States, the joint impact of information, accountability, and sanctions improved outcomes for poorly performing public schools (Rockoff & Turner, 2008; Chiang, 2009), and information on school rankings altered parental choices (Hastings & Weinstein, 2008). In health, hygiene report cards in restaurants improve food safety, particularly in those establishments that start off with poor results (Jin & Leslie, 2003) and a study from Uganda showed large impacts of an intervention that bundled information and community empowerment (Björkman & Svensson, 2009).

[2] In India and Pakistan, 40% of primary enrollment is in private schools (ASER India, 2012; ASER Pakistan, 2012). In Sub-Saharan Africa, private primary enrollment is estimated at 16.5%, despite these administrative data having been shown to underestimate participation in private schooling (Dixon, 2012).

environments. As outlined below, depending on the initial equilibrium, additional information provision may lead to diametrically opposite effects on provider price, quality and the new market equilibrium. Our setting allows us to more broadly test the theoretical predictions derived from such models and thereby better understand the impact of information provision in these more complex, but realistic and increasingly prevalent, environments.

We analyze data from a market-level experiment in 112 Pakistani villages that increased information exogenously through the dissemination of school and child level test scores. Importantly, these villages contain both public and private schools, with an average of 7 schools per village. We provided households and schools in treatment villages with a report card that displayed the test scores of their own children and all schools in the village. Each village can be regarded as an "island economy," as children rarely attend schools outside of the village. This context combined with limited central regulation means private school prices are independently determined within each village. Thus, each village forms a unique competitive market. Since the village is also our unit of treatment, we are therefore able to study the average impact of information on the schooling market as a whole in addition to the heterogeneous impact on particular schools.

We first confirm that parental knowledge indeed improved as a result of the intervention. In treatment villages, perception of quality became better aligned with school test scores than in control villages. Learning also greatly improved: In treatment villages, the average child's test score increased by 0.11 standard deviations or 42 percent of the average yearly gain; this persisted for at least two years after the intervention. In addition, (private) school fees fell in treatment villages by 17 percent relative to schools in control villages. Faced with improved quality and/or lower fees, overall enrollment among primary-age children rose by three percentage points in treatment villages. Moreover, (private) schools with low baseline test scores were more likely to shut down in treatment villages, with

3

their students shifting into alternate schooling options.[3] These average impacts are substantial relative to a range of (typically costlier) educational interventions in other low-income environments (McEwan, 2013).

This market level experiment in the presence of private (fee-charging) providers also allows us to interpret our results and explore further heterogeneous impacts in the context of canonical models of optimal pricing and quality choice in markets with asymmetric information (Wolinsky, 1983; Shapiro, 1983; Milgrom & Roberts, 1986). These models recognize that even in the absence of third-party information, consumers receive partially informative signals of firm (in our context, school) quality and that quality is malleable and can be increased through costly investments.[4]

A standard result is that when the original signal is somewhat informative, the initial equilibrium may be separating with high quality schools charging high prices. This equilibrium is sustained through even *higher* prices for the high quality schools relative to the counterfactual of perfect information.[5] Furthermore, the distribution of school quality may differ than under perfect information; particularly, it could display excess quality differentiation. In such a separating equilibrium, when information improves, such as through the provision of report cards, the markup required to sustain separation falls: Therefore, report card provision *decreases* average price, with greater declines for higher quality schools. The models would additionally predict that this price effect reduces the

---

[3] The fee and school closure results are for private schools only since public schools do not charge any fees and rarely (if ever) shut-down.

[4] In our context, parents rely on informal monitoring, the schools' own tests and their own assessments of child performance to judge school quality. In one example, an illiterate mother gauged literacy skills by reciting poems to her child, asking the child to write them down and then read them back to her. In addition, schools typically improve their quality level through investments in teachers, infrastructure and educational materials and methods.

[5] To induce a separating equilibrium, the high achieving school must earn a markup over and above the cost of production. Schools have incentives to not cheat (i.e. charge high prices while producing low quality) because they lose the markup on the fraction of consumers who correctly infer quality from their information signal and therefore choose not to buy the product.

price-quality gradient and excess differentiation in the market.[6] Moreover, if the cost of increasing quality is convex in initial quality levels, then the reduction in excess differentiation will likely come from quality improvements among initially low achieving schools. We find support for all these predictions.

First, both the baseline and experimental evidence suggest that schools were initially in a separating equilibrium. We find that the schools' baseline test scores were highly correlated with both their baseline price and the households' perception of school performance, even after accounting for village fixed-effects and a set of parental attributes. Second, given that parents (correctly) update their beliefs as a result of the intervention, we can directly test the predictions of the model by exploring whether heterogeneous impacts exist by school baseline test scores. We indeed find that the price-quality gradient declines in treatment villages and that this is due to greater declines in the prices of high achieving private schools. Finally, we also find a reduction in excess differentiation that is driven by the initially low quality schools. The test scores of children in initially low achieving private schools rose by 0.31 standard deviations relative to the control, while those in high achieving schools did not experience any changes.

We also document additional test score gains of 0.10 standard deviations for children in public schools in response to the intervention. Although public schools face few market or administrative disciplining mechanisms, social (non-price) disciplining actions that the community can take may alter teacher effort and other inputs in these schools. Importantly, these test score gains did not differ by the baseline test scores of the public school (if anything, initially low quality public schools have smaller gains compared to initially high quality public

---

[6] If the original information signal is very imprecise, the markup required to sustain separation for any two "similar" quality levels may be too high given the underlying demand for the product. In this case, the initial equilibrium collapses to a "pooling" equilibrium. If so, providing better information would (in contrast) lead to (some) schools changing their price and quality levels, inducing greater quality separation between schools and the emergence of a price-quality gradient.

schools), suggesting that the private school impacts were driven by the structure of the market equilibrium rather than a mechanical relationship between baseline test scores and subsequent improvements following the provision of information or simply due to changed parental investments.[7]

Our study contributes to the literature on information and the quality of services (see Dranove & Jin, 2010 for a recent review in education and health sectors) by providing, to our knowledge, the first *market*-level experimental evidence on the impact of information on market outcomes.[8] Banerjee et al. (2010) and Bjorkman and Svensson (2009) assess whether information leads to consumers demanding better services from public providers. Similarly, Hastings and Weinstein (2008) demonstrate parental responsiveness to information (when provided with school rankings, parents with high scoring alternatives nearby change their declared choice toward such higher scoring schools, leading to higher test scores). Our paper builds on this work by showing the role of information in markets where private providers change prices to optimally adjust to an improvement in the quality signal.[9]

---

[7] The combination of test score and price changes suggests that schools altered their investment as a consequence of the report cards. However, test scores are jointly produced by school and household/child-level investments. Using household survey data, we can also examine parental involvement in education, time use and expenditure. We find no evidence of parental input changes (see Online Appendix Table VI). We also study changes in school inputs (also see Online Appendix Table VI). While we find no evidence of changes in teacher quality or indices of infrastructure in private schools, we do find evidence of reduction in break time in low-achieving private schools. This is plausible given the context of constraints in teacher supply and limited financial resources, which imply that changing teachers or improving infrastructure may be less feasible; however, low-achieving schools may still be able to increase quality relatively cheaply by reducing break/play time in order to increase overall instruction time.

[8] A second class of papers looks at the impact of information in markets where there is uncertainty over supply and demand. Jensen (2007), for instance, shows that the introduction of cell phones leads to price convergence across different markets for fish, because fishermen are better able to choose the market that they take their catch to.

[9] Hastings and Weinstein's empirical strategy—using choice models to look at demand-side responses to information provision—also follows from their specific context, where prices are administratively determined and school-level responses are unlikely in the short-term. In contrast, our setting allows for supply-side responses both in prices and quality and among both public and

Three examples of non-experimental but related studies in similar market settings are Figlio and Lucas (2004), Mizala and Urquiola (2013), and Jin and Leslie (2003). Figlio and Lucas (2004) show that house prices are sensitive to school rankings, suggesting that parents are willing to pay more to send their child to schools with higher test scores. However, Mizala and Urquiola (2013), using a regression discontinuity design, find little impact of school rank on enrollments or prices. They rightly caution that they cannot capture the effect of new information in markets since their unit of comparison is not *across* markets that differ in the extent of third-party information, but rather across providers *within* a market, which faces the same informational environment. A closer analogue to our paper is Jin and Leslie's (2003) study on the introduction of restaurant hygiene cards in a single market, Los Angeles, and its impact on food safety and restaurant revenues. They document positive effects, especially among initially poor performers. While our results use an experimental design from a different setting that is able to exploit variation across multiple markets, it is nevertheless noteworthy that the pattern of heterogeneity they uncover is similar to ours.

In short, we show that prices are indeed a key component of how markets react when information improves, and that price changes are differential across initial school quality in a way that is consistent with the predictions of canonical models of asymmetric information. These insights are important in helping develop a more nuanced understanding of the impact of informational provision in markets with multiple (public and private) providers, and how this may vary based on the pre-existing informational environment. Information provision in our setting improves consumer (parent) welfare by lowering markups and inducing lower quality schools to improve quality or shutdown. Public schools also respond positively by raising quality and overall village enrollment increases.

---

private schools, leading us to focus more on the overall structure of the market and how it changes with the provision of information.

7

The remainder of the paper is structured as follows: Section I provides details on the data, the context and the report card intervention. Section II describes the conceptual and empirical framework. Section III presents the findings. Section IV discusses these results further and concludes.

I. Data, Context and Intervention

Private schooling has increased dramatically in low-income countries. In India, private school enrollment doubled between 1993 and 2005, with a market share of above 50 percent in urban areas (Andrabi et al, 2010). Similar patterns are observed in African countries; for example, in The Gambia, private school primary enrollment increased from 14 to 28 in just over 10 years (World Bank Data, 2014). In Pakistan, the setting of this study, the number of private schools increased sharply from 3,800 in 1983 to 47,000 in 2005; such schools currently account for a third of all primary school enrollment. These private schools are co-educational and instruct children in English, Mathematics, and Urdu using a curriculum and textbooks similar to that in public schools. In contrast to public schools however, private schools face little government oversight or regulation and operate in (de facto) lightly regulated markets with no administrative guidance on pricing. Sixty percent of the rural population in the province we study resided in a village with at least one private school in 2001, with on average 7.3 schools in these villages. Thus, parents face substantial choice for their children's education. We designed our study around the particular opportunities and challenges represented by this increasingly common choice-rich environment.

A. Data

The data come from the Learning and Education Achievement in Punjab Schools Project (LEAPS, www.leapsproject.org), a multi-year study of education

in Pakistan. For the LEAPS project, we sampled 112 villages across three districts in the Punjab province, the largest state with a population of 70 million in 2010.[10] Using a household census of schooling choices, we verified that these villages were effectively "closed" markets with children attending the schools in the village and school populations drawn from children in the village.[11] We included all schools in these villages that offered primary education in our sample, resulting in over 800 public and private schools (see Online Appendix I.A for sampling details).

In each of these villages, we conducted a series of annual surveys starting in 2004. First, we tested children who were in Grade 3 in the first survey round and continued to track and test them in each subsequent round. Around 12,000 children were tracked and tested in each round with a tracking rate over 96%.[12] Second, we conducted yearly surveys for all schools. These surveys contained a number of modules, including a facility survey, roster data on around 4900 teachers and detailed surveys for head teachers and Grade 3 teachers. Third, for all tested grades, we administered a short child questionnaire to 10 randomly selected children in order to collect household-level information (6,000 children).

---

[10] One district was chosen from each of the province's three socioeconomic zones. 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).

[11] The villages are geographically separated by farmland, forests or wasteland, and distance to school is the primary determinant of enrollment and school choice. The census confirmed that 92% of children in the sampled villages attended the schools in our study; the remaining 8% were typically older children who had graduated to middle school by the time of our study (2% were studying in locations much farther away, including outside the country).

[12] We developed norm-referenced tests for English, Mathematics and Urdu (the vernacular) to maximize variation across the population of test-takers. The test was piloted to improve the validity and reliability of the instrument and ensure that it was easy to understand, administer, and unbiased in terms of socioeconomic background. Preference was given to formats familiar to children and items were refined using differential item functioning (Andrabi et al., 2002, discuss the psychometric properties of the test, including tests of reliability). 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 (see Das and Zajonc, 2010).

Finally, we conducted surveys with parents separately from the schools. This household questionnaire, with an extended focus on educational investments, was fielded for 1,800 households in the sample villages and stratified to over-sample students eligible by age for Grade 3 (the tested grade). These three data sources allow us to triangulate self-reported data from multiple sources and investigate the role of school and household inputs. Here, we use data from the first two rounds of the LEAPS surveys, augmented to check for longer-run effects with data from the third round. Online Appendix I.B provides further details on the content of and respondents for the different school- and household-based surveys as well as their timing across rounds.

On average, there are 631 households in a village with an adult literacy rate of 37.3 percent (Table 1). Among children between the ages of five and fifteen, baseline enrollment rates (public and private) were 76% for boys and 65% for girls in 2004. Schools in our sample enroll an average of 166 children, with 18 children enrolled in the tested Grade 3. Teacher quality is relatively low: just over half the teachers in these schools report more than a secondary education. The enrolled children in Grade 3 are on average 9.7 years old and 55.7% are male.

B. Patterns in the baseline data

Households spend 3-5 percent of their monthly budget on each child's schooling, with private school fees averaging about Rs. 1,200 per year. Analysis of choice suggests that while parents take into account school fees and infrastructure, the distance to school remains a major determinant of their choices (Carneiro et al., 2013). For example, increasing the distance of the nearest school from the home by 500 meters (adjusting for demographics) reduces enrollment by 1.5 to 3 percentage points for boys and 9 to 11 percentage points for girls (Andrabi et al., 2007), a large effect that is also replicated in the specific choice of school (Carniero et al., 2013). The importance of distance, documented across

numerous studies, underscores why these villages are effectively closed educational markets, thereby allowing us to study market-level interventions (Andrabi et al., 2007; Alderman, Orazem, & Paterno, 2001; Burde and Linden, 2012).

There are strong indications that the environment is competitive, with schools offering vertically differentiated products. Private schools locate within denser settlements in villages; the average private school has at least three schools around it; the Herfindahl index is consistent with a competitive environment; and, the median profits of Rs. 14,580 ($200) of private schools is similar to the wages of a male teacher with secondary education and therefore the appropriate option value if the entrepreneur were to shut down the school.[13] Although the student population differs slightly across schools, there is little evidence that these are segmented markets, either by wealth, parental education, or social variables such as caste (Andrabi et al., 2007).

While learning levels are generally low, there is substantial variation in test scores and prices with most of the variation *across schools* and *within villages*. Variation in test scores *within villages* accounts for 83% of the total test score variation in our data.[14] Part of this variation is driven by differences across public and private schools, but even across private schools, the inter-quartile range for test scores lay between –0.08 and 0.78 standard deviations in Mathematics, with similar results for other subjects. Similarly, within the same village there are large differences in the prices offered by private schools.

[13] We compute a Herfindahl index of 0.20 for the sampled villages. With an average of 7 schools in every village, exactly equal enrollment shares (the most competitive scenario) would imply a Herfindahl value of 0.14.

[14] 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*."

Average prices are low, with a median annual fee of Rs.720 ($1 per month) in the year 2000, which was lower than the daily wage rate for unskilled labor (PEIP, 2000).[15] The inter-quartile range of prices for private schools lies between Rs.650 and Rs. 1,350 (per year), and 45% of the variation in prices is *within* rather than across villages.

Previewing a subsequent discussion on whether these markets are in a separating equilibrium prior to the information intervention discussed below, test scores and fees are positively correlated at baseline. A one standard deviation increase in baseline test scores is associated with a 0.45 standard deviation (Rs 369) increase in school fees (Table II, Column 1). The result is similar if we include village fixed effects and demographic characteristics including household wealth and education. Test scores better predict school fees than infrastructure. However, infrastructure type does matter: The equivalent impact of a one standard deviation increase in an index of basic infrastructure (rooms, chairs, blackboards, and other non-building material) leads to a 0.07 standard deviation increase in fees (Rs 55), while a one standard deviation increase in advanced infrastructure (such as a library, fans, or computer facility) leads to a 0.17 standard deviation (Rs 141) increase (Table II, Column 2).

C. Intervention and Experimental Protocol

In 2004, we tested all children in Grade 3 in all the schools in our sample. We then experimentally allocated half the villages (within district stratification) to receive report cards on child and school performance. The two-page report card reported raw test scores for the child in English, Mathematics and Urdu as well as

---

[15] Low fees are maintained by keeping costs—primarily teachers' salaries—low. Salaries for teachers in the private sector are 20-25% of those in the public sector and, in 2003, the median salary was Rs.1100 ($11) a month. This model relies on the availability of locally-resident secondary school-educated women as potential teachers since geographic and occupational immobility for women leads to a 30% wage discount relative to men (Andrabi et al., 2013).

her quintile rank across all tested children on the first page. The second page reported scores for all the schools in the village, with their quintile rank (across all schools tested in the sample) and the number of children tested. The report cards were delivered to schools and parents at a school meeting, which confined itself to only explaining the information on the report cards and not to advocate or discuss any particular plan of action. The meetings were held in September 2004, after the summer break and prior to the next regular admission cycle in April 2005.

The timing of the report card delivery has implications for child switching behavior. While children can switch schools right after summer break (the timing of our delivery), most choose to do so when the new school year starts in April. Consequently, our timing decision may imply less switching relative to delivery before the new school year. However, the gap between information revelation and the next year's admission decisions also gave parents sufficient time to absorb the information and schools sufficient time to respond to it. From a welfare and policy point of view, it may be more desirable to give schools time to respond to information by altering their price and investing in quality, as opposed to encouraging parents to immediately exit schools with low test scores.

At the time of distribution, schools and households were explicitly informed that the exercise would be repeated a year later to ensure that educational investments would be captured in future test scores. This implied that parents and schools would be able to verify how test scores changed over the year, allowing parents to give a school more time to improve before withdrawing their children. Nevertheless, a limitation of our study is that we cannot assess the sensitivity of our findings to the timing of information provision.

Online Appendix I.C provides the detail of the experimental protocol including the design, content, and delivery of the report cards along with a discussion of the validity and the reliability of the test score measures. We also confirm that the baseline values of outcomes and control variables are balanced

13

across the treatment and control villages and attrition is unlikely a concern (see Online Appendix I.D). Specifically, while we are able to track over 96 percent of children between years, we find that absenteeism leads to somewhat lower retesting rates (82 percent of children tested in the baseline are retested in the second year). We confirm further that there is no evidence of differential attrition or any compositional (demographic or baseline test score) differences between attriters in treatment versus controls villages.

II. Conceptual and Empirical Framework

A. Conceptual Framework

In order to understand how report card delivery can impact the market, we outline a canonical framework of market equilibrium under asymmetric information. Online Appendix II illustrates this using Wolinksy's (1983) model.

The main theoretical insight is that the impact of the intervention depends on the pre-existing informational environment (regarding school quality) and, more specifically, on whether schools were pooling or separating on quality in the initial equilibrium. The theory leads to testable predictions on how the price-quality gradient changes due to treatment and, relatedly, whether we would expect differential fee impact by school quality. The theoretical predictions on how test-scores (school quality) respond to information and whether such responses differ by initial school quality are ambiguous since they depend on the specific structure of demand. Nevertheless, the theory will help in interpreting such additional results. In this section, we summarize the intuition and refer the reader to the online appendix for further details.

Consider a case where a school is deciding whether to produce at *H(igh)* or *L(ow)* quality, with associated costs of production, $C_H > C_L$. There is a pre-existing option (public or private) that offers L quality at a price P_L. Suppose that

there is sufficient heterogeneity in parent's quality valuation such that if quality were perfectly observable, this school would offer quality *H* at $P_H > P_L$.

First, consider the scenario where parents' signals of school quality are extremely poor, such that they can never receive an information signal sufficiently precise to distinguish H from L. In such a case, no parent will be willing to pay the higher price required and so the market will have a single product in both schools, quality L with price P_L; the general case of such a market collapse with multiple schools is Akerlof's (1970) model of lemons. In such a situation, report cards, provided they are sufficiently informative, can allow a school to credibly signal its higher quality and as such allow for a new equilibrium where a higher priced and higher quality school emerges. In this case, both average price and quality in the market would increase as a result of information provision.

Now consider what we believe is a more realistic scenario (and more consistent with evidence from our setting) —that even in the absence of third-party information, parents receive (imperfect) signals of school quality. A separating equilibrium can now arise, where a school can (profitably) produce at quality *H* and charge a higher price P_H (i.e. we are initially in a separating equilibrium). The central intuition in the canonical models of asymmetric information is that prices can substitute for the lack of information in determining the market equilibrium and schools can (credibly) signal higher quality by charging a higher price. In this separating equilibrium, schools can be induced not to "cheat" (charge a higher price, but produce a lower quality) provided that they earn a markup above the price under perfect information and there exists a set of parents who receive the correct signal and will leave if the school cheats. This is due to the usual incentive compatibility constraint that equates the marginal loss from cheating to the gain. Mathematically, we can show that this implies that $P_H = C_H + (C_H - C_L)(1 - D/D)$, where D is the fraction of parents who receive the correct (fully revealing) signal.

This simple framework suggests that not only will higher quality schools charge a (greater) price markup initially, but to the extent that report card delivery improves parents' ability to perceive true quality (i.e. "D" increases), the markup needed to sustain the equilibrium will also fall. In other words, report card provision will "flatten" the price-quality gradient in the market and result in higher quality schools decreasing their prices more. In this case, average price in the market overall will also fall.

Beyond these price predictions, one can also obtain, under plausible conditions, excess differentiation in the market in a separating equilibrium. This follows if parents lack the ability to distinguish between two similar quality schools (i.e. D is small for similar quality values). Thus, the markup that is required to sustain separation may be high enough that no parents will choose the higher quality school. However, as these quality levels diverge and parents can better differentiate quality, a separating equilibrium becomes possible. In this case, report cards reduce this excess quality differentiation by allowing parents to better distinguish between (nearby) quality levels. Predicting whether such excess differentiation reduction will be due to lower quality schools upgrading or higher quality ones downgrading depends on the underlying structure of the demand and cost conditions. For example, if quality improvements are cheaper at lower quality levels and there is parental demand for intermediate quality, the report card can lead to lower quality schools improving their quality (see Online Appendix II). This implies that even in the case of a reduction in excess differentiation, we obtain ambiguous predictions on the impact on average school quality in the market.

Finally, the canonical models have little to offer about the performance of public schools without knowing their objective function and the extent to which they face market discipline. To the extent that public schools face little competitive or regulatory pressure to perform, information may have little or no

16

impact (as in Banerjee et al., 2010). Alternatively, even in the absence of such reward systems, there maybe be non-monetary mechanisms such as social/community pressures on public school teachers that may induce performance improvements if a school is revealed to have low test scores.[16]

Classic models of asymmetric information can help us understand the overall impact of providing information through school report cards, as well as provide predictions for how different types of schools may react to improved parental knowledge. While for outcomes like schools fees we have clearer predictions based on the initial informational environment and equilibrium, for others, like the distribution of school quality and the public sector response, the predictions are more ambiguous. We therefore explore these issues empirically.

B. Empirical Framework

We estimate the causal effect of the report card treatment on key outcome variables, such as test scores, fees or enrollment.

We present our main results at the village level. Our preferred estimating equation is:

$$Y_{i2} = \alpha_d + \beta \cdot RC_i + \gamma \cdot Y_{i1} + \delta \cdot X_{i1} + \varepsilon_i$$

where Y_{i2} is the outcome of interest, for example, average (across all children in the village) test scores from the post-intervention year (year 2) in village i; RC_i is the treatment dummy assigned to village i; α_d are district fixed-effects; Y_{i1} is the baseline measurement of the outcome variable; and X_{i1} is a vector of village

---

[16] The standard model allows us to incorporate the public school as an "outside" option whose quality is directly affected by the report card through social or regulatory pressures. Any such change in public school quality has a direct impact on the distribution of private school quality and price; in particular, given that the public schools are typically of lower quality in our data, their presence in the market will generate greater pressure on the low-quality private schools to increase quality. Because these low-quality private schools are also in a separating equilibrium relative to the public schools, in this case we would also expect their prices (i.e. markups) to decline for the same reason they did for high quality private schools (although not by as much as for the latter).

level baseline controls (village size, village wealth, adult literacy etc.). Under random assignment, β is an unbiased estimate of the impact on test scores associated with the report card intervention. While we prefer including baseline controls to improve the precision of the estimates, we also present more parsimonious specifications (without any controls, and only controlling for baseline value of the dependent variable) for completeness. We include district effects in all specifications since the randomization was stratified by district.

The conceptual framework suggests that the reaction to the information will differ by the schools' baseline quality. To examine this, we also estimate models with treatment effects separately for the school's type (private or public) and baseline test score. These specifications are estimated at the school level with standard errors clustered by village.[17] A generic specification is:

$$Y_{ij2} = \alpha_d + \beta_0 RC_i + \beta_1 GOV_{ij} + \beta_2 HIGH_{ij1} + \beta_3 RC_i \cdot GOV_{ij} + \beta_4 RC_i \cdot HIGH_{ij1}$$
$$+ \beta_4 RC_i \cdot GOV_{ij} \cdot HIGH_{ij1} + \gamma \cdot Y_{ij1} + \delta \cdot X_{i1} + \varepsilon_{ij}$$

where Y_{ij2} represents the outcome of interest such as fees (private schools only), test scores, enrollment etc. for school j in village i in time period 2 (post-intervention year). As before, RC_i is the treatment dummy assigned to village i, α_d are district fixed-effects; Y_{ij1} is the baseline of the outcome variable; and X_{i1} is the vector of baseline village level controls. GOV_{ij} is a dummy indicator for whether the school is a public school, and $HIGH_{ij1}$ is an indicator for whether the school baseline score was above ("high") or below ("low") a pre-defined baseline test score threshold.

---

[17] In reality, we also separate out the very small number (16) of schools run by Non-Governmental Organizations in the specification. However, we have suppressed these terms in both the specification given above and when we present our estimated effects since the NGO-run sample is too small to allow for meaningful comparisons.

III. Results

The conceptual framework demonstrates that the impact of information provision depends on whether the initial equilibrium was separating or not. As we discussed previously, the correlation between prices and quality in the baseline data (Table II, Columns 1 and 2) suggested a separating equilibrium. We now turn to the empirical evidence, starting with the report card impact on household perceptions. We then examine the impact on school fees, test scores and enrollment at the village level. Finally, we turn to the more specific model predictions, including heterogeneous impacts across different types of schools, and interpret them in light of the conceptual framework above.

A. Impact on Perceptions

We first test whether perceptions/signals of school quality are correlated to school test scores at baseline. Table II, Column 3, finds that a one standard deviation increase in test scores is associated with a 0.22 (0.44 standard deviation) increase in the perception of school quality (elicited on a Likert scale of from 1= very poor to 5 = very good).[18] This shows that parents are (somewhat) informed at baseline and is consistent with an informational environment that would sustain a separating equilibrium. This also suggests that parental perceptions likely have room for improvement and may potentially reflect other dimensions of quality beyond those captured by test scores.

---

[18] 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 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 if we also include schools where a different set of households provided perceptions across the two time periods or run these regressions at the household X school level rather than aggregating at the school. We prefer aggregation 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.

19

In Column 4, we now test whether providing report cards leads to a stronger relationship between parental perceptions and test scores. We indeed find that in villages that received a report card, the relationship of perceptions with test scores (controlling for baseline perceptions) is stronger, i.e. the coefficient on the interaction term (Report Card* School Score) is 0.114 (Column 4). This is a sizeable effect and represents a substantial increase in the sensitivity of parental perceptions to test scores relative to the control villages.[19]

B. Impact on Market Outcomes

We now examine the impact of report card provision on school fees, test scores, and enrollment at the village/market level. Recall, our conceptual setup suggested a fall in average fees but predictions on school quality required further structure on parental demand and the costs of enhancing quality and is therefore ultimately an empirical question.

I. Fees

Columns 1-3 in Table III show that there were substantial changes in private school fees due to the provision of report cards (recall public schools are essentially free). Panel A presents the specification without any controls, Panel B adds baseline of the dependent variable as a control, and Panel C adds additional

---

[19] Column 4 also highlights limited learning over time in the absence of report cards: Once we account for baseline perception and fees there is no relationship between baseline score and Year 2 perceptions in control villages (i.e. the coefficient on school score is no longer significant). An alternative test of the increased sensitivity between perceptions and test scores would be to see whether there is an increase in the perception-test score *gradient* between years in treatment relative to control villages. Since the perception-test score gradient is the same across treatment and control villages at baseline (regression not shown), this result is already implied in the Column 4 result. This is confirmed if we run the Column 4 regression in a panel format interacting baseline score with treatment and a (post-treatment) time dummy (regression not shown). The Column 4 result also shows that while report card provision increased the sensitivity of parental perceptions to test scores, there is no overall treatment effect, suggesting that parents were not systematically over or underestimating school quality.

village level controls. Our preferred specifications are those in Panel C (the sparser specifications are to demonstrate that the causal effect is not sensitive to the choice of controls). Using our preferred Panel C specification, Column 1 shows that private schools in treatment villages decreased their annualized fees relative to those in the control by an average of Rs.187 in response to the report card intervention, representing 17 percent of their baseline fees.[20] We find a similar effect when we weight by the number of children enrolled in private schools, implying that this effect likely reflects the impact on the fees paid by the average private school going child in the village (Column 2). Finally, in Column 3, we show that the estimated effect is robust to using household reports on school fees, rather than self-reports by the school.[21]

II. Test scores

Columns 4-6 in Table III now examine the impact of report card provision on average test scores in the village. The dependent variable is the average test score in the year following the provision of report cards. We find tests scores improved by 0.11-0.13 standard deviations depending on the specification used.[22] Column 5 shows that these effects persist two years after the provision of the report cards.[23] In Column 6, we replicate the analysis from Column 4, but restrict

---

[20] The fee regressions have slightly fewer observations at the village (104 instead of 112) and school level (274 instead of 303) due to school closures in Year 2 (15 schools), missing data (3 schools), and cleaning major inconsistencies in fee data across grades within years (11 schools).

[21] The dependent variable in Column 3 is the village mean of school fees as reported by households that had a child enrolled in one of the private schools. Since this does not ensure all schools may get a fee report, the number of villages falls. While the magnitude of the fee effect is somewhat smaller than that in Column 2, we cannot reject equality.

[22] Online Appendix Table III shows the results separately for the three different subjects (English, Urdu and Math). We find that report cards increased test scores for the average child in treatment villages by around 0.10 to 0.15 standard deviations, depending on the subject and specification. While point estimates are highest for Math, we cannot reject equality of coefficients across the three subjects, and, therefore, in the paper, we focus on results for the score averaged across the three tested subjects.

[23] Some caution is warranted in interpreting the size of the effect relative to the first year both because the intervention changed (in the second year, we also reported information on score gains)

the village test scores to children tested in both years. The results show that the test score gains were not driven by compositional changes, which is unsurprising given that attrition was low and not differential by baseline test score (Online Appendix I.D). The results are particularly encouraging because the report cards were inexpensive to provide (costing \$1 per child, likely an upper bound for the cost if the program were scaled up) and because the two previous papers on educational report cards in low-income countries showed no positive impact (Banerjee et al., 2010; Mizala & Urquiola, 2013).[24]

III. Enrollment & Switching

Columns 1-4 in Table IV examine whether the report cards led to changes in enrollment and switching at the village level. To the extent that there is a drop in average prices and quality increases, one may expect increased enrollment in treatment villages.

Column 1 looks at overall child enrollment rates in the village. We find that the enrollment rate increases by 3.2 percentage points in treatment villages (around 40 additional children - a 4.5% increase given baseline enrollment rates).[25] These enrollment gains compare quite favorably in terms of cost per additional child enrolled to a range of other interventions such as Conditional (on

---

and because of test score depreciation that may imply the treatment gains are larger than that suggested by the point estimates. Andrabi et al. (2011) show that the coefficient on the lagged test scores for subjects such as Mathematics is less than 0.5 (i.e. gains depreciate over time), suggesting that for the level gains to have remained the same over the two-year period the treatment effect continued to grow between years 1 and 2.

[24] The impact size is also within the range of positive estimates from other educational interventions in low-income countries. The gains are higher than those obtained from reducing class sizes in Kenya and India (Duflo et al., 2012; Banerjee et al., 2007) and are similar to those obtained by providing school grants (Das et al., 2013) or teacher incentives (Muralidharan & Sundararaman 2011; Glewwe et al. 2010).

[25] We note that the Column 1, Panel A, result is close to significance with a p-value of 0.142. Once we control for baseline enrollment rates in Panel B and C, the enrollment result becomes highly significant.

enrollment) Cash Transfers (CCTs) programs that have been used to bolster school participation around the world.[26]

While we find overall enrollment gains, Columns 2 and 3 show that there is little change in the overall switching or dropout rates in treatment villages (i.e. number of children who switch schools/drop out in the village as a fraction of children enrolled at baseline). However, as we will examine later, this lack of an overall impact hides heterogeneous results across schools.

Finally, the lack of evidence of differential switching or dropouts suggests that the test score gains were driven primarily by students who remained in the same school. In Column 4, we restrict the sample to children who were tested in both periods (as in Table III, Column 6) but *also* exclude any children who switched schools. The results confirm that the test score gains for these children remain essentially the same as in Table III (Columns 4 and 6 - show effects of 0.114 and 0.109 respectively, and we now obtain 0.113).[27]

---

[26] Across a number of studies, the gains in enrollment range from 5-10%, but at a cost that can range from $450 per marginal child enrolled in Pakistan (Chaudhury & Parjuli, 2010) to more than $9,000 in Mexico (de Janvry & Sadoulet, 2006). Given the per-child cost of $1, the cost per marginal child enrolled is $20, a number that compares favorably to one of the lowest cost interventions documented thus far—providing information to parents on the returns to schooling (Jensen, 2010).

[27] 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). 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 switchers (R_S^T) see 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%. The decomposition also provides our bounding number. Assuming no treatment effect for non-switchers ($\delta =0$) and no differential switching between treatment and control villages ($R_S^T = R_S^C$), switchers in treatment relative to control villages would need to experience a gain (β) of 0.108/0.048 = 2.25 standard deviations. While one could lower this somewhat by assuming

C. Impact by Provider Quality and Type

We now examine some of the more specific predictions highlighted in the framework in Section II. Since the primary predictions are on school fees, we start with those, but also examine heterogeneous impact on test scores and enrollment.

I. School Fees

With prima facie evidence that schools were likely in a separating equilibrium, we should expect higher price declines among initially high achieving schools - or more specifically - there should be a flattening of the price-quality gradient in treatment villages. Our results in Table V show that this is indeed the case.

Column 1 first regresses (log) fees on test scores before and after the provision of report cards in treatment and control villages. Our interest is in the triple-interaction term, RC*Score*Post. As predicted by the framework, there is a large and significant decline in the price-quality gradient in treatment villages relative to control villages, as a consequence of the report cards.

Columns 2-4 now directly examine how the impact of reports cards on school fees varies by baseline school test scores. Column 2 shows that if a school in a treatment village has a one standard deviation higher test score at baseline, it experiences a Rs 281.6 greater fee drop. Column 3 illustrates the same result using a binary quality measure, which was constructed by dividing schools into high and low-scoring ones (since private schools have generally higher test scores and are fewer in number than public schools, we define a school as high scoring if it is in the top 40th percentile of school test scores in the baseline sample and low scoring otherwise). Online Appendix Table IV shows that our results are similar if

switchers in general also gain, since there are only 0.9% more switchers in treatment villages (and it is not significantly different from control villages), one would still need to assume implausibly large learning gains from switching to generate the observed overall treatment effect.

24

we split at the sample median or use analogous *within* village quality thresholds.[28] Private schools with high baseline test scores show larger price declines (a Rs 294 drop, which is around 25% of their baseline fees) as a result of report card provision, as compared to initially low scoring private schools. Column 4 further validates this by showing the same results hold even when we use fees as reported by households.

Heterogeneity in price declines thus supports the canonical model intuition that the educational markets in this context were inducing separation in price and quality by providing markups to higher score/quality schools and that once information improved through the report cards, this markup fell.

II. Test Scores

While the asymmetric information model provides specific predictions on prices, the predictions on quality depend on the nature of market demand. Moreover, the impact on public school quality is also ambiguous since it depends on the nature of competitive, regulatory and/or social pressures such schools face.

---

[28] Our results are very similar regardless of which binary 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 13%). Our preferred estimate sets the 40th percentile as the threshold and uses the full sample of schools to define the threshold. This leads to a more even division of private schools classified as initially low achieving. Further, because each village has on average seven schools, 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. We note that while the theoretical framework in section II only considers schools in a given village, this does not mean that quality has to be defined *within* village. Specifically, the theory does not require that every village have multiple schools; it only requires that every school can choose its quality level and parents then receive a noisy signal of this quality. The markup required to sustain a high quality school even if it were the only one in a village in an equilibrium with asymmetric information would be similar, although the precise results would depend on assumptions made about the outside option. Moreover, when we present further results on the impact of the report card on test scores by school quality, it is clear that this depends on the cost function schools face in upgrading quality and that is likely to rely on a more absolute notion of quality (rather than just within village quality). In practice though the within sample or within village classifications of quality are not that different given that there is a lot of variation in school quality within villages (and in fact more than the across village variation).

Thus, we empirically test how the report cards affect quality of different types (provider and baseline quality) of schools.

Table VI shows that the quality improvements in private schools observed in the aggregate data were primarily a result of improvements in scores —by 0.31 standard deviations—among the initially low-scoring private schools. Private schools that were high performing in baseline show no improvement (the point estimate is negative). In contrast, Column 1 allows public schools to have differing effects by initial quality but obtains very similar results for both high and low scoring public schools (if anything, point estimates are slightly higher for high scoring public schools but not statistically different from low scoring ones). Column 2 therefore combines both types of public schools and obtains similar results with an overall 0.097 standard deviation increase in public school test scores.

These results suggest that there was excess product differentiation in the initial private market equilibrium. An increase in the precision of quality signals due to report cards allowed for a finer quality gradation such that the low quality schools could now increase their quality. Improvements among public schools also suggest that better information provides non-price incentives for the public sector. The lack of detectable heterogeneous impacts in public schools supports that differential responses among high and low scoring private schools were not mechanically driven by the nature of the production process. Instead, they likely reflect responses to information improvements within the context of a market equilibrium.[29]

---

[29] Online Appendix Table V further shows that heterogeneous responses across schools do not reflect heterogeneous responses across initially low/high scoring *children*. Separating out low/high-scoring children (defined either as above/below the overall sample median score or the child's school's median score) from low/high scoring schools shows that in low-scoring private schools, both low and high-scoring children increased their scores, while in high-scoring schools private schools, neither type of child improved.

III. School Enrollment

Previously, we documented evidence of an aggregate increase in enrollment but little impact on aggregate switching behavior. The latter hides potential heterogeneity, examined further in Table VII. We should caution that these results are (statistically) weaker and therefore presented more as suggestive evidence of movements of children. Columns 1 and 2 consider the impact on total enrollment in the schools (Grades 1 to 5) and for the tested cohort only. The point estimates for both broadly suggest that low-scoring private schools lose children, public schools gain children and high-scoring private schools see little impact. However, results are statistically significant only for total primary enrollment in public schools (Column 1) and for children in the tested cohort for low-scoring private schools (Column 2). With an average baseline enrollment of around 18 children in the tested grade, these are nevertheless reasonably sized effects.

Columns 3 to 4 decompose changes in the tested cohort into children moving into schools (switching in and new children), and those moving out of schools (switching out and dropouts).[30] We see that the loss in net enrollment in low-scoring private schools is primarily driven by children switching or dropping out (in regressions not shown, separating between the two shows equal sized effects). Notably, while the net gain in high scoring private schools was minimal, this masks churning within these schools with children both switching in and newly enrolling (one additional child) countered by an increase (of half a child) in switching or dropping out. This churning likely reflects both heterogeneous responses across parents and within these schools.

---

[30] We can only do so for children in the tested cohort by using a child-tracking exercise that followed every child (in the tested grade) enrolled in year 1 through the subsequent years. Of the 13,735 children in the roster in year 1, we were able to definitively track 12,990 children (95 percent). Out of the 12,110 *tested* children in year 1, we definitively tracked 11,580 children (96 percent).

Column 5 shows that, consistent with some of the enrollment changes, the treatment also increased the incidence of closure among schools, with low-scoring private schools 12.5 percentage points more likely to close in treatment villages. Given the smaller number of low-scoring private schools, this increased rate of closure reflects an additional six such schools closing in treatment villages. Was the decline in enrollment in low-scoring private schools a result of these closures? If we re-estimate Columns 1 and 2 and exclude any schools that closed, we confirm that this is indeed the case (regressions not shown). However, since we don't know the timing of child switching and school closure, it is hard to tell whether a large number of children switching out lead to some schools closing or vice versa.[31]

IV. Discussion and Conclusion

While the literature has shown that providing information can change individual behavior, we have limited rigorous empirical evidence on how markets, especially for services, adjust when the informational environment improves, particularly in contexts where there are few administrative constraints or requirements on quality and pricing. This paper informs that question, showing that providing education report cards to parents changes the price-quality menu offered in the market. Specifically, the informational improvement results in a

---

[31] Closures could affect our interpretation of the quality increase in low-scoring private schools if the schools with the lowest expected gains shut down. This is less of a concern because even if a school closed, we were able to track the child when they re-enrolled in another school. Since we assign children to their initial school (the timing of report card provision makes it likely they spent more time in their initial school), we can still (partially) consider gains in closed schools. In addition, for children in closed schools that we are unable to retest, we can conduct a bounding exercise. This bounding exercise shows that for the observed gains in low-scoring private schools to be driven entirely by selective school closure one would need to have the schools that shut down experience an extremely large test score decline of more than three standard deviations. More plausibly, assigning children in such schools to either the worst score gain of any private school in our sample or the gain of a school closest to them in baseline test scores does not alter the point estimate.

more favorable price-quality tradeoff offered to parents - with initially high quality schools dropping fees and low-quality ones improving quality. This is accompanied by quality improvements in the public sector and overall enrollment gains. Information on test scores thus seems to improve efficiency and equity simultaneously.

Our paper also underscores how the impact of information provision depends critically on the initial informational environment and equilibrium. A commonly held, though admittedly casual, view is that providing information should allow the (initially) better quality providers to benefit more by increasing prices and/or seeing higher revenues. Yet standard models of asymmetric information can readily obtain the opposite results - that better quality providers can lose (informational) rents when the information environment improves. Our paper demonstrates this empirically. In fact, this may be the more realistic scenario to the extent that the initial informational environment is likely to neither be so poor so as to be uninformative (i.e. only pooling is possible) nor so rich so as to be fully revealing. Our paper thus provides a more informed and realistic perspective on what we could expect as market participants receive better information.

The magnitudes of the impacts we find are also quite large. The gains in learning alone compare favorably to other interventions, both in absolute terms and relative to normal yearly gains (the treatment effect is 42 percent of the average yearly gain experienced by children in our sample).[32] Similarly, the gain in enrollment represents a cost per marginal child enrolled of $20, which is significantly lower than several programs that are currently regarded as quite successful in low-income countries (Akresh et al., 2013). In terms of overall

---

[32] A recent meta-study (McEwan, 2013) of over 70 educational intervention studies from developing countries finds that the largest mean effects were around 0.15 standard deviations (for interventions with computers or instructional technology).

welfare implications, it is noteworthy that the cost of providing information was similar to the decline in school fees. Specifically, the cost of the report card exercise was $1 per child, and the fee savings were approximately $3 per child in private schools. With nearly one-third of all children enrolled in private schools in these villages, the total cost of providing information for all children is comparable to the decline in fees. This partial analysis would suggest that the entire improvement in test scores is free of cost if only the welfare of households is considered.[33]

Finally, our results also help inform the ongoing debate on public versus private education in low-income countries where public sectors failures are common. Much of the debate about how to improve learning focuses on public schools. In fact, it is especially telling that many interventions and studies focus on the public sector. Yet, increasingly, parents face choice over public versus private, as well as choice over the type of private school. In this context, market-level interventions that can improve the performance of the schooling sector as a whole can yield rich dividends. What we have been able to show here is that the dissemination of credible and comparable information on learning quality is an intervention that can improve performance in the private sector and simultaneously strengthen the public sector. Government policies that do not account for such multiplicity of providers may miss key opportunities for reform. Fixing market failures in the private sector should remain a priority—and in doing so, can yield broad improvements across the public and private sectors, both on efficiency and equity.

---

[33] A comparison with other cost-benefit calculations in the low-income country educational literature shows the usual practice is to focus only on the welfare of households and their children. For instance, in cases where improvements have come through greater effort by teachers, their welfare cost is not accounted for in the cost-benefit analysis. However, a complete welfare analysis would exclude the decline in fees, since it is a direct transfer from the schools/teachers to parents and focus on the enrollment and test score gains alone. Even by that criteria though, the returns to this intervention are significant.

References

Akresh, R., De Walque, D., & Kazianga, H. (2013). Cash transfers and child schooling: evidence from a randomized evaluation of the role of conditionality. Working Paper, World Bank: Washington, DC.

Allcott, H., & Mullainathan, S. (2012). External validity and partner selection bias (No. w18373). National Bureau of Economic Research.

Alderman, H., Orazem, P. F., & Paterno, E. M. (2001). School quality, school cost, and the public/private school choices of low-income households in Pakistan. *Journal of Human Resources* 36 (2): 304-326.

Andrabi, T., Das, J., Khwaja, A., Vishwanath, T., & Zajonc, T. (2002). Test Feasibility Survey, PAKISTAN: Education Sector. Harvard Kennedy School Working Paper.

Andrabi, T., Das, J., Khwaja, A., Vishwanath, T., & Zajonc, T. (2007). The Learning and Educational Achievement in Punjab Schools (LEAPS) Report. *World Bank, Washington, DC*.

Andrabi, T., Bau, N., Das, J., & Khwaja, A. I. (2010). Are Bad Public Schools Public "Bads"? Test Scores and Civic Values in Public and Private Schools. Working Paper.

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.

Andrabi, T., Das, J., & Khwaja, A. I. (2013). Students Today, Teachers Tomorrow: Identifying constraints on the provision of education. *Journal of Public Economics*.

ASER India. Annual Status of Education Report India (2012).

ASER Pakistan. Annual Status of Education Report Pakistan (2012).

Bagwell, K., & Riordan, M. H. (1991). High and declining prices signal product quality. *The American Economic Review*, 224-239.

Banerjee, A.V., S. Cole, E. Duflo & L. Linden (2007). "Remedying Education: Evidence from Two Randomized Experiments in India". *The Quarterly Journal of Economics* 122 (3): 1235-1264.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of Participatory Programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, *2*(1), 1-30.

Björkman, M., & Svensson, J. (2009). Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, *124*(2), 735-769.

Burde, D., & Linden, L. L. (2012). The effect of village-based schools: Evidence from a randomized controlled trial in Afghanistan (No. w18039).

National Bureau of Economic Research.

Carneiro, P., Das, J., & Reis, H. (2013). Parental valuation of school attributes in developing countries: Evidence from Pakistan. Working Paper. Retrieved from: http://ipl.econ.duke.edu/bread/papers/1013conf/carneiro.pdf

Chan, Y. S., & Leland, H. (1982). Prices and qualities in markets with costly information. *The Review of Economic Studies*, 49(4), 499-516.

Chaudhury, N. and D. Parjuli (2010). "Conditional cash transfers and female schooling: the impact of the female school stipend programme on public school enrolments in Punjab, Pakistan". *Applied Economics*, Volume 42, Issue 28

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9), 1045-1057.

Das, J., & Zajonc, T. (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, *92*(2), 175-187.

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

De Janvry, A. and B. Sadoulet (2006). "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality". *World Bank Economic Review*, 20 (1): 1-29.

Dixon, P. (2012). Why the denial? Low-cost private schools in developing countries and their contributions to education. *Econ Journal Watch*, *9*(3), 186-209.

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is More Information Better? The Effects of Health Care Quality Report Cards. *Journal of Political Economy*, *111*, 555-88.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, 48(4): 935-63.

Duflo, E., Dupas, P., & Kremer, M. (2012). "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools" (No. w17939). National Bureau of Economic Research.

Figlio, D. N., & Lucas, M. E. (2004). Whats in a Grade? School Report Cards and the Housing Market. *The American Economic Review*, *94*(3), 591-604.

Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system?. *Advances in Applied Microeconomics*, *14*, 35-49.

Glewwe, P., N. Ilias, and M. Kremer (2010). "Teacher Incentives" *American Economic Journal: Applied Economics*, 2(3): 205-27.

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

Hoxby, C. M. (2002). The cost of accountability (No. w8855). National Bureau of Economic Research.

Jacob, B. A., & Levitt, S. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating.". *Quarterly Journal of Economics*. 118(3): 843-877.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quarterly journal of Economics*, *122*(3), 879-924.

Jensen, R. (2010). "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 125(2), p. 515-548

Jin, G. Z., & Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, *118*(2), 409-451.

Judd, K. L., & Riordan, M. H. (1994). Price and quality in a new product monopoly. *The Review of Economic Studies*, *61*(4), 773-789.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91-114.

McEwan, P.J. (2013). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. Retrieved from: http://academics.wellesley.edu/Economics/mcewan/PDF/meta.pdf

Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *The Journal of Political Economy*, 796-821.

Mizala, A., & Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103, 313-335.

Muralidharan, K., & Sundararaman, V. (2011).Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39-77.

PEIP (Private Educational Instituions in Pakistan) (2000). Pakistan Bureau of Statistics, Government of Pakistan.

Rockoff, J. E., & Turner, L. J. (2008). Short run impacts of accountability on school quality (No. w14564). National Bureau of Economic Research.

Rogosa, D. (2002). Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores". *Retrieved from:* http://www-stat.stanford.edu/~rag/api/kscresst.pdf

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics*, *98*(4), 659-679.

Wolinsky, A. (1983). Prices as signals of product quality. *The Review of Economic Studies*, *50*(4), 647-658.

World Bank (2004). World Development Report: Making Services Work for Poor People.

World Bank (2014). World Bank Data Platform. Retrieved on: April 17, 2014.

Table I: Summary Statistics

	Mean	Standard Deviation	N
Panel A: Village Level			
Village Wealth (Median Monthly Expenditure - Rupees)	4,641.5	1,575.2	112
Number of Households in Village	631.3	383.9	112
Percent of Adults (>24) Literate in Village	37.3	11.9	112
Village enrollment % (All)	70.8	16.9	112
Village enrollment % (Boys)	76.2	15.6	112
Village enrollment % (Girls)	64.8	19.7	112
Herfindahl Index of Schools in Village	0.194	0.076	112
Panel B: School Level			
School Average Test Score	0.029	0.727	804
School Fees (private schools only - Rupees)	1,184.4	811.5	289
Number of Students Enrolled at School (All grades)	166.2	150.7	804
Number of Students Enrolled at School (Grades 1-5)	88.8	70.6	802
Grade 3 Enrollment at Baseline (Number)	17.6	15.0	802
Percent of Teachers with More than a Matriculate Education	0.561	0.302	784
Panel C: Child Level			
Child Average Test Score	-0.018	0.913	12,110
Female Child	0.443	0.497	13,735
Child Age	9.7	1.5	13,733
Child Time in School or School Prep (minutes per day)	420.8	65.3	983
Child's Time Spent on School Work, not in school (minutes per day)	96.7	61.5	982
Perception of School Quality (Likert scale: 1 to 5)	3.3	0.5	619
Parents' Spending on School Fees (Rupees)	302.5	531.3	954
Parents' Education Spending, other than School Fees (Rupees)	969.1	822.2	988
Parents' Time Spent Teaching Child (hours per week)	3.4	5.2	964

Notes:

This table presents baseline summary statistics for outcome and control variables in the main regression tables and the online appendix tables, as well as other background variables mentioned in the text of the paper. Panel A displays variables at the village level. All variables have 112 observations, which is the number of villages in our sample. Panel B displays variables at the school level. Most variables consider those 804 schools which had non-zero Grade 3 enrollment; missing data reduces the number of observations in some cases. Some variables consider data only from private schools; there are 303 private schools in the sample, but missing/inconsistent data further reduce the number of observations. Panel C displays variables at the child level. These variables derive from three different sources: (i) child roster data from testing at the school (variables with greater than 12000 observations), (ii) child and parental data from household survey for all children in the household data that were matched to the school testing roster (variables with observations in the 900s), and (iii) household data on perceptions averaged at the school-level (variable with 619 observations - we have fewer than 800 observations, the number of schools in the sample, because parents were not asked to provide perceptions for schools they did not know about and could respond with "don't know").

Table II: Fee-School Quality Relationship and Impact on Perceptions

| | Fees (Year 1) | | Perception | |
| | | | Year 1 | Year 2 |
	(1)	(2)	(3)	(4)
School Score	369.2***	316.3***	0.216***	-0.0279
	(95.07)	(107.2)	(0.0239)	(0.0347)
School Fee				0.000129***
				(2.26e-05)
Baseline Perception				0.228***
				(0.0365)
Report Card				0.00798
				(0.0364)
Report Card * School Score				0.114**
				(0.0438)
Basic Infrastructure Index		54.93*		
		(33.08)		
Extra Infrastructure Index		141.1*		
		(79.67)		
Controls	Village	Village Fixed Effects	Village	Village
Observations	289	289	610	588
R-squared	0.337	0.137	0.116	0.315
Baseline Depvar (mean)	1184.360	1184.360	3.288	3.275

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

This table presents results on the association between school fees and quality, and some findings on perception of school quality. Columns 1 and 2 show the relationship between school characteristics and school fees for private schools; there are 303 private schools in our sample, but we have fewer observations due to missing data. The dependent variabes in Columns 3 and 4 are constructed by taking the average of all parental perceptions, ranked of a five point scale, for a given school. This ensures schools are equally represented (one observation per school). Column 3 shows the correlation between school test score and parental perception in year 1. Column 4 considers perception in year 2 (open schools only). We have fewer than 800 schools in Columns 3 and 4 because we restrict to only those household-school combinations where we have perceptions data for both rounds. All regressions have standard errors clustered at the village level, and include district fixed effects. All regressions include baseline of outcome variable as a control as well as, where appropriate, additional village controls (village wealth [median monthly expenditure], number of households in village, Herfindahl index of schools in village, and percent of adults [>24] literate in village). Baseline Depvar (Mean) displays the baseline mean for the sample for all outcome variables.

	Village Average Fees (Year 2)			Village Average Test Scores		
	School Report		Household Report			
		Weighted by				Year 2
	Basic	Children	Basic	Year 2	Year 3	(Same Kids)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: No Controls						
Report Card	-288.4***	-334.1***	-193.9*	0.128**	0.140**	0.129**
	(92.58)	(107.9)	(99.97)	(0.0624)	(0.0584)	(0.0599)
Observations	104	104	83	112	112	112
R-Squared	0.336	0.473	0.259	0.328	0.292	0.399
Panel B: Baseline Control Only						
Report Card	-191.8***	-194.9***	-128.2*	0.107**	0.122***	0.103**
	(65.18)	(55.92)	(73.46)	(0.0448)	(0.0428)	(0.0395)
Baseline	0.750***	0.799***	0.780***	0.710***	0.648***	0.719***
	(0.104)	(0.0865)	(0.0859)	(0.0628)	(0.0742)	(0.0603)
Observations	104	104	83	112	112	112
R-Squared	0.719	0.808	0.644	0.687	0.625	0.746
Panel C: Baseline and Village Controls						
Report Card	-187.0***	-175.2***	-141.7*	0.114**	0.123***	0.109***
	(65.91)	(62.12)	(74.35)	(0.0455)	(0.0435)	(0.0401)
Baseline	0.764***	0.842***	0.742***	0.706***	0.644***	0.718***
	(0.104)	(0.102)	(0.0831)	(0.0624)	(0.0754)	(0.0596)
Observations	104	104	83	112	112	112
R-Squared	0.726	0.816	0.665	0.692	0.631	0.749
Baseline Depvar (Mean)	1080.699	1234.479	998.964	-0.032	-0.032	-0.008

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

The outcome variables are: Year 2 village average private school fees from school survey data - in levels (column 1), in levels and weighted by children in school (column 2); Year 2 village average private school fees, in levels, from household survey data (column 3); Year 2 village average (across all three subjects-Math, English, Urdu) test scores (column 4); Year 3 village average test scores (column 5); Year 2 village level average test score using only those kids tested in years 1 and 2 (column 6). All regressions include district-fixed effects and robust standard errors. Panel A considers no additional controls; Panel B includes a baseline control of the outcome variable; and Panel C includes baseline of the outcome variable and additional village controls, which are the same as in Table II. Columns 1-3 have fewer than 112 observations due to private school closure in Year 2 and missing/inconsistent fee data in some villages. Column 3 has 83 villages because we only consider those villages where we can match children who attend private school from the household survey to the testing roster. Columns 3-6 are run on all 112 sample villages. Baseline Depvar (Mean) displays the baseline mean for the sample for all outcome variables.

Table IV: Enrollment & Switching - Impact on Market Outcomes

	Village Enrollment (Year 2)			Village Average Test Scores - Same Kids, No Switchers (Year 2)
	Primary Enrollment Rate	Switching Rate	Dropout Rate	
	(1)	(2)	(3)	(4)
Panel A: No Controls				
Report Card	0.0390	0.009	0.009	0.129**
	(0.0263)	(0.007)	(0.006)	(0.0608)
Observations	112	112	112	112
R-Squared	0.473	0.0561	0.377	0.397
Panel B: Baseline Control Only				
Report Card	0.0351**			0.107***
	(0.0140)			(0.0402)
Baseline	0.973***			0.711***
	(0.0470)			(0.0595)
Observations	112			112
R-Squared	0.851			0.742
Panel C: Baseline and Village Controls				
Report Card	0.0324**	0.009	0.007	0.113***
	(0.0137)	(0.0074)	(0.0056)	(0.0408)
Baseline	1.037***			0.711***
	(0.0690)			(0.0587)
Observations	112	112	112	112
R-Squared	0.853	0.083	0.429	0.745
Baseline Depvar (mean)	0.71	-	-	-0.012

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

The outcome variables are: Year 2 village primary enrollment rate (column 1); switching rate and drop out rate at the village level (columns 2 and 3); and Year 2 village average test score for those kids who did not switch schools between years 1 and 2 (column 4). All regressions include district fixed effects and display robust standard errors in parantheses. Panel A considers no additional controls; Panel B includes a baseline control of the outcome variable; and Panel C includes baseline of the outcome variable and additional village controls, which are the same as in Table II. Baseline Depvar (mean) displays the baseline mean for the sample for all outcome variables. Note that we do not observe baseline rates for switching and dropout. Columns 1-4 are run on all 112 sample villages.

Table V: School Fees - Impact by Provider Quality and Type

	Fees (Year 2)		Fees (Year 2)		
			Continuous Quality	Binary Quality	Binary Quality (Household Report)
	(1)		(2)	(3)	(4)
Report Card (RC)	-0.139	Report Card (RC)	-111.6	-42.70	78.58
	(0.0916)		(76.40)	(88.65)	(145.2)
School Score (Score)	0.244**	School Score (Score)	195.9		
	(0.114)		(162.9)		
RC * Score	0.0389	RC * Score	-281.6*		
	(0.150)		(163.0)		
Score * Post	0.0544	High * Private		232.2*	530.2***
	(0.129)			(121.3)	(189.0)
RC * Score * Post	-0.368**	RC*High*Private		-293.8**	-511.4**
	(0.179)			(129.0)	(207.1)
Post	-0.177	Baseline	0.683***	0.681***	0.488***
	(0.323)		(0.122)	(0.117)	(0.125)
RC * Post	0.121				
	(0.109)				
Controls	Village		Village	Village	Village
Observations	555		274	274	238
R-Squared	0.311		0.584	0.585	0.402
SUBGROUP POINT ESTIMATE, F-TEST p-VALUES IN BRACKETS					
		Low private school		-42.70	78.58
				[0.631]	[0.590]
		High private school		-336.5	-432.9
				[0.000]	[0.000]
Baseline Fee (mean)	6.911		1188.482	1188.482	1047.882

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

The outcome variables are: Year 2 private school fees - in logs (column 1); Year 2 private school fees from school survey data - in levels (column 2 and 3); and Year 2 private school fees from household survey data (column 4). Column 1 data is from the school survey and is constructed in a panel format to test whether the price-quality gradient falls as a result of the intervention; we thus see roughly double the number of observations in Column 1 compared to Columns 2 and 3. The coefficient of interest in the triple interaction terms (RC*Score*Post). Column 2 considers continuous quality whereas Columns 3 and 4 consider binary quality with schools defined as high quality if they are in the top 40th percentile of the test score distribution. The number of observations is less than 303 private schools due to missing/inconsistent fee data and private school closure in Round 2. Column 4 has even fewer observations because we only use data from those households with children in private schools who we tested and were able to match in our testing roster. All regressions include district-fixed effects and cluster standard errors at the village level. Additional village level controls, the same ones listed in Table II, are used in all regressions. 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 fee mean for the sample across all regressions.

Table VI: Test Scores - Impact by Provider Quality and Type

	Test Scores (Year 2)	
	By School Type and Quality	Government Schools Combined
	(1)	(2)
Report Card (RC)	0.315**	0.313**
	(0.122)	(0.122)
RC*Government (Gov)	-0.241**	-0.217*
	(0.120)	(0.123)
RC* High	-0.389***	
	(0.133)	
RC*Gov*High	0.505**	
	(0.218)	
RC * High * Private		-0.389***
		(0.133)
High	0.211***	
	(0.0736)	
Government	-0.208***	-0.225***
	(0.0516)	(0.0593)
High * Private		0.218***
		(0.0757)
Baseline	0.542***	0.534***
	(0.0608)	(0.0630)
Controls	Village	Village
Observations	780	780
R-Squared	0.65	0.648

SUBGROUP POINT ESTIMATE, F-TEST p-VALUES IN BRACKETS

	Low Private School	0.315	Low Private School	0.313
		[0.011]		[0.011]
	High Private School	-0.0738	High Private School	-0.0755
		[0.194]		[0.181]
	High Gov School	0.190	Gov School	0.0965
		[0.262]		[0.044]
	Low Gov School	0.0742		
		[0.111]		

Baseline Test Score (mean)	-0.021	-0.021

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

The outcome variable is Year 2 school average (across all three subjects) test score. Column 1 separates the effect by school type and by school performance. Column 2 combines government school and focuses on private school type. Regressions are weighted by number of children tested in the school, and include district fixed effects and cluster standard errors at the village level. The same village controls as in Table II are included in both regressions. Regressions 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 school test score mean for the sample, weighted by number of children tested.

Table VII: School Enrollment - Impact by Provider Quality and Type

	Primary Enrollment (Year 2)	Tested Cohort Enrollment (Year 2)	Tested cohort children going into schools (Year 2)	Tested cohort children going out of schools (Year 2)	Private School Closure (Year 2)
	(1)	(2)	(3)	(4)	(5)
Report Card (RC)	-4.472	-1.474*	-0.410	1.296**	0.125**
	(3.815)	(0.846)	(0.483)	(0.537)	(0.0486)
Government	7.315***	1.628*	0.155	-0.698**	
	(2.655)	(0.838)	(0.666)	(0.305)	
High* Private	3.216	-0.792	-1.293**	-0.192	0.0336
	(3.241)	(0.801)	(0.570)	(0.303)	(0.0237)
RC*Gov	9.424*	2.180**	0.989	-1.413**	
	(4.769)	(1.063)	(0.752)	(0.580)	
RC*High*Priv	3.906	1.706	1.428**	-0.794	-0.111*
	(4.853)	(1.072)	(0.665)	(0.604)	(0.0599)
Baseline Enrollment	0.961***	1.065***	0.169***	0.0485***	
	(0.0254)	(0.0491)	(0.0407)	(0.0109)	
Controls	Village	Village	Village	Village	Village
Observations	801	802	798	802	303
R-Squared	0.904	0.863	0.203	0.151	0.0378
SUBGROUP POINT ESTIMATE, F-TEST p-VALUES IN BRACKETS					
Low Private School	-4.472	-1.474	-0.410	1.296	0.125
	[0.244]	[0.084]	[0.397]	[0.018]	[0.011]
High Private School	-0.567	0.232	1.017	0.502	0.0141
	[0.836]	[0.714]	[0.043]	[0.073]	[0.633]
Government School	4.952	0.706	0.578	-0.117	
	[0.013]	[0.273]	[0.335]	[0.557]	
Baseline Depvar (mean)	88.774	17.562	-	-	-

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

The outcome variables are: Total primary, Grades 1-5, enrollment in Year 2 (column 1); Tested cohort enrollment, Grade 4, in Year 2 (column 2), these are children now in Grade 4 who were originally tested in Grade 3; Number of children in the tested cohort going into schools (column 3) includes children who were confirmed to have switched into or newly enrolled in a school; Number of children in the tested cohort going out of schools (column 4) includes children who were confirmed to have switched out or dropped out of school, and untracked children from closed schools; and school closure by private school type (column 5). For columns 1, 2 and 5, we use data from school surveys. For colums 3 and 4, we use child tracking data for the tested cohort. Columns 1-4 are run on all 804 schools in 112 villages; some missing values reduce the number of observations. Column 5 is run on all 303 private schools in the sample. All regressions include district-fixed effects and are clustered at the village level. The same village controls as in Table II are included in all regressions. 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 sample for all outcome variables.

ONLINE APPENDIX

<u>Appendix I: Experimental Protocol</u>

A.   Sampling Design

Our sample includes 112 villages across three districts in Punjab, Pakistan, where the three districts were chosen to represent the different socioeconomic zones within the province (one each from the North, Center and South of the province). Within these districts, 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. The random selection of villages into the experimental frame helps alleviate "partner and location selection" concerns of the sort raised by Allcott and Mullainathan (2012). 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 typical village in our sample. The school type icons are labeled (government/public or private) and the shopping cart icon designates retail stores.

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.

Figure 2: School Sample Selection
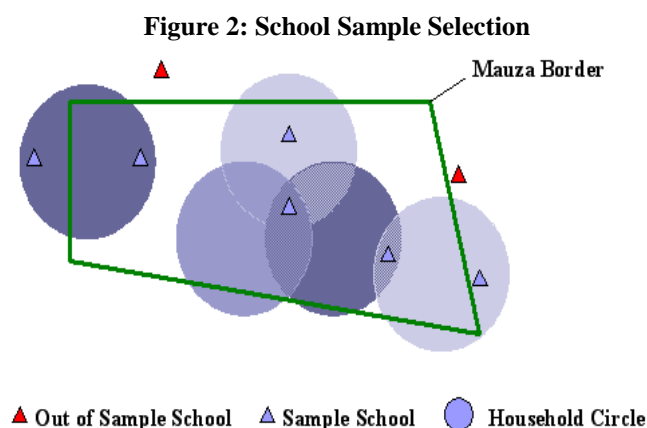
B.   Survey Instruments and Timeline

We use data from a range of surveys over a 2-year period of the LEAPS project. In the first year (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 again roughly 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 Listing	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.   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.

ii. 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.

iii. Child Tests

We administered tests of English, Mathematics and Urdu (the vernacular) 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 non-absent children in Grade 3 (12,110 children) in the sample schools 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.D. 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.

iv.   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.

Household-based Surveys

This survey is administered to over 1,800 households. These households are selected based on a census conducted in the sample villages in 2003, which listed over 80,000 households. We stratify by enrollment status and over-sample households with children who might be eligible by age for (tested) Grade 3.

---

[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.

The household survey 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.   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, Pandey & Zajonc, 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
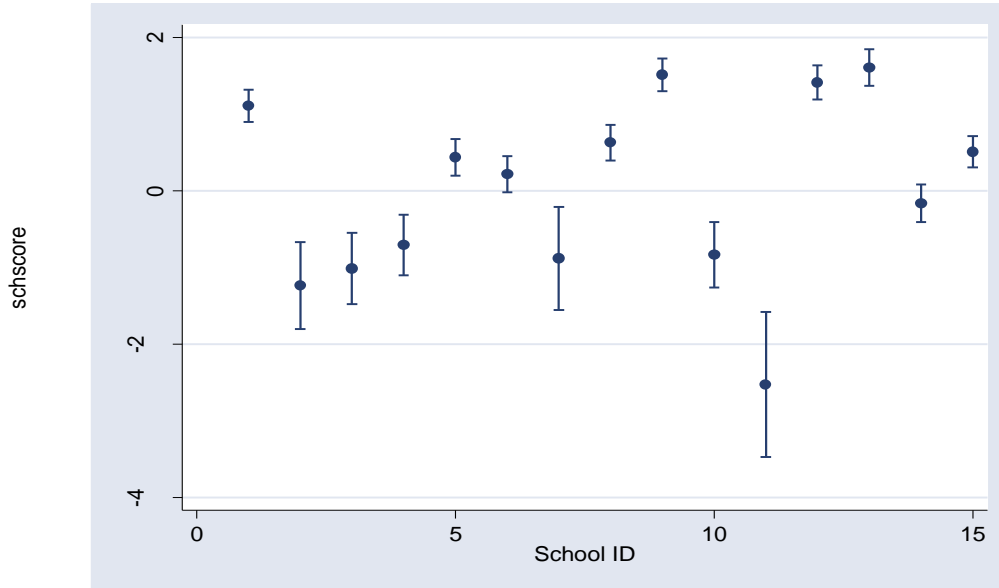
A template of the report card is shown in Figure 4. Card 1 (the top image) reports the score of the child in English, Mathematics, and Urdu, as well as her quintile rank across all tested children. Quintile rank is described as: 1 – "Very Good"; 2 – "Good", 3- "Satisfactory", 4 – "Needs Improvement" 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 (2002).

Figure 4: Report Cards

Card 1

Learning and Educational Achievement in Punjab Schools

رپورٹ کارڈ برائے تعلیمی کارکردگی

| School Name | | | Child Name |
| Grade | | | Father Name |

Math		English		Urdu		Ranking
Rank	Obtained Marks (Total Marks 100)	Rank	Obtained Marks (Total Marks 100)	Rank	Obtained Marks (Total Marks 100)	1st: Very Good 2nd: Good 3rd: Satisfactory 4th: Needs Improvement 5th: Needs Significant Improvement
						Child Performance
						Average School Score
						Average Village Score of all schools

Mauza
Exam Date

Card 2

Learning and Educational Achievement in Punjab Schools

کے تمام سکولوں کے بچوں کی اوسط کارکردگی

Math		English		Urdu		Number of Tested Students	School Name
Rank	Obtained Marks (Total Marks 100)	Rank	Obtained Marks (Total Marks 100)	Rank	Obtained Marks (Total Marks 100)		

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).

D.   Balance and Attrition

I.   Baseline Balance

We confirm that baseline values of outcomes and control variables are balanced between treatment and control villages (Online Appendix 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 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; the other children were absent the day of the test. 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 high student absence on a day-to-day basis.[4] 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 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 (Online Appendix Table IIA), and unable to detect any differences in child or parental characteristics and child test scores for attriters in the treatment and control groups (Online Appendix Table IIB, Columns 1-2). Separating attriters into those who were absent on the day of the test, those who definitely dropped out, and those who left the village or whose status we were unable to confirm 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.

---

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

Appendix II: Theory

In order to interpret and expand on the results, we refer to canonical models that examine the link between information provision and market equilibria (prices and quality) under asymmetric information. In such models, even in the absence of third-party information, consumers receive signals of quality and schools optimally account for the information environment in deciding how much to invest in quality.[5] The central intuition that follows is that prices can substitute for the lack of information in determining the market equilibrium.

Here, we sketch this basic insight to understand how the report card intervention likely impacted the market for education. This exercise suggests dimensions of school type and quality along which we expect impact heterogeneity.

Stylized Theoretical Setup

Our stylized framework of market equilibrium under asymmetric information draws entirely on Wolinksy (1983), but the market structure when prices can signal quality has been explored in several contributions, notably Chan and Leland (1982), Milgrom and Roberts (1986), Shapiro (1983), Bagwell and Riordan (1991) and Judd and Riordan (1994).[6] The framework provides a general template of what to expect in situations where consumers receive a noisy signal of quality, and firms can set quality to respond to the overall informational environment. We examine market equilibrium in the *absence* of third-party information and plausible impacts on the market when the precision of the information signal increases. Within this stylized framework, we discuss two predictions relating the provision of information to the price structure and distribution of school qualities in the market, which are feasible to examine directly in our data.

Analogous to Wolinsky's (1983) setup, let school s derives profits which depend on the cost of producing quality, $c(q_s)$, the price (p_s), the expected sales volume, x_s and a fixed cost of entry, z.

$$\pi_s = (p_s - c(q_s))x_s - z$$

---

[5] Informal mechanisms, learning by parents, and tests that schools themselves conduct are all potentially informative, albeit noisy, signals of school quality. Schools can affect their quality by changing their investments in teacher quality and spending on infrastructure and educational materials.

[6] The models differ in the specific handling of information (whether consumers receive a signal of quality on product inspection, as in Wolinsky [1983] or can purchase information at different costs [Chan and Leland, 1982]); the market structure (competitive as in Wolinsky [1983] or a monopolist as in Judd and Riordan [1994]); and the extent to which dynamic aspects are introduced. Of particular interest is Shapiro's (1983) model where sellers invest in reputation and higher markups for high quality firms act both to prevent cheating as in Wolinsky (1983) but also as an incentive to invest in reputation building in early periods. The central result that markups can substitute for information remains an underlying theme across all these models.

There are m consumer types (parents) who each consume one unit of the good with a value function that depends on the good's price (p), quality (q) and their valuation of quality, θ.[7] Consumer i's preferences are given by $U = u_i(q,\theta) - p$. This is the basic formulation in many models of market equilibrium when firms vary by quality. For instance, one parametric form of the utility function used in several models is $U = \theta q - p$ with θ distributed as $F(\theta)$. This model under perfect information will lead to appropriate cutoffs of θ, such that those with higher valuations (higher θ) will choose higher q and will pay a higher p. Thus, without any information asymmetry the model leads to a hedonic equilibrium—parents will choose the quality of the school they wish to send their children to depending on the price and their valuation of quality (Rosen, 1974).

To incorporate information into this model, Wolinksy (1983) presents an intuitive approach that allows information to be treated in an analytically convenient manner. Specifically, the information structure is such that for any quality level q, there is always a lower bound on the quality signal that the parent can receive; any signal below this lower bound fully reveals that the school cannot have produced at quality q. Formally, parent i receives a signal of quality for school s prior to choosing a school where the cumulative distribution of the signal is given by:
$$D(t,q) = prob\ (d^i_s \leq t \mid q_s = q)$$

Assume that for every q, there is a at least one t such that $D(t,q) = 0$. Define t_q^* as the maximum t such that $D(t,q)=0$. That is, for every school producing at a particular quality level, there is a single scalar t_q^*, such that *no parent* can ever receive a signal lower than t_q^* if the school produces at $q_s = q$. If any parent does receive such a signal, it reveals with probability 1 that the school cannot have produced q. This setup is illustrated in Figure 5. We plot hypothetical density functions of the signal that parents receive when quality is L(ow) or H(igh) respectively. The key assumption is that there is a t_H^*, denoted by the dashed line, such that $D(t,H)=0$ for all signals below t_H^*. For a school that produces H, parents can never receive a quality signal to the left of the dashed line; if they do, it will reveal to them with certainty that the school cannot have produced H.

---

[7] To avoid technical issues arising from the type of market collapse in Diamond (1971) and issues of costly sampling (Grossman & Stiglitz, 1980), the model requires (i) that consumers pay a small fixed cost, k, for each firm they visit and that information arrives the moment a firm is visited and (ii) that quality is produced in discrete units.

Figure 5: Density Functions of Signal with Low (L) or High (H) Quality

d(t, L)

d(t, H)

t^*_H: D(t, H) = 0

d_s^i

Increase in t* implies a better information
environment
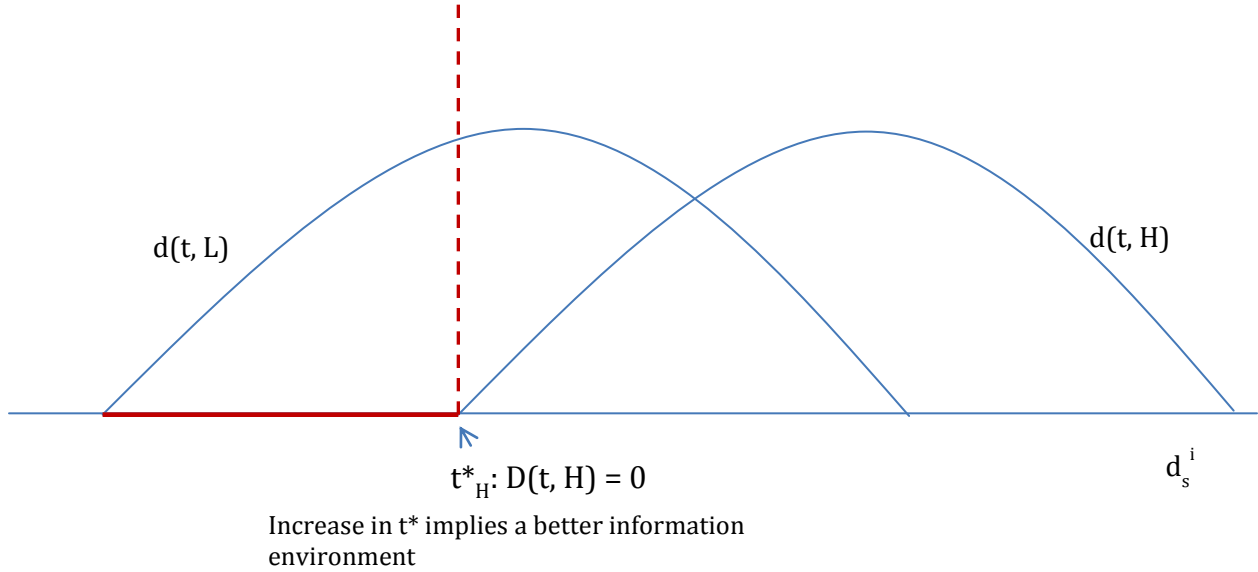
This figure plots hypothetical density functions of the signal that parents receive when school quality is
L(ow) or H(igh) respectively. The key assumption is that there is a tH*, denoted by the dashed line, such
that D(t,H)=0 for all signals below tH*. For a school that produces H, parents can never receive a quality
signal to the left of the dashed line; if they do, it will reveal to them with certainty that the school cannot
have produced H.

Initial Equilibrium

Of interest for our empirical exercise is how the improvements in the signal can alter equilibrium
in this model. Specifically, we are interested in the conditions under which schools will *separate*
so that prices reveal quality. We present a simplified decision process to highlight the main
intuition that abstracts from consumer choice across schools and focuses purely on the
information problem. For a full treatment, we refer the reader to Wolinsky (1983). Consider the
decision process for a single school, deciding whether to produce H or L quality, faced with a set
of x parents who would choose the school for sure if they knew its quality were H. In a
separating equilibrium, every quality is associated with a different price and the choice of p
completely reveals the choice of q. For this separation to hold, it must be the case that the
choices of p and q are incentive compatible. Suppose that a school tries to deviate by charging p_H
but producing q_L. In this case, relative to producing q_H,

$$Gains = (c_H - c_L)[1 - D(t_H^*, L)]x$$
$$Loss = (p_H - c_H)[D(t_H^*, L)]x$$

The gains represent first the fact that by producing quality L, for every unit produced the school saves c_H-c_L. At this new quality level, the fraction of parents who will receive a signal consistent with H are those whose signal is greater than t_H^*, that is *[1-D(t_H^*,L)]*. These parents will enroll their children in the school. The loss from charging p_H but producing L are the markups from every parent when producing H, *(p_H-c_H)* multiplied by the fraction of parents who receive a signal inconsistent with the school producing H, i.e. *[D(t_H^*,L)]*. For the separating equilibrium to hold, it must be that the gains are no greater than the loss, so that

$$p_H \geq c_L + (c_H\text{-}c_L)/D(.) \text{ or, } p_H = c_H + (c_H\text{-}c_L)(1\text{-}D)/D$$

Thus, school H must earn a markup above his/her marginal cost, c_H, to induce separation in the market.[8] As the precision of the signal declines, t_H^* decreases, the markup required to sustain separation increases. Intuitively, the mass of parents who receive an inconsistent signal when the school charges p_H but produces L is smaller. For separation to hold, it must be that the losses from cheating are larger to compensate for the gain in the number of parents who are "fooled" and pay the high price for low quality. The only instrument available to increase these losses is p_H, and therefore, in equilibrium, the p_H that can sustain a separating equilibrium must increase as the signal deteriorates.

When does separation in the market fail? At sufficiently imprecise signals, the markup required to sustain separation between any two given quality levels will be very high. But, at this very high price, the demand for the H schools may be too low to sustain equilibrium and one of two things can happen. Either the market will pool at a quality level determined by the distribution of valuation over quality in the population. Thus, we will observe a single quality and price combination in the market. Or, the schools will choose different quality levels at a greater distance than what would arise under perfect information. With greater product differentiation, parents are better able to distinguish between the schools, which would reduce the required markups to sustain separation. Thus, imprecise information can lead to excess quality differentiation in the market.[9]

The Impact of Report Cards

This canonical formulation highlights the link between the impact of third party information, such as report cards, and the nature of the initial equilibrium. When the initial equilibrium is separating, additional information provision through report cards should lead to a *greater* price

---

[8] See Wolinsky (1983) for an equilibrium refinement that narrows the set of equilibria to prices where the inequality holds exactly; this refinement is similar to a Cho-Kreps criterion.
[9] As school quality differences narrow, there are two opposing effects. On the one hand, cost differences narrow, reducing the incentive to cheat; on the other, the benefits of cheating increase as the fraction of consumers who receive an inconsistent signal declines. If the former is outweighed by the latter, then for a given information environment and demand, separating two schools that are closer in quality requires higher markups, leading to excessive differentiation in equilibrium.

decline for the H schools—that is, schools with initially higher test scores. Thus, as a consequence of the intervention, the price-quality gradient should *decline* in villages that receive report cards relative to control villages. Further, we should observe that some schools will occupy a "middle" quality space, reducing excess differentiation in the market, particularly at relatively higher levels of quality.

Alternatively, when the initial equilibrium is pooling (because signals are very imprecise to begin with) additional information through report cards will lead to an increase in quality differentiation and a positive price-quality gradient will emerge. In this case, report cards would lead to *greater* separation with better quality schools charging *higher* prices - the opposite of what one would find in the case of an initially separating equilibrium.

<u>Further predictions and public schools</u>

Further predictions of the impact of report cards on the distribution of school quality and on which schools will choose to alter their quality will depend on the structure of underlying demand and cost conditions. To see this, consider the following case where we observe the same equilibrium prior to the arrival of information, but different outcomes once the additional (report card) information is given. Suppose schools can choose from *L*, *M* and *H* levels of quality but the initial (asymmetric) information equilibrium we observe has only *L* and *H* schools—i.e. there is excessive differentiation. What happens after third party information is introduced depends on the nature of parental demand and the cost of changing quality.

If consumers have high valuation for quality, under perfect information the unique equilibrium may be *(M, p_M)* and *(H, p_H)*. With asymmetric information, *M* and *H* together may require too high a markup to be sustainable and so we observe *L* and *H* instead. In this case, report cards will lead to *L* schools upgrading their quality to *M*, with those schools that cannot undertake such quality investments shutting down. Alternatively, if consumers have low valuation for quality, under perfect information, the optimal distribution may be *L* and *M*. Again, this may require too high a markup under asymmetric information so that we observe *L* and *H* instead. In this case, the report cards would allow *H* schools to reduce their quality to *M* and some *L* schools to move to quality *M* depending on their cost structure. Finally, if consumer valuation is such that the equilibrium under perfect information supports *L*, *M* and *H* (but as before we observe only *L* and *H* under asymmetric information), depending on the underlying cost structure, after report card provision some *L* schools may upgrade to *M* or some *H* schools may downgrade to *M*.

The canonical models also have little to offer about the performance of public schools without knowing their objective function and the extent to which they face market discipline. Given that their objective function may be he hard to define/ambiguous, we may expect that the information has no impact (as in Banerjee et al., 2010); alternatively, if there are non-monetary costs such as social exclusion of teachers when public school quality is low, such schools may choose to

increase quality. What the standard model *does* allow for is the incorporation of the public school as an "outside" option whose quality is exogenously affected by the report cards. Any such change in public school quality then has a direct impact on the distribution of private school quality and price; in particular, given that the public schools are typically of lower quality in our data, their presence in the market will generate greater pressure on the low-quality private schools to increase quality. Because these low-quality private schools are also in a separating equilibrium relative to the public schools, we also expect their prices (i.e. markups) to decline for the same reason they did for high quality private schools, although not by as much as for the latter.

Ultimately, on both these fronts—the change in distribution of school quality and the public sector response—the canonical models outlined above can help us interpret the results, but without imposing further structure, do not offer specific predictions on what these results may be.

Online Appendix 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:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

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.

Online Appendix Table IIA: Child Attrition by Treatment

Type of Attriter	(1) Report Card
All Attriters	0.008
	(0.014)
Absentees	-0.001
	(0.011)
Dropouts	0.006
	(0.004)
Untracked/Left Villages	0.002
	(0.007)

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

This table presents attrition checks by treatment status for different types of attriters. Column 1 presents regression estimates with district fixed effects and standard errors clustered at the village level. We note there is no differential attrition by treatment status.

Online Appendix Table IIB: Differential Attrition Checks Across Baseline Characteristics

	Attriters		Absentees		Dropouts		Untracked or Left Village	
	Control	Difference	Control	Difference	Control	Difference	Control	Difference
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

This table checks for differences across baseline child and parental characteristics for attriters, i.e. children tested in the first year but not the second year. Column 1 shows the mean of the control group across various child and parental characteristics. Column 2 tests for differences across treatment and control (treatment-control) after clustering standard errors at the village level and including district fixed effects since randomization was stratified by district. Columns 3-8 repeat Columns 1 and 2 while considering different types of attriters. We note that attriters are not different across a range of child and household characteristics.

Online Appendix Table III: 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:
* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level
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: Round 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.

Online Appendix Table IV: Fees - Impact by Provider Quality using Different Definitions of Quality

	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*Private	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*Private	-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)
Village 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 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 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]

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

This table runs robustness checks on the heterogenous fee results in Table V by considering different definitions of high and low-achieving schools and a series of controls. Columns 1 to 3 define high as top 40th percentile within the village. Columns 4 to 6 define high as above the village median. Columns 7 to 9 define high as above the median 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/inconsistent data in rounds 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.

Online Appendix Table V: 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 * 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 * 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 Private		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 * 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 private school, low ach child in sample — 0.383	Low private school, low ach child in school — 0.336
	[0.0286]	[0.060]
	Low private school, high ach child in sample — 0.204	Low private school, high ach child in school — 0.267
	[0.005]	[0.001]
	High private school, low ach child in sample — 0.0113	High private school, low ach child in school — -0.0393
	[0.871]	[0.460]
	High private school, high ach child in sample — -0.0588	High 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:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

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 inititally high or low achieving 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 refers to a dummy variable for high-scoring schools; 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 have extended controls, which 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.

Online Appendix Table VI: Channels

	Household Inputs (Year 2)				School Inputs (Year 2)			
	Child's Time Spent in School and Preparing for School (Minutes per day)	Parental Time Spent on Education with Kids (Hours per week)	Child's Time Spent on School Work Outside of School (Minutes per day)	Parental Spending on Education Not Including School Fees (Rupees)	Break/Recess Time (Minutes per day)	Percent of Teachers with Matric Qualifications	Basic Infrastructure Index	Extra Infrastructure Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Report Card (RC)	12.90	-1.566	17.36	-163.1	-12.39*	0.0129	-0.0510	-0.0231
	(20.96)	(1.051)	(13.40)	(160.4)	(6.875)	(0.0499)	(0.246)	(0.202)
Government School (Gov)	32.59***	-1.142	-19.42*	-105.6	-1.932	0.0248	-0.722***	-0.516***
	(9.609)	(0.791)	(10.68)	(134.1)	(5.760)	(0.0344)	(0.172)	(0.138)
RC * Gov	-13.47	1.367	-17.43	29.63	14.03*	0.0118	0.0774	0.0978
	(21.93)	(1.067)	(13.26)	(174.9)	(7.102)	(0.0534)	(0.257)	(0.208)
High* Private School	15.25	-0.0202	-13.39	132.2	-6.244	0.0314	-0.111	0.106
	(15.42)	(0.961)	(13.03)	(193.8)	(6.008)	(0.0374)	(0.197)	(0.132)
RC * High* Private School	-4.884	0.740	-0.295	-74.74	15.05*	0.0211	0.0204	-0.0175
	(24.85)	(1.281)	(17.98)	(239.2)	(7.884)	(0.0539)	(0.294)	(0.232)
Baseline	0.0694	0.179***	0.105***	0.334***	0.105*	0.792***	0.726***	0.715***
	(0.0489)	(0.0317)	(0.0350)	(0.0687)	(0.0552)	(0.0266)	(0.0430)	(0.0257)
Controls	Village	Village	Village	Village	Village	Village	Village	Village
Observations	899	930	898	953	782	783	776	783
R-Squared	0.061	0.0910	0.126	0.136	0.038	0.659	0.62	0.727
SUBGROUP POINT ESTIMATES, F-TEST p-VALUES IN BRACKETS								
Low private school	12.90	-1.566	17.36	-163.1	-12.39	0.0129	-0.0510	-0.0231
	[0.539]	[0.139]	[0.198]	[0.311]	[0.074]	[0.796]	[0.836]	[0.909]
High private school	8.019	-0.826	17.07	-237.8	2.658	0.0340	-0.0306	-0.0407
	[0.550]	[0.242]	[0.164]	[0.212]	[0.462]	[0.177]	[0.840]	[0.691]
Government school	-0.566	-0.199	-0.0673	-133.5	1.632	0.0247	0.0264	0.0747
	[0.937]	[0.511]	[0.991]	[0.104]	[0.450]	[0.0708]	[0.678]	[0.308]
Baseline Depvar (Mean)	421.251	3.438	96.748	971.005	32.64	0.561	-0.067	-0.008

Notes:

* significant at 10 percent level **significant at 5 percent level *** significant at 1 percent level

This table examines treatment effects on several school and household inputs as possible channels of impact for changes in test scores. The outcome variables are: Time spent in school and preparing for school (column 1); Parental time (reading and helping) spent on education with kids (column 2); Child time spend on school work outside of school (column 3); and parental non-fee spending on education (column 4); Break Time (column 5); Percent of teachers with matric qualification (column 6); Basic infrastructure index - rooms, blackboards, and other non-building material (column 7); Extra infrastrucutre index - library, computer facility, or fans (column 8). This data comes from the school survey and the household survey for children who were matched to the school testing roster; the household data is at the household X school level. The observations from schools surveys are less than 800 due to school closure and missing data; we restrict to open schools only in these regressions. The observations from household survey differ slightly across regressions because of missing LHS/RHS values. All regressions control for baseline value of the dependent variable as well as additional village controls, the same ones used everywhere. All of the above regressions have standard errors clustered at the village level and include district fixed effects. 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 of the dependent variable for the sample in these regressions.