

*A self-ordained professor's tongue
Too serious to fool
Spouted out that liberty
Is just equality in school
"Equality," I spoke the word
As if a wedding vow
Ah, but I was so much older then
I'm younger than that now.*

Bob Dylan, My Back Pages

Introduction¹

In India, as elsewhere, there are increasing proposals for an increase in the localization of schooling, either through decentralization of responsibilities for the operation of schools from state governments to elected lower tiers of government (Pande and Pritchett 2007) or various types of “money follows the student” schemes. These proposals are, in part, a response to the accumulation of results from independent testing showing that the actual learning achievement in the government schools in India was frequently abysmally low. A small study by Pratihci Trust (an NGO founded by Amartya Sen) in West Bengal found that among third and fourth grade students in government schools (who had not hired private tutors) only 7 percent could write their own name. Another study in Andhra Pradesh revealed only 12 percent of students tested in government schools in standards (grades) 2 to 5 could do *single digit* subtraction (e.g. 9-2=?). The nationwide ASER results show an enormous state and district heterogeneity in learning outcomes, including both extremely high and shockingly low achievement. Das and Zajonc (2009) use assessments of 14 year olds in Rajasthan and Orissa to construct international comparisons and find that, while India has some of the best results in the world, it also has much of the worst.

¹ We would like to thank, without implication, Michael Clemens, Deon Filmer, Ruth Levine, Ricardo Hausmann, Michael Woolcock, Asim Khwaja Jishnu Das, and Sangeeta Goyal for helpful discussions.

Yet in my three years working around education specialists (both Indian and international) they consistently dismissed proposals for localization (either through decentralization or through “money follows the student schemes”) as “too risky.” They (claimed to) fear that localization, even if it might improve results in some places, or even on average, would lead to greater heterogeneity and inequality in the quality of schooling. They believe that the direct hierarchical control of large scale systems allows the enforcement of standards, uniformity, and the (re)allocation of resources they believe is necessary to equalize quality across schools. The pursuit of equality has been an important argument for large-scale, state-run, highly centralized (e.g. nation-state or state/province) educational systems with control exercised directly through the physical ownership and direct operation of schools. The combination of the increased importance of formal education as a key to life chances, the spread of social equality (in aspiration and rhetoric and, more hesitantly, in practice), and the deepening of democratic accountability of governments, has made equal opportunity in education a central *rationale* for direct and hierarchical government control of schools². But does hierarchical control of basic education really lead to more equality across schools than alternative more localized arrangements? How does one square the seemingly “common sense” view that equality requires top-down control with the empirical evidence about heterogeneity?

James Scott’s (1998) *Seeing Like a State* has a revealing sub-title (“How Certain Schemes to Improve the Human Condition Have Failed”), a compelling metaphor, and a

² Note that this is a claim about “arguments” and “rationales” for centralized direct production of schooling rather than “reasons.” The relationship between rationales/arguments and reasons is complex as rationales can just be rhetoric that masks true reasons. A claim that something has been a “rationale” for X is merely descriptive about claims that have been advanced, whereas the assertion something is the “reason” for X is a much stronger claim about actual causes and effects.

key insight. Scott's compelling metaphor is "vision" as what we "see" is not raw external physical phenomena but is structured by biological processes³. Scott's key insight is that "bureaucratic high modernism," which characterizes the administrative apparatus of the modern nation state, (re)constructs social realities through dramatic simplifications to make life "legible" or "visible" to the state. An explicit goal of "bureaucratic high modernism" is to blind the state to characteristics on which localized social processes had historically depended--kith, clan, class, caste, religion, race, gender—in order to produce equal treatment of all citizens by the state's administrative apparatus. Like many key insights, the notion that different organizational structures and sizes can differentially "see" and process information has been (re)discovered many times, with different names and in different disciplines⁴.

A popular tool in education, particularly for support by external assistance agencies, is the development and roll-out of EMIS (Education Management Information Systems)⁵. At best these systems make the EMIS-visible characteristics of schools and teachers

³ Human optical and brain processes both filter out phenomena and filter in structure that can lead us to "see" things that are not really "out there." For wonderful examples of how optical-neural processing produces visual experiences that are not part of any physical input see Hoffman's *Visual Intelligence* (1998) or to actually see them for yourself: <http://www.cogsci.uci.edu/~ddhoff/index.html>.

⁴ Indeed, integral to Weber's definition of a bureaucracy—the consistent application of rules—is that only those characteristics specified in the rules legitimately affect decision making. Jensen and Meckling (1992) distinguish between "specific" and "general" information based on how easily the information can be transferred across agents within an organization. Stein (2002) uses the distinction between "hard" and "soft" information to explain banking practices. An enormous literature on institutions based on transactions costs explores the implications for organizational design and markets versus hierarchical control based on the ease with which information is transmitted and verified (Williamson). Aghion and Tirole (1997) argue the ease with which an agent's information can be verified affects "formal" versus "real" authority in organizations.

⁵ That attempt to create comprehensive data bases for the management of education systems has been supported by the World Bank and other donors in more than 40 countries. Moreover, beyond the specifically "EMIS" the approach to facilitating the measurement and expansion in hard inputs is a hallmark of many externally supported approaches to education, including some would argue current initiatives like the MDGs and the Fast Track Initiative.

available to the organizational hierarchy with greater clarity and frequency⁶. The stress on “management” through EMIS focuses on “standards” for the EMIS-visible aspects of schools (e.g. size of the building, availability of toilets, number of enrolled students, attendance, availability of textbooks, etc) and teachers (e.g. age, seniority, qualifications, training).

How important are the EMIS-visible aspects of schools? Are the inputs that are “seen” and hence capable of being equalized by a public sector, centralized, hierarchical, organization sufficiently important in the process of achieving educational goals that control produces substantial equalization or does EMIS-visible equalization produce the merely the illusion of equality?

The consequences of blinding of the state by operating of schools within large hierarchical bureaucracies for the level and variability of student performance depend on (a) the nature of the education and (b) the strength of the implementation environment. Blinding weak states by relying on centrally controlled “bureaucratic high modern” educational systems in weak implementation environments could be “lose-lose” and produce *lower* average performance with *higher* variance and produce on the illusion of equality⁷.

⁶ That is, even when they are successfully implemented, the data is actually accurate and up to date, and actually used in decision making, which has been rare in practice.

⁷ We alert the reader that this paper is all about education and not really about education at all, or at least principally as an example of a broader phenomenon. Education happens to be the area of “development” endeavors in which this tension between centralizing control over inputs in order to achieve (or appear to achieve) goals about equality and the corresponding loss of control over outputs when the relevant dimensions cease to be visible, particularly when state and social capability is low, can be the most easily (though still controversially) quantified. But this tension is the key to understanding the history of “development” and why the “one size fits all modernization” has failed where the rubber hits the road in so many “discretionary transaction intensive” activities (e.g. curative health care, rural water supply, irrigation, policing, tax collection) (Pritchett and Woolcock 2004). Finding the practical way forward in promoting development on the balance between transplanted established best practice and “accelerated modernism” versus simply allowing for local processes to play themselves is a intellectual challenge of our time (Andrews, Pritchett, and Woolcock, forthcoming).

This paper has three sections. The first section uses recent detailed (but not internationally comparable) data from Pakistan and India to show frequently *much* higher inequality in (adjusted) learning achievement across schools in the government sector compared to a low-quality, inexpensive, effectively unregulated private sector. The second illustrates analytically and with simulation results the connection between educational outcomes and centralized versus localized processes of decision making. The third uses data from the PISA examinations to simply describe and compare the variation across schools in measured student learning achievement (netting out student background characteristics) between public and private schools (as a crude proxy for “centralized” versus “localized” processes). We find that in weak states the variability in school quality is frequently *much* higher than the variability in school quality among private providers (while in strong states the state schools are often much more equal)⁸.

⁸ We are using measures from standardized tests simply as convenient proxies for school quality not making a normative assertion that this is how the output of the education sector or equality should be measured. Education is the process of preparing children to fulfill their potential in their adult roles in society, polity and economy, a process of training and socialization that involves parents, relatives, society at large, peers and schools. Schooling is one element of education and by “school quality” we mean a *school’s causal increment to educational goals*. This conceptual definition already reveals why any proposed measure of quality will be controversial. First, school quality likely differs across education goals. Some schools are better at math, some at promoting citizenship, while others at encouraging creativity. Therefore any proposed single measure of “school quality” is in itself an assertion about the relative importance of various educational goals, which is bound to be controversial. Second, school quality has to distinguish between the performance of students in that school and the school’s *causal contribution* to that performance, its “value added.” Since student background characteristics (e.g. parental education, household wealth) are often important correlates of measured educational performance in many dimensions many “good schools” simply have good students. Third, school quality is student specific (e.g. better match for students learning style(s)) and can vary across initial ability. Some schools may be better at bringing out the best in the best, others at bringing lagging performers up to snuff. Any single measure of school quality specifies (a) the relevant domains on which performance is to be measured, (b) how that performance is to be assessed, (c) how the various domains are to be aggregated (if at all), and (d) how the distribution of performance increments across students is to be aggregated (e.g. the average increment, the increment to the best students, worst students. We, of necessity of convenience not commitment, will mostly settle for simple measures of school value added on learning achievement, aware all the while this focus on learning achievement runs the risk of simply “seeing like an economist” as if these scores were the value of education., which we emphatically do not believe.

I. Blinding weak states

*Thus we have sacrificed what was ours, but have
not gained in its place anything that might be
considered its equivalent; we have lost our world,
but we have not entered another.*

Dewantara (about Indonesian schooling)

We start with empirical illustrations from existing studies of the variability of school quality in the Punjab state of Pakistan and in the states of Rajasthan and Orissa in India. These are “weak” states in two senses. First, there is low general capacity for policy or program implementation, as is indicated by various measures of government effectiveness, corruption or “progressive deterioration of public services.”⁹ Second, with regard to education in particular, elsewhere (Pritchett and Viarengo 2008b) we classify these countries as having a relative “weak” commitment to the expansion of schooling as a means of expanding basic skills in the population¹⁰. In both of these countries there is a large and growing education private sector, largely without government support or effective regulation. This allows us to compare the variability across schools in the public and private sectors.

⁹ While this is perhaps obvious for Pakistan, for a discussion of India as a “flailing state” see Pritchett (2009). On the Fund for Peace 1 to 10 (10 the worst) ranking of “Progressive deterioration of public services” in 2007 Pakistan was 7.1, India was 6.7. In the World Bank governance indicators of “Government Effectiveness” India was in the 57th percentile, Pakistan the 28th, and in “Control of Corruption” India was in the 47th percentile and Pakistan in the 21st.

¹⁰ The difference is that generally “weak” states may have a strong commitment to schooling for ideological grounds stemming from a desire to control socialization (e.g. China in the Cultural Revolution). In Pritchett and Viarengo (2008b) we argue that in countries with a weak commitment to schooling either on ideological grounds or as a means of expanding skills then: (i) the overall level of schooling would be responsive to the demand for schooling, not the supply, so if demand is low then expansion would be low and only if private returns are high overall levels would increase and (ii) the share of private schooling would be relatively high, both because the state would tolerate such schooling and because the state supplied schooling is likely to be of poor quality (since broad based quality education would take a back seat to other agendas, including the agenda of politically mobilizing teachers and patronage).

I.A) Pakistan

The Learning and Educational Achievement in Punjab Schools (LEAPS) survey was carried out in 812 private and public schools in 112 villages in the Punjab province of Pakistan. The overall sample includes learning outcomes for 12,000 students in grade III in Urdu, English and Mathematics together with detailed information on schools, teachers and family background. This report provides three key facts about learning quality in primary education in Pakistan (Punjab). First, learning achievement is low in public schools and is lower than in private schools even after controlling for student background characteristics. Second, the variance in estimates of school quality, measured as value-added in learning achievement, is substantially higher in the public than in the private sector. Third, this larger variance in the public sector is due to the existence of some extremely bad schools in the government sector—while the best public schools are roughly as good as the best private schools, the worst public schools are much worse.

The overall students' performance is low. The overall learning outcomes are not only poor by international standards (Das et al. 2006) but they are also far below the standards set at each grade-level (LEAPS 2007, p. iii). For example, 35 percent of students in grade I cannot subtract single-digit numbers, 76 percent of students in grade II cannot read and write the time (for example, when the clock shows 3:40), and 99 percent of students at the end of grade IV are not able to subtract fractions (e.g., $7/5-5/4$)¹¹.

There has been a massive expansion of private schooling in the Punjab with as many as one-third of children enrolled in the private sector. This sector is almost entirely “Mom and Pop” single schools run on a for profit basis at extremely low costs (the traditional elite private schools and the religiously oriented *madradas* are a tiny share of

¹¹ LEAPS (2007, p.iii, table 1).

enrollment). This private sector has flourished largely without state support (e.g. there are no significant “voucher” or “scholarship” like programs) and mostly without effective government regulation of any kind.

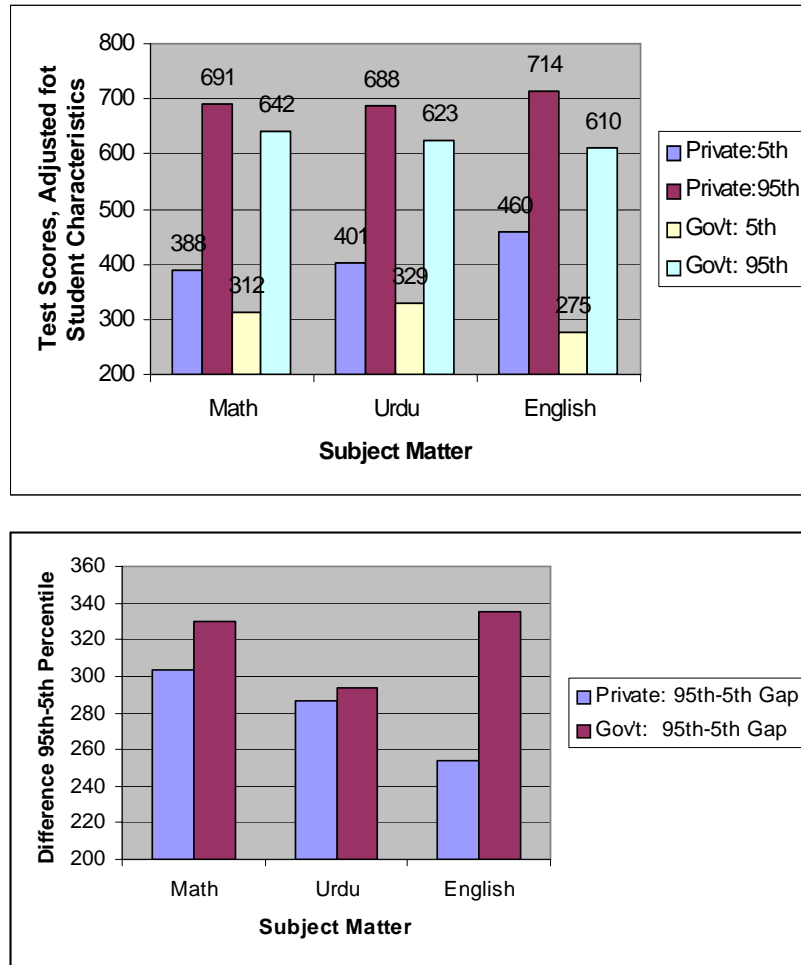
Analysis of the data shows that children in public schools are behind students in private schools by between 1.5 and 2 years (a child in private school in Grade III has the same performance as a child in public school in Grade 5) in the three tested subjects of English, Mathematics and Urdu (LEAPS 2007, p.31). Unlike in many countries, in which the most important correlates of differences in student achievement are student and student household characteristics, in LEAPS the largest gaps are between schools—much of this between government and private schools. The learning achievement gap of being in a private school versus a government school is 12 times as large as the difference between a child from a rich versus a poor family (Das, Pandey, and Zajonc, 2006)¹².

One can imagine that the centralized government schools might perform worse *on average* than the completely localized and unregulated private schools but that the central control allowed the achievement of greater *equality* across schools. The opposite is true. The researchers used regression methods to back out an estimate of adjusted school quality, which is the difference between the schools actual average learning achievement and that which would have been expected based on the characteristics of students and their parents—often, for better or worse, referred to as school “value added.” Figure 1 shows the scores in the top adjusted learning achievement schools (the 95th percentile) and the worst (bottom 5th percentile) for Math, Urdu, and English. This particular is

¹² Of course, it is difficult (many would argue impossible without a randomized controlled experiment) to know how much of this gap is because private schools have more value added and how much because of children more likely to do better even with the same school value added choose to attend private schools, but analysis of the data using plausible methods of identifying the causal effect of private schooling suggest that most of the observed difference is causal.

normed to a mean of 500 and the student standard deviation to 150, so that the difference in average learning achievement between students in the 95th and 5th percentile schools of 300 hundred or more is around two full student standard deviations, which, as we shall see below with evidence from a larger number of countries, are enormously large differences across schools.

Figure 1: The gap between the better and worst government schools is substantially larger than the gap in private schools



Source: Derived from Das, Pandey and Zajonc, 2006, table 7

In each of the three subjects, the gap between the best and worst government schools is larger than the gap between the best and worst private schools. In mathematics, for instance, the best private schools score 691 while the best government score 642, a gap of 49 points, while the worst private schools are at 388 (303 points behind the best private). The 5th percentile government schools score 312, which is 330 points behind the 95th percentile government schools. As seen in Figure 1, compared to

mathematics these “gaps between the gaps” are smaller for Urdu and much larger for English.

One implication of the larger performance gap, combined with the lower averages, is that nearly all of the worst schools in adjusted learning achievement are government schools. In mathematics 91 percent of the schools performing in the bottom 10 percent are operated by the government (LEAPS 2007, p.31)).

To put these in perspective, the *additional* disadvantage of a student being in the worst versus the best government school compared to the worst versus the best private school is about 27 points in Mathematics (330 less 303). Comparing this to other correlates of performance, the score gain associated with having *both* parents literate was 16 points and from being in the richest versus poorest third of households was 18 points. The *additional* variability of government schools puts students at substantially greater risk of having lower learning achievement than the disadvantage of coming from a poorer or less educated household.

Two points are worth noting. First, one cannot explain the private-public gap as a result of the public sector being starved for resources versus “elite” private schools with ample inputs. In fact the opposite: government schools spend *twice as much* per student on average than private schools with a median cost of Rs. 2039.2 in the public sector and Rs. 1012.1 in the private sector (LEAPS 2007, p.40)¹³. Second, one cannot attribute the poorer and more variable performance of the government sector to under-paid or under-qualified teachers. Public sector teachers are paid on average *five times* as much as their private counterparts (LEAPS 2007, p.73). By EMIS-visible measures of teacher quality

¹³ Moreover, as this study, like many others, did not find a consistent relationship between school expenditures per student and learning outcomes, hence one cannot “explain” in a proximate sense the greater government gap with larger gaps in expenditures (or inputs).

such as formal qualifications, training, certification the public sector teachers appear if anything more qualified (and there is less variance in their qualifications).

The LEAP report argues that the private and public sectors learning gap is due to two main differences in management, both of which we return to in more detail in Section II). The typical public sector school is part of a huge and highly centralized bureaucracy--the government of Punjab employs roughly 300,000 teachers¹⁴--while the typical private school is a stand along enterprise. The LEAP report argues the second major difference in management is that the focus of the private sector appears to be on the observable outcomes, whereas the focus of the government sector seems to rely more on the observable inputs¹⁵.

III.B) India

The typical quality of the Indian schooling system is low both by international and national standards, but with outstanding performance at the very top. Das and Zajonc (2009) used the test questions administered in two states of India, Rajasthan and Orissa, to construct estimates of learning achievement that are as comparable as possible to the test scores in the Trends in International Mathematics and Science Study (TIMSS)¹⁶. The

¹⁴ Mahmood (2009).

¹⁵ For instance, in the private sector, teachers' salary increases by students' performance and teacher test scores (which is based on their knowledge of the curriculum they have to teach), decreases by the level of teacher's absenteeism whereas in the public sector teachers' compensation does not appear to be related to students' and teachers' test scores and more absence is actually associated with *higher* pay. It appears to be the case that teachers with higher levels of seniority are those who receive higher levels of compensation in the public sectors and are also those with higher rates of absenteeism (a finding, higher teacher status correlated with more absences, found in other countries as well Chaudhury et al 2006). Teachers' absenteeism appears to be less important in the public sector and teachers rarely face any disciplinary measure whereas in private schools teachers are penalized (LEAPS 2007, p.71).

¹⁶ Although we stress, as do Das and Zajonc, that these are *not* TIMSS scores and should not in any case be referred to as TIMSS scores as the organization responsible for quality control played no role in the administration of the exam or processing of the results and these Indian states did not choose to participate in TIMSS.

scores of eight-grade students in these two states score ranked 45th and 42nd compared to the 49 TIMSS participating countries. These constructed test scores also reveal the great variation in the distribution of students' outcomes. The 95th-5th percentile spread was 344 points in Orissa and 329 points in Rajasthan following only South Africa¹⁷ (Das and Zajonc 2009).

In addition to the ASER results mentioned above, there are an increasing number of assessments of student performance, confirming both the low averages and the variability across states. Pandey et al (2008) show that by the end of the 4th grade only 8 percent of students in Uttar Pradesh can divide a three digit number by 1 digit number correctly (e.g., 585/9), only 11 percent of students in Madhya Pradesh, but 41 percent of students in Karnataka. Even division of 2 digit by a 1 digit number can be performed correctly by only 14, 19 and 60 percent of standard (grade) 4 students in UP, MP and Karnataka respectively (Pandey et al. 2008, p.8).

In India, as in Pakistan, there has been a massive expansion in the number of, and enrollments in, private schools. A recent household survey showed that 58 percent of children in urban areas and 24 percent in rural areas attend private schools, which are predominantly private unaided schools¹⁸ (Desai, Dubey, Vanneman, and Banerjee 2008). Recent studies in two Indian states allow the comparison of the *variability* of private and public schools controlling for the observed characteristics of the students. Again, as in Pakistan, the school attended accounts for more than half of the variation in test scores (Goyal 2007a,b). Figure 2 shows the distribution (kernel density) of test scores adjusted

¹⁷ The TIMSS scores are normed so that the OECD mean is 500 and the OECD student standard deviation is equal to 100.

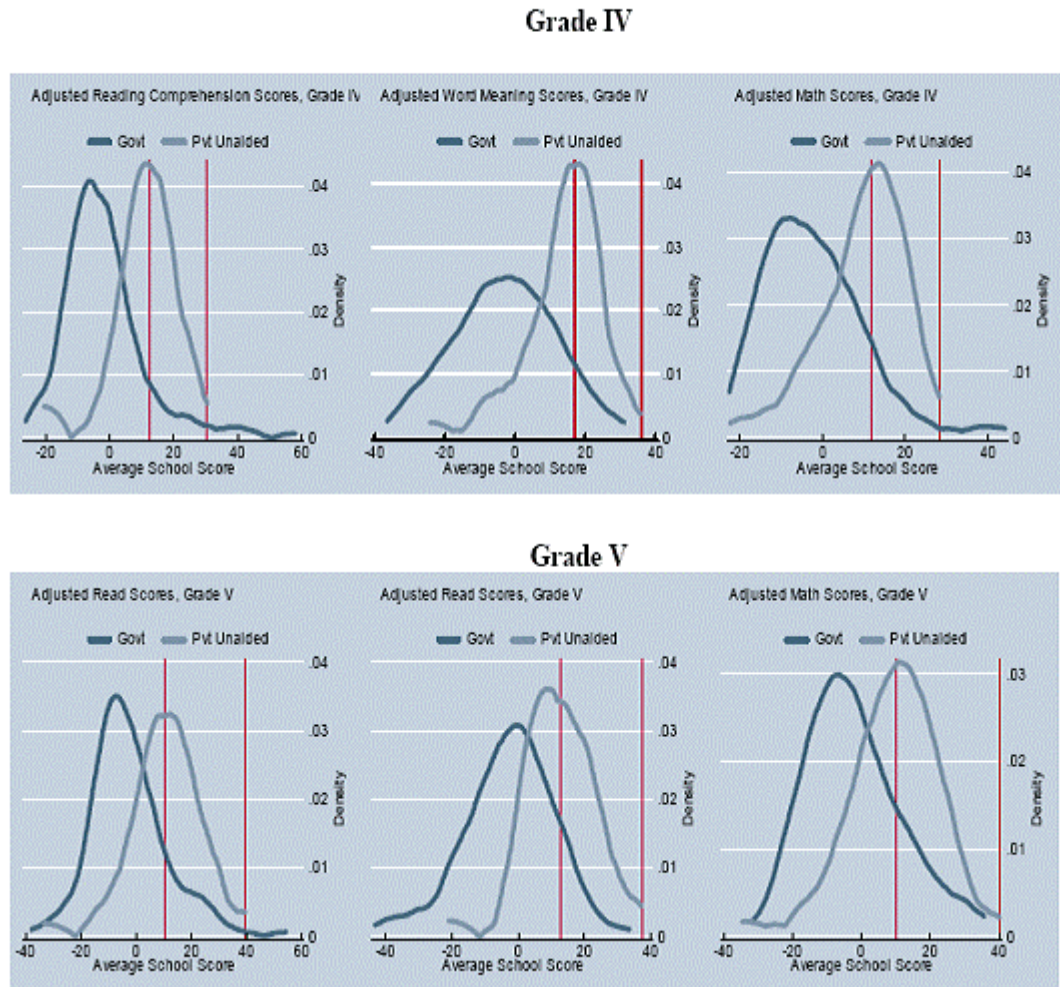
¹⁸ Of course no generalization is true of all of India, in Kerala there is a historically a substantial private aided sector.

for student background characteristics with red lines indicating the performance of the best and median private unaided schools in adjusted learning achievement¹⁹. In Orissa private schools outperform public schools in both 4th and 5th grades and the variation in schooling achievement among government schools appear to be larger than in the private sector. The grade 4 and grade 5 adjusted reading scores are particularly striking, as there is visibly larger variance amongst the government schools. In grade 5 essentially *all* of the schools in the lower tail of performance (more than 20 points below the mean) are government schools.

¹⁹ Since these are “adjusted” scores they are centered on zero as they are essentially the school averages of the residuals.

Figure 2: Higher variability of learning achievement in government schools than private unaided schools (adjusted for student characteristics) in Orissa India

Kernel density distribution of adjusted School Average Scores for Government and Private Unaided Schools in Orissa India, grades 4 and 5, Reading, Word Comprehension and Math



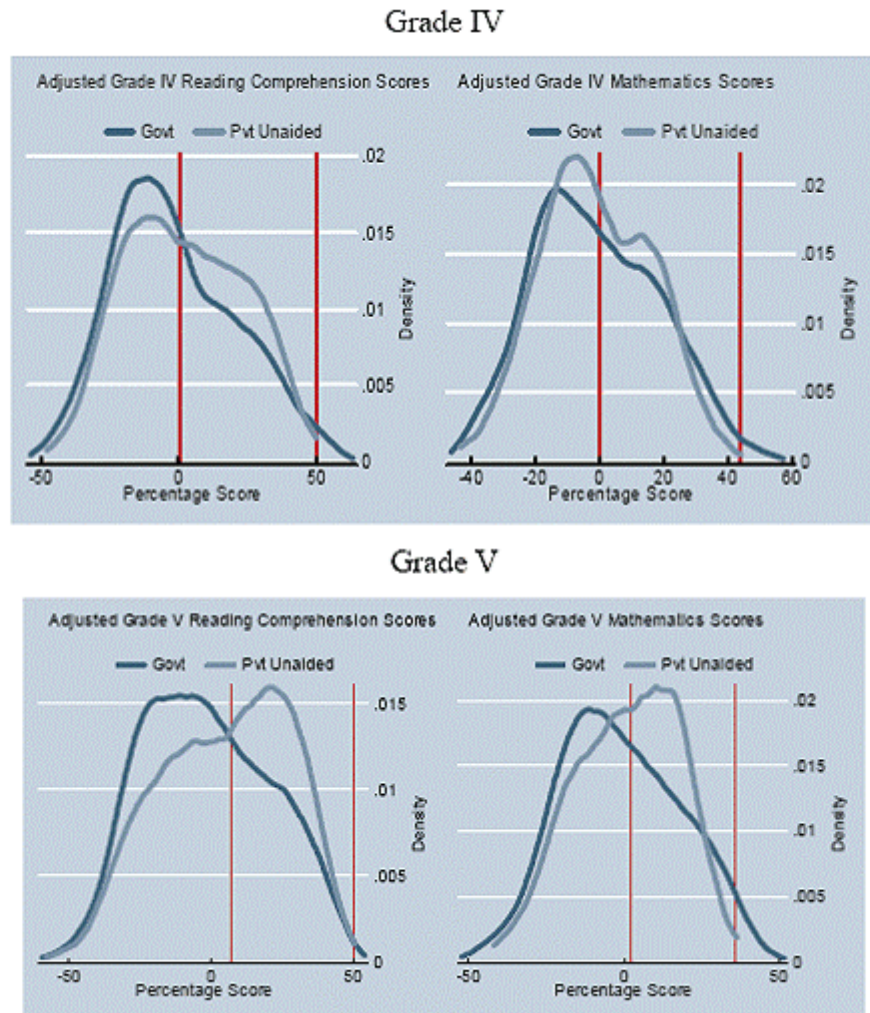
Source: Goyal (2007b)

Interestingly, Rajasthan shows quite distinct results. While the private sector outperforms the public sector on average, the *variability* is quite similar and the higher average of the private sector is due to a larger concentration of schools just above the

average (neither a long tail of high performance in the private sector nor a long tail of poor performance in the public).

Figure 3: Similar variability of learning achievement in government schools than private unaided schools (adjusted for student characteristics) in Rajasthan, India

Kernel density distribution of adjusted School Average Scores for Government and Private Unaided Schools in Orissa India, grades 4 and 5, Reading Comprehension and Math



Source: Goyal (2007a)

A collection of recent studies both nation-wide (DDVB 2008) and in four states (Goyal 2007a,b, 2008a,b) show four results which are similar to each other and similar to the LEAPS results for Punjab, Pakistan. First, the students in the private (unaided) sector

outperform those in government schools by a substantial margin,²⁰ a gap that cannot be explained by characteristics of the students (e.g. parental education, wealth, etc.)²¹.

Second, none of this difference can plausibly be explained by resources or input use. As in Pakistan, the typical cost per student in public schools is roughly twice as high as in private schools. Third, none of this difference can be explained by underpaid or uncertified teachers. Teachers in the government sector make at least twice as much, and up to four to 8 times as much, as teachers in the private unaided sector (Sharma 2009). Teacher qualifications are similar in the two sectors, with private teachers having on average more education but less teacher specific training or certification while public sector teachers having much greater seniority²². Fourth, private schools are managed based on locally observable outputs while the government sector is operated at the state level with typically quite small *de jure* autonomy at the local level over key decisions (until quite recently).

II. Illustrating Potential Illusions of Equality

The results that large, hierarchically controlled, school systems could actually produce larger *variance* across schools in quality than schools that are autonomous,

²⁰ Interestingly, DDVB (2008) shows the private-public (adjusted for student characteristics) gap state by state and there is large variability. There are a small number of states in which the public sector appears equal or better (by about a sixth of a student standard deviation in Maharashtra and Tamil Nadu). In contrast, there are states with a private sector advantage of over a third of a student standard deviation in ten states including all the BIMARU states and in four states the adjusted gap is over ½ a SSD.

²¹ Again, we do not need to make any strong causal claims, but using plausible identification techniques, appears to be causal. The key issue is of course the identification of the private school impact versus the possibility that better learners (for whatever reason) sort into private schools. DDBV (2008) find nearly identical impact effects on learning in reading and mathematics, roughly equal to a grade level, with instruments as in OLS (e.g. in reading .392 in OLS, .363 with switching regression (being in one grade higher .341), or .307 using family fixed effects (being in one grade higher .229)).

²² As Murgai and Pritchett (2007) show, the wages of teachers in the public sector are sufficiently high that there is very little “churning” or turnover of teachers—those who manage to get a public sector job tend to hold it until retirement.

completely without ties to each other, and effectively unregulated is striking. After all, the public sector hierarchical controlled system, at least in principle, has equality as an explicitly stated goal. And, at least in principle, these systems have a full array of instruments to achieve uniformity and equality across schools: (i) explicit programs to ensure uniformity in inputs (at least through each school meeting minimal infrastructure standards), (ii) the ability to reallocate teachers across schools, (iii) control of the budget to reallocate funds across schools, (iv) control of a supervision function to monitor each school. In contrast, no schooling provider in the private sector has equality across all schools as a goal, each school has to provide its own infrastructure, each school has to hire its own teachers, there is generally no reallocation of funds from “rich” to “poor” schools, and there is little or no effective regulation.

While there is an enormous literature about whether private schools outperform public schools *on average*, the inefficiency costs to centrally controlled schools *on average* might be worth bearing to provide *equality*. While competition in the private sector might lead to more pressure for performance (in those dimensions of performance in which parental choice creates pressure) one might still believe this would produce greater *variability* across schools in the private sector, with some elite but also some very bad schools.

Why might this intuition that hierarchically controlled public school systems should always produce more equality than completely localized systems go astray? Think of schooling in countries with unreliable postal services as this raises all three issues:

- How much does increased organizational size and/or hierarchical structure in the public sector lead to control based solely on “hard” information?
- How important are the “hard” versus “soft” controllable inputs to education?

- Does the loss of *de jure* operational control based on soft information by blinding the state to rely only on EMIS-visible inputs (which may lead to **greater** *de facto* teacher discretion and “real authority”) lead to better or worse outcomes?

II.A) Large and Small, Hard and Soft

The state-run post-office is a modern marvel of centralization, but when the private sector delivers mail it often does so through huge centralized organizations as well. While the United State Postal Service has 785,000 employees, the United Parcel Service has 425,000, FedEx 240,000. The centralized organization is amenable to the activity. Delivering the mail can be carried out with just an address: a single piece of hard, easily encoded, third party verifiable, information about the intended recipient. Everything but the address (name, location) can be invisible to the delivering: whether the address is a huge office building, a tiny box, a mansion, a hovel, whether the addressee is tall or short, nice or mean, rich or poor, or for that matter even a person and not a church or a corporation. The ease with which *all of the* relevant information can be transmitted plus economies of scale in coordination leads postal services to be huge organizations as it is productively efficient and organizationally viable.

In contrast, nearly all organizations providing services, especially professional services, are extremely small organizations. The largest law firms in the United States have less than 4000 lawyers (the largest in all of Latin America has 444). Most dentists have traditionally worked in practices of one or two dentists²³. A survey of architects in the USA showed three quarters worked in practices with five or fewer partners. In

²³ A study of “economies of scale” in dentists practices in the USA did not have sufficient observations on practices with more than 5 employees to be able to look at economies of scale above 5.

contrast, in occupations or sector (or even products within sectors)²⁴ in which the quality of the service provided requires detailed adaptation to a specific case, such that the quality of the service provided is based on information that is complex, difficult to encode, hard for a third party to verify, then managing organizations with large numbers of employees becomes very difficult. Hence unless some other positive economies of scale are sufficiently powerful to offset this, organizations will tend to be small, with relationships handled without complex and rigid rules or organizational policies (including human resource policies), with performance assessed directly, and high powered incentives easier to create (e.g. small business owners).

A cost of a large size organization is that it becomes more dependent on “hard” information in its products, policies, and personnel decisions as it is increasingly difficult and costly to convey and utilize “soft” information²⁵. If organization size is exogenous, (or chosen for reasons besides organizational efficiency or efficacy²⁶) one can get mismatch between organizational size and the intrinsic nature of the activity: with services that require the extensive utilization of “soft” information being run in large organizations that can only use “hard” information.

II.B) Is Education Like Delivering the Mail?

We assert education is more dentistry than delivering the mail? The consolidation and centralization of the provision of education into “bureaucratic high modern”

²⁴ For instance Stein (2002) shows that banks change their product lines as they get larger, doing less small business lending, which is contingent on “soft” information. One can also think of segments of sectors, like the difference between low-end “fast food” restaurants which can be very large organizations (through often run as franchises to maintain the incentives of the small units) and high end restaurants in which quality is important, which can only be “chained” with difficulty.

²⁵ A literature uses the costs and benefits of size as a positive model to explain the size distribution of firms across sectors or activities assuming that organizations can explicitly choose their size.

²⁶ The obvious example in the case of schooling is where the size of the jurisdiction for schooling is chosen on political and ideological grounds with a desire for homogenous nation-state socialization (Pritchett 2003).

organizations and systems has benefits, but also costs. Large hierarchical organizations (both public and private) tend to rely on bureaucratic processes and are incapable of explicitly relying on the discretion of their local agents to respond to the full richness and complexity of individual situations. Less and less is “visible” to large organizations. Goals are reduced to organizationally visible and controllable. Centralization creates organizational pressures to simply assert that educational quality *is* the organizationally visible: *input* indicators (e.g. physical infrastructure, inputs), the *formal qualifications* of teachers, and *process compliance*. But how important are the hard, organizationally manageable, EMIS-visible characteristics of schools?

How important are hard non-teaching inputs? Suppose there are N different schools each with school quality (Q^j) where school quality is a mix of K_H hard variables and K_S soft variables and we suppose for simplicity the relationship is linear and there is some idiosyncratic variability as in equation 1²⁷:

$$1) Q^j = \alpha + \beta H^j + \theta S^j + \varepsilon^j$$

If the H (a N by K_H matrix) and S (a N by K_S matrix) variables are uncorrelated then the proportionate reduction in the variance of school quality available from equalizing on hard variables is the simply the regression partial R-squared of all of the H variables:

$$2) \frac{\sigma_Q^2 - \sigma_{Q|H=\bar{H}}^2}{\sigma_Q^2} = R_H^2$$

²⁷ This is much less restrictive than it seems as we can approximate a non-linear function with a linear function powers of H (and their cross-products for interactions) to approximate a wide variety of functional forms.

Appendix 1 uses a simulation model to illustrate that, with low explanatory power of hard variables, even a nearly complete equalization of inputs across schools leads to almost no reduction in the variance of school quality.

What is the partial R-squared in a typical learning achievement regression of the totality of the EMIS-visible characteristics of schools? Unfortunately, it is impossible to review the literature in this way as, for some good reasons and bad²⁸, there has been more emphasis on tests of significance of individual variables and less on the total importance or explanatory power of variables or collections of variables. Moreover, a complete review of the literature would take us too far a field and there is no expectation that this would be uniform across countries or contexts. Nevertheless, we can say that there are many studies in which the *entirety* of the EMIS-visible characteristics of schools provide close to zero explanatory power of differences in student learning achievement (Appendix 2 provides one example that, based on our reviews of the literature for other work (e.g. Pritchett 2004) we feel is typical of hundreds published and unpublished studies)²⁹.

How important are the “hard” characteristics of teachers? A new wave of controlled field experiments in schooling in developing countries is producing rigorous

²⁸ Two good reasons for the reduction in emphasis on the R-square are that (a) it was used quite indiscriminately as a “model selection” tool leading to *ad hoc* and unreliable data mining and (b) if there are identification and endogeneity issues then estimation techniques that maximize R² such as OLS lead to inconsistent estimates of impact.

²⁹ See analytical reviews of empirical literature in Filmer and Pritchett (1997) or Pritchett (2004) or empirical reviews by Hanushek (2003) about the impact of measurable “hard” school inputs on quality as measured by learning achievement. One of the debates we in particular want to avoid is that this literature typically, out of necessity in comparing results across multiple studies uses a standard of “statistical significance” to cumulate results. This leads to sterile debates about zero as the null hypothesis and about statistical significance (Ziliak and McLoskey 2009). T-statistics give no direct information on explanatory power as (a) with sufficient precision in estimation even very small impacts can be “statistically significant” and (b) even variables with practically large effects on learning can have small scope for reducing variance if these are already nearly equal (e.g. many studies find that specific inputs, like having a chalkboard, are strongly correlated with outcomes but that few schools lack a chalkboard).

and reliable estimates of the causal impacts of various interventions, which serves to raise old puzzles with fresh evidence. A recent experiment in Kenya for instance provided funds to schools to reduce class sizes in first grade from their very high levels, which had increased to around 80 children per class in first grade after the introduction of “free” schooling (Duflo, Dupas, and Kremer 2007). In this experiment the newly assigned teacher was either a civil service teacher or a teacher hired directly by the local community on an annual contract. If the children were assigned to a civil service teacher the large reductions in class size made possible had *no* demonstrable impact on either learning of mathematics or literacy or grade progression of children. However, when those same children were assigned a first grade contract teacher, who were hired from a pool of applicants with the same academic qualifications as a civil service teacher, their learning was .2 standard deviations higher and they were much more likely to reach grade 3 than either children in larger classes or children in the same size classes with a civil service teacher.

How is it that community control of the process of hiring teachers can have such enormous impacts on learning outcomes? Everyone remembers teachers from their own schooling, either a teacher from whom they feel they learned a lot, or a teacher who inspired and motivated them, or sparked curiosity, or who they just enjoyed having. We assert *no one* remembers a teacher because of their seniority or what degree they had completed. We remember teachers positively because of their kindness, their inspiration, their excitement and depth of knowledge about the subject, their concern for the students, their creativity, their ability to maintain a lively classroom without chaos, and even endearing personal quirks. We remember teachers negatively for being tedious, boring,

regimented, unable to control the classroom without oppressive discipline, and ill-tempered. All of these “soft” characteristics of teachers are easily *observable* to the untrained (and perhaps even more to the trained) eye but are difficult for high modern bureaucracies, even with functional performance assessment and certainly much less without, to see and act on.

Centralized control over the operation of schooling systems almost inevitably sets in motion a process of blinding the state to the quality of teachers as “soft” characteristics become less and less organizationally actionable. Hard characteristics like a teacher’s age, seniority, formal educational qualifications, participation in pre-service or in-service training, and (perhaps) teacher compliance with schedules, reports, and regulations are easily visible to the state³⁰. Kindness, creativity, “with-it-ness” are not easy for large formal organizations to act on. This can lead to teachers hired based exclusively on hard criteria, like teacher training or certification with the “soft” information based on their actual classroom performance not utilized in retention decisions. Appendix 3 presents a simple simulation showing that a “hard qualifications” based approach to teachers can produce *larger* variance across schools than a strategy of hiring teachers irrespective of qualifications but basing retention decisions on their observed performance (even if performance is imperfectly measured).

De jure and de facto control over decision making. The association of the “public sector” with centralized control and “private sector” with complete localization is of course merely a broad brush characterization (if not caricature) that fails to capture the complex reality for two reasons.

³⁰ As we emphasize below, with weak implementation capacity even simple performance characteristics like attendance become invisible.

First, there are enormous variations across countries in how actual decision making authority is allocated. Some public systems are very centralized over nearly all decisions while even in some large government systems many decisions are delegated to the school or teacher. Woessmann (2007) uses TIMSS data to study how the reported *de jure* allocation of decision-making over various aspects of schooling (e.g. teacher selection, curriculum) affects school performance. Studies of decentralization of school discuss which functions of education are to be retained at which level (Pande and Pritchett 2003)—e.g. countries can choose to move all operational decisions to a very local level while maintaining central control of regulation, monitoring and evaluation (either of inputs or outputs).

Second, the actual pattern of decision making in practice may diverge quite significantly from the reported regulations. Large hierarchical organizations with large spans of control can easily lead to a situation in which the rules and policies are specified in a quite bureaucratic way but the organizational practice *de facto* provides substantial latitude for the exercise of individual discretion. In a very interesting study, Vegas (1999) compares teacher's directly reported sense of autonomy to act and utilize discretion between the public and private sectors in Chile. Strikingly, most of the differences were not between the two sectors, but *within* the sectors, as teachers *reported* very different levels of control over decision-making within the public sector. This leads directly to the question of the next section: how is available *de facto* discretion used?

II.B) Implementation Capacity: Is “work to rule” a threat or an ideal?

Whether centralized systems will be capable of producing greater equality than localized systems depends on the implementation capacity of the state. Some states are strong and some states are weak in policy implementation capacity, an aspect of countries that is not unique to education that cuts across various sectors. Policy implementation capacity depends on (at a minimum) (i) individual capacities, (ii) organizational capabilities, and (iii) the background institutions of governance (legislators, civil service), law (e.g. judicial enforcement), and civil society (e.g. professional associations, social norms). Policy implementation capacity has to be invoked because *de jure* organizational and institutional design cannot explain why systems isomorphic in design have such very different performance. Just mail a letter in different countries. The formal organizational structures, policies, procedures, and standards of post offices are often nearly identical and yet in one country you would put nearly anything in a letter with little or no concern while in other countries you would never send anything of any value by post. When centralized schooling organizations are embedded in otherwise strong states and background institutions (e.g. professionalized teachers) then the system may deliver outcomes *better* than the design. However, a combination of a state organization blinded by large size and bureaucratic structures with a weak policy implementation capacity can lead to disaster: where even the post office works badly (a weak implementation environment) a post office approach to education (a large hierarchical organization) may wreak havoc.

All of us are familiar with organizations that operate much better than their rules, exactly to their rules, or worse than their rules. A large literature in public administration

in the USA seeks to explain why organizations actually work much *better* than their rules, as when the actions of “street level bureaucrats” are at variance with formal procedures it is often to further the goals of the organization. In contrast, in weak implementation environments actions of the street level bureaucrats differ from formal procedures in ways that *worsen* outcomes (e.g. rampant absenteeism/ghost workers, an “effort deficit”, corruption). In high implementation capacity countries “work to rule” is a threat and a recognized “job action” whereas in most countries of the world “work to rule” seems an unattainable ideal.

One example is attendance. In most countries of the world teachers do not use all of their allotted days for sick leave, even though this leave is available and there would be almost zero risk to faking illness to take more days of leave. In contrast, in India and Pakistan absences are much *higher* than available leaves as absenteeism has been rampant. Efforts to limit this have been almost completely ineffective as attempts to discipline workers for unexcused absences run up against the elaborate protections embedded in large public sector organizations. One study in India found absence rates of over one quarter of teachers in any given day. At the same time, in over 3000 interviewed schools people were asked if any teacher had *ever* been disciplined for absence and there was only *one* case mentioned, *ever*. In Pakistan, a calculation suggests that even if someone complains about teacher absence in the public sector, the odds of an adverse action against the teacher is .002. These systems with weak implementation capacity work much worse than their rules.

This matters for the comparison of centralized and localized systems as these alternative arrangements may not be similarly robust to weaknesses in implementation

capacity. An evolutionary metaphor that has been used in connection with organizational robustness is between a spider and a starfish (Bachman and Brofman 2007). A spider spins a web over which information is transmitted to the spider, which then can allocate its limited resource to where the need is greatest. A starfish, in contrast, is a localized creature with no central nervous system whose motion and activity is governed by the very loosely uncoordinated responses of its individual parts. The differential robustness of the two organizational approaches is obvious: the same centralization which potentially allows gains from coordination and directed resources also means that if the spider dies the entire web is worthless and the organization suffers catastrophic failure. In contrast, in many species if you cut the legs off of a starfish you actually get more starfish, as constituent parts are capable of independent operation.

It is an open empirical question whether *directed* systems will be able to achieve greater equality in actual outcomes, even if this were an explicit goal, than *localized* systems. The result will depend on the specifics of individual activities as the outcome depends on the trade-off between the capability of directed centralized systems to produce equalized inputs, the loss of visibility across a range of inputs, and the importance to final outcomes of the localized knowledge that becomes invisible in large organizations. For certain activities, like delivering mail, large hierarchical control is the optimal solution as all that matters for effectively delivering the mail really can be reduced to just three lines of text. For other activities, like marriage, complete localized control is understandably the norm. Is schooling more like mail or marriage?

Table 1: Classification of the “visible” (hard, objective, third party verifiable, encodable) versus “invisible” and the more or less important

	<p>What characteristics are “visible” (legitimate basis for action) is dependent on the organization and its institutional and social context</p>	
	<p>Hard information visible to “bureaucratic high modern” organizations (e.g. age, seniority, formal qualifications, internal process compliance (attendance)) is most important to organizational/unit performance</p>	<p>“Soft” information that is locally observable (e.g. “with-it-ness”, kindness, dynamic, likeable, concerned) is most important to organizational/unit performance</p>
<p>Environments with strong implementation capacity</p>	<p>“Spider” ideal, “Starfish” forgoes economies of scale, coordination (e.g. post office in Germany)</p>	<p>“Starfish” might be ideal, especially with citizen heterogeneity (e.g. Netherlands) but “Spider” will also work adequately (e.g. France)--especially on “harder” dimensions of performance (e.g. schools in Korea)</p>
<p>Environments with weak implementation capacity</p>	<p>“Spider” will be weak but “Starfish” forgoes needed elements of coordination, economies of scale</p>	<p>“Starfish” is a robust organizational form can avoid worst outcomes, weak “Spider” systems may produce “lose-lose”: low average and highly variable performance</p>
<p>“Spider” refers to large, centrally controlled, hierarchically structured organizations/systems, “Starfish” refers to localized systems with many small, only loosely connected units (Bachman and Brofman 2007).</p>		

III. Comparing Public and Private Sector Equality of School Quality

The conceptual distinctions above are only background to the purely descriptive exercise of this section: *how does the variance across schools in adjusted school quality (with learning achievement as a proxy for one dimension of school quality) differ across between the public and private sectors within countries?* This is a *descriptive* exercise in that we are not pretending to test a theory, nor are we asserting causality, we are just presenting new facts.

III.A) Methods

The OECD's Programme for the International Student Assessment (PISA) measures students' conceptual mastery of three subject matters—mathematics, science, and reading—by assessing their ability to apply knowledge to real situations³¹. The PISA study design ensures as much comparability as possible among in testing representative random samples of 15-year-old students in 57 countries. We use the data for the 35 of those countries had a sufficient number of schools of each type for our analysis³². We primarily report the scores on the PISA 2006 Mathematics assessment, with a total

³¹ The main dataset used in this paper is the PISA 2006. This is not motivated by the fact that we believe PISA is better than TIMSS. The two international assessments are different. PISA measures “what 15 year-olds can do” with the skills acquired in school whereas TIMSS assesses the curriculum that students should know when they are close to the end of compulsory schooling (i.e., 8th; we are not going to focus on the 4th grade in this analysis). This is due to data availability. That is, the “school sector” variable is not available in TIMSS 2007. We could derive this variable for only for a limited number of countries from the indicator variables used for sample stratification. An explanation on how this variable has been constructed is available in the appendix. In appendix 5 we present the results of the analysis carried out with the TIMSS data for those countries for which we could construct the school sector variable.

³² Given the relatively small number of schools chosen to represent enrollment many countries had few schools in the private sector (and one country in the public sector). We, somewhat arbitrarily, chose to exclude countries with fewer than ten schools in either sector.

sample size of 290,259 students. We also implemented the analysis reported below for the science and reading assessments, and the overall results were very similar³³.

By using these assessments of learning achievement we are not arguing that “school quality” can be reduced to standardized tests, as educational systems have many legitimate goals in preparing youth for adulthood, but learning achievement is *one* goal of every educational system³⁴. Moreover, in discussions of equalization, gaps in learning achievement across social categories (boys/girls, black/white, rich/poor, etc.) are a key concern. If there were comparable measures of other objectives of schooling systems, such as citizen socialization, amenable to empirical analysis we might get very different results in the variability between the public and private sectors.

School “value added.” Estimating a school’s causal impact on the distribution of student learning achievement would be an extraordinarily complex task as the competences of 15 year olds are the cumulative result of a long-term dynamic obviously influenced by many factors—genetics, early nutrition, family background and encouragement, aptitudes--beyond the school. We are going to do what we can, not what we wish we could do. We take as a measure of an individual school’s learning “value added” the school fixed effects from regressions that control for student background both at the individual and school average level (to allow for peer effects). The students’ achievement file has been combined with data from the background, and

³³ The cross-national correlations of our statistics comparing differences in public and private variability are .84 for reading and .90 for science so the results are quite robust to the subject matter assessed (appendix 4).

³⁴ We wish to stress that by using these are measures we are also not taking on a position of the use of “high stakes” testing (at either the student, school, or system level) nor on the optimal design of assessment from a school system management perspective. Koretz (2008) points out some of the limitations such as: measurement error, test score inflation, sampling error at the aggregate level, possible bias in the assessment of specific groups of students such as special need students or students with limited English proficiency. So that test score gains may reflect changes different from schooling quality.

school and teachers' questionnaires. We simply regress achievement of student i in school j linearly on measures of individual (e.g. age, sex) and family background (i.e., parents' education, occupation status, socio-economic status, number of books at home) characteristics. We also include the school specific averages of those characteristics so not just the student i 's family background but also the family background of other students in the school are controlled for in estimating school "quality."

$$(1) S^{i,j} = \alpha + \beta * X^{i,j} + \delta * \bar{X}^j + \varepsilon^{i,j}$$

From the estimation of the model above, we extract the estimated residual ($\hat{u}^{i,j}$), we regress those on a school fixed effect (f_j) and an error term:

$$(2) \hat{u}^{i,j} = f^j + v^{i,j}$$

This school fixed effect from this regression³⁵ is the average learning achievement for the students in school j compared to what the achievement that would have been predicted (in this simple linear model) for an observational equivalent (in the dimensions measured) student body. This is what we call "adjusted school quality" and we denote by $Q^j(M)|B$ as this measure of the "quality" (Q) of the j^{th} school is not omnibus but specific to the learning domain tested (M for mathematics) and which background characteristics are controlled for (and how) (denoted by |B, or "conditioned on background characteristics B).

Comparing Public and Private. Given that we have observations on both public and private schools we can compare, country by country, measures of the variability of school quality by computing the standard deviation across of a specific measure of adjusted school

³⁵ And there is nothing special about having done it in two stages which was a programming convenience.

learning achievement across schools separately for public and private schools: $\sigma_{Q^j(M)B}^{Pub}$ and $\sigma_{Q^j(M)B}^{Pvt}$.³⁶

Let us be the first to stress how enormously problematic this comparison is. First, since the PISA sampling results are not designed with this purpose in mind there are certainly a variety of ways in which this is a heuristic rather than precise statistic for a well-defined population. Second, we are not particularly interested in the question of public versus private, we are interested in the question of “bureaucratic high modern” (centralized control, rules based, impersonal, driven on state “visible” characteristics) versus “localized” control of schools in which more of the observable becomes visible. The “public” versus “private” is the crudest of proxies for that distinction because in some cases the public sector is quite localized (e.g. in some parts of the US system), the degree of control within the public sector varies widely across countries (e.g. Woessmann 2004). Similarly, what the “private sector” is varies enormously from place to place. In some countries (e.g. the Netherlands, Chile) the private sector receives most of its funding from the government and is regulated. In some countries the private sector is primarily alternative religious education (e.g. Catholic schools). In other countries the private sector is a small sector primarily for the elite. But nevertheless we feel this distinction is a useful descriptive distinction to make. While we would like in principle to compare “centralized” and “localized”, the private sector is at least predominantly “localized”—e.g. schools are managed with much greater degrees of autonomy and discretion in managing their schools and are not constrained to act exclusively

³⁶ There is nothing particularly sacred about the standard deviation as we certainly could have used “centered” measures like the coefficient of variation, or measures of dispersion like the 90th-10th spread, particularly as it appears from the kernel density graphs that often the underlying shapes of the distributions are different (e.g. the “long tails” in some public sector).

on what is “visible” to the state even where many private schools may themselves have some broader affiliation (e.g. “Catholic” schools).

The comparison with the private sector is itself complex, as students and their parents make choices to be in the private versus public sectors and which private school to enroll in. This means that the distribution of students across schools is the result of the actions of parents and students purposively acting on the basis of their own knowledge (e.g. about student aptitude and interests) and within their constraints (e.g. income, distance to possible alternatives). On the one hand this dynamic could drive the variability across public schools to be large (as parents can afford very different qualities of school) but at the same time there is some dynamic at work such that private schools to survive economically must attract students, which at least potentially leads new schools to enter and some failing schools to exit and might induce smaller variability. Without a fully articulated model of the equilibrium we cannot develop predictions about the conditions under which public and private would be theoretically expected to have larger variance. Instead, lets just see.

III.B) Outcomes of the comparison

There are two principal results of our comparisons.

First, across the 34 countries the variability across the private and public sectors is almost exactly the same—the *average* difference ($\sigma_{Q(M)|B}^{Pub} - \sigma_{Q(M)|B}^{Pvt}$) across all of the countries is almost exactly zero. There is no *general* pattern that direct public sector control of the operation of schools results in lower variability in school quality than across the private schools in the same country. This is not a test of a theory or an assertion of causality, just a statement about (relatively) simple descriptive statistics of the observed outcomes: take student scores, net out student background effects, compute school averages and calculate the standard deviation of those school averages across

schools separately for those operated as public versus private schools. Across the countries in this sample the average difference in those summary statistics of variability is almost exactly zero: .01 on a scale with a student standard deviation of 100³⁷.

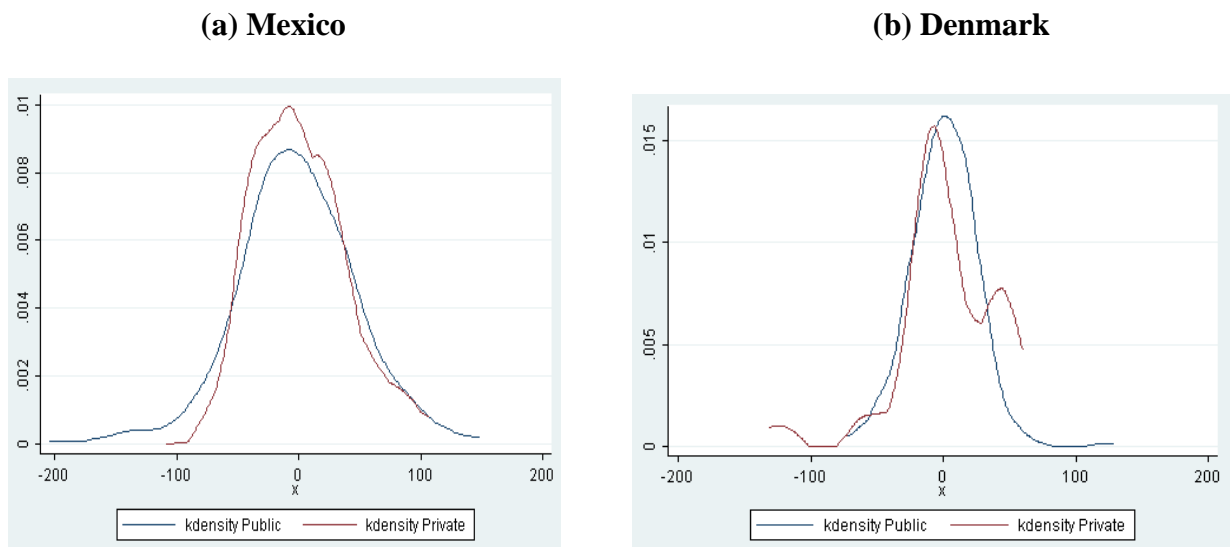
The second descriptive fact is that while *on average* public and private variability is equal, there is variation across countries in the comparison. In some countries the variability is more than 10 points *lower* in the public sector (e.g. Belgium, Germany, Switzerland, Colombia), in some countries variability is more than 10 points *higher* in the public sector (e.g. Uruguay, Greece, Chile, Argentina, Mexico), while in others the variability in the two sectors is nearly equal (e.g. Sweden, Spain, Japan, USA). While the average is zero, the standard deviation of the *difference* in variability between the two sectors ($\sigma_{Q(M)|B}^{Pub} - \sigma_{Q(M)|B}^{Pvt}$) across countries is 11 points.

³⁷ The median difference is -.75, still less than one point on a hundred point student standard deviation scale.

Table 2: Comparisons of the variability in adjusted school average learning achievement across schools in the public and private sector							
Country/Region	Excess of public over private (sorted)	Std. Dev. of adjusted school average mathematics score		Share of enrollment in private schools	Overall standard deviation of scores	Percentage share of the total variability in scores between schools:	
		Public	Private			Public	Private
Hong Kong	18.02	44.7	26.7	92.0	87.1	48.6	30.9
Czech Republic	18.01	31.4	13.4	5.0	89.8	34.7	18.1
Uruguay	14.96	27.0	12.1	21.2	86.0	31.2	14.9
Greece	14.78	22.3	7.5	6.5	82.9	26.8	10.1
Argentina	14.02	32.2	18.2	31.3	88.1	36.1	22.1
Chile	13.94	26.7	12.7	57.7	73.7	34.1	18.3
Slovak Republic	9.34	25.7	16.3	6.9	79.9	31.9	21.4
Mexico	8.91	30.1	21.2	12.1	79.9	39.4	31.5
Ireland	7.08	12.8	5.7	61.9	72.3	17.0	8.1
New Zealand	5.04	8.6	3.6	5.5	81.7	10.5	4.8
Portugal	4.50	19.6	15.1	11.6	79.5	24.7	19.7
Indonesia	4.20	33.9	29.7	32.7	73.8	45.9	41.7
Jordan	3.73	13.6	9.9	14.1	73.1	18.6	15.1
UK	3.33	12.8	9.4	5.0	75.3	17.0	13.1
Hungary	2.44	26.2	23.7	15.6	74.0	35.6	31.6
Japan	1.43	32.6	31.1	28.8	81.5	40.4	38.7
Sweden	1.12	10.9	9.7	7.8	78.8	13.8	12.4
Korea	0.38	22.0	21.6	47.3	82.1	26.1	27.3
Spain	-0.31	9.5	9.8	43.6	78.0	12.1	12.9
USA	-1.21	13.6	14.9	8.9	77.9	17.6	20.0
Italy	-1.56	36.1	37.6	6.4	87.0	41.6	41.7
Israel	-1.92	29.8	31.8	28.5	95.3	31.0	33.9
Austria	-3.96	30.6	34.6	8.7	83.7	36.8	39.8
Netherlands	-4.66	33.3	37.9	68.0	76.1	45.7	48.9
Chinese Taipei	-4.96	24.7	29.7	40.4	90.0	28.6	34.2
Canada	-5.69	13.4	19.1	7.2	78.1	17.4	24.8
Denmark	-8.58	7.8	16.4	23.6	76.6	10.5	20.6
Brazil	-9.20	24.8	34.0	15.9	78.9	32.7	41.3
Thailand	-10.02	13.3	23.3	12.6	73.3	18.5	30.8
Germany	-12.01	28.8	40.8	5.0	80.7	40.6	48.5
Belgium	-16.82	22.8	39.6	68.2	91.0	26.8	43.1
Poland	-17.78	9.4	27.2	6.7	77.1	12.2	32.9
Switzerland	-21.22	24.2	45.4	3.8	84.8	28.7	48.8
Colombia	-25.59	11.2	36.8	18.7	78.8	14.9	41.2
Average	-0.01	22.5	22.5	24.4	80.8	27.9	27.7
Median	0.75	24.5	21.4	14.9	79.7	28.6	29.0
Std Dev	11.02	9.6	11.6	22.8	5.8	11.3	12.8

The graphs of the kernel density of the adjusted school quality reveal a variety of differences. Mexico has higher public sector variability due to a long-left tail (Figure 4a)—essentially all of the schools more than 100 points below the average are public. In contrast, Denmark has lower public than private variability due to more upside variability in the private sector with a concentration of schools (a bump in the distribution) about 50 points above the mean (Figure 4b).

Figure 4: Kernel Distributions of Public and Private Schools:



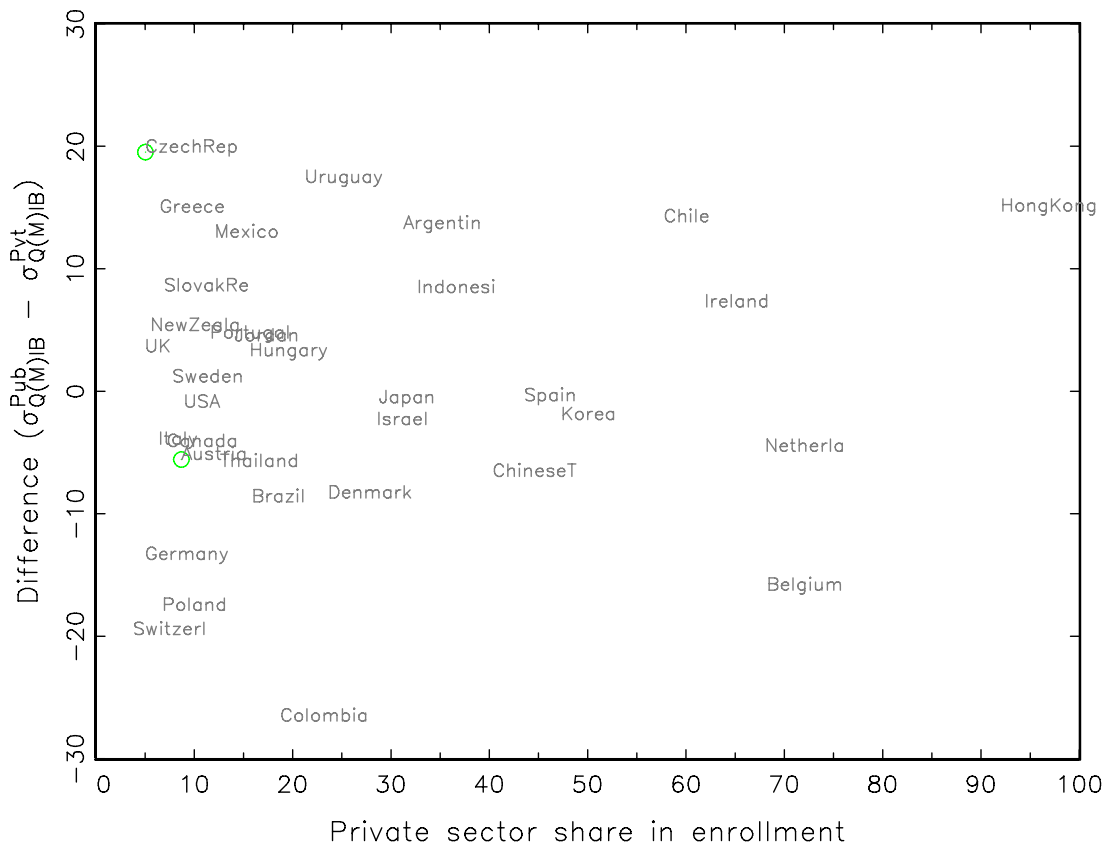
III.C) Variability by public-private education system

As a pure exploratory exercise, we present graphs of the difference between public and private sector variability in school quality (expressed so that a positive number implies the public sector variability is larger) across two characteristics of countries: (i) share of the private sector in enrollment and (ii) a classification of the “type” of schooling system mix of public and private pioneered by Woessmann (2007).

Figure 5 shows the difference in variability between the public and private sectors and the share of the private enrollment in the total enrollment. One might think that

perhaps countries with very small enrollments might have small variance, or perhaps that countries with large private enrollments would have large variance, but there appears to be no relationship at all. At low shares of enrollment in the private sector there are large differences—e.g. Czech Republic has public sector variability 20 points *higher* than private, Switzerland has public sector variability 20 points *lower* than private. But this is also true at very high private sector enrollments, both Chile and Belgium have more than 50 percent of students in private schools but Chile’s public sector variability is much higher than its private sector, in Belgium much lower.

Figure 5: Difference in variability and private share in enrollment



Following Woessmann (2006) we distinguish the share of public sector involvement in *management* of schools from the public sector’s share in *financing*

schools. This allows for various forms of public-private partnership, in some countries most private schools receive a significant share of their funding from public sources but are operated privately, while in some countries publicly managed schools rely on private funding. We divide countries into those with more or less than 75 percent of students enrolled in schools public managed, and those with more or less than 75 percent of school funding coming from public sources³⁸.

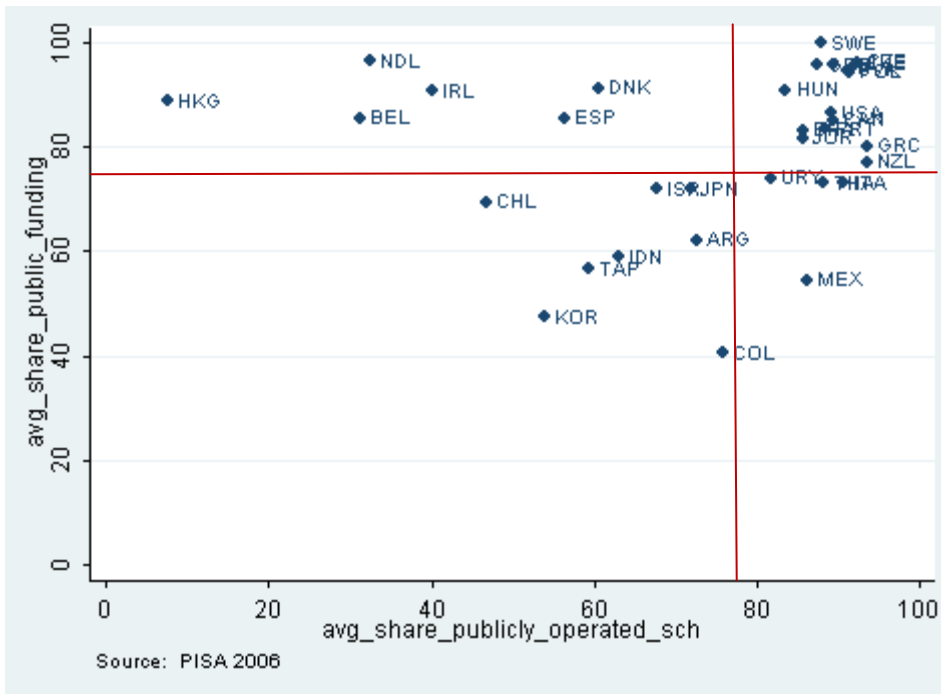
This two-way distinction creates four categories. The upper right (northeast) quadrant are countries where the public sector provides both most of the funding and the management of schools is also more than 75 percent public. This is the “typical” category and includes the most countries, including: Brazil, Canada, Czech Republic, Hungary, Germany, Greece, Jordan, New Zealand, Poland, Portugal, Sweden, United Kingdom and the United States. The upper left (northwest) have systems in which schools receive a high share of public funding (more than 75 percent) but schools are mostly privately operated which includes Hong Kong, the Netherlands, Belgium, Ireland, Spain and Denmark³⁹. In the lower left (southwest) are countries with public management and funding of schools both lower than 75 percent: Argentina, Chile, Chinese Taipei, Colombia, Israel, Japan and Korea. The lower right (southeast) combines countries with a significant share of public management but lower levels of public funding. In contrast to systems in which “public” money flow into privately operated

³⁸ Using PISA 2000 data Woessmann (2006) uses as cut-off points: public funding: 80 percent and share of publicly operated schools: 77 percent, cut-off points that were driven in part by the desire to have at least 4 countries in each category

³⁹ Pritchett and Viarengo (2008b) have a description of the educational public-private partnership in these countries (and a positive theory of why these countries are in the category they are in). Interestingly, Chile just misses being in this category as, although private enrollment is high and public support is high, it is not quite high enough. From the graph it is clear another more relaxed definition, say 65 percent public funding, would add Chile, Israel and Japan into this group.

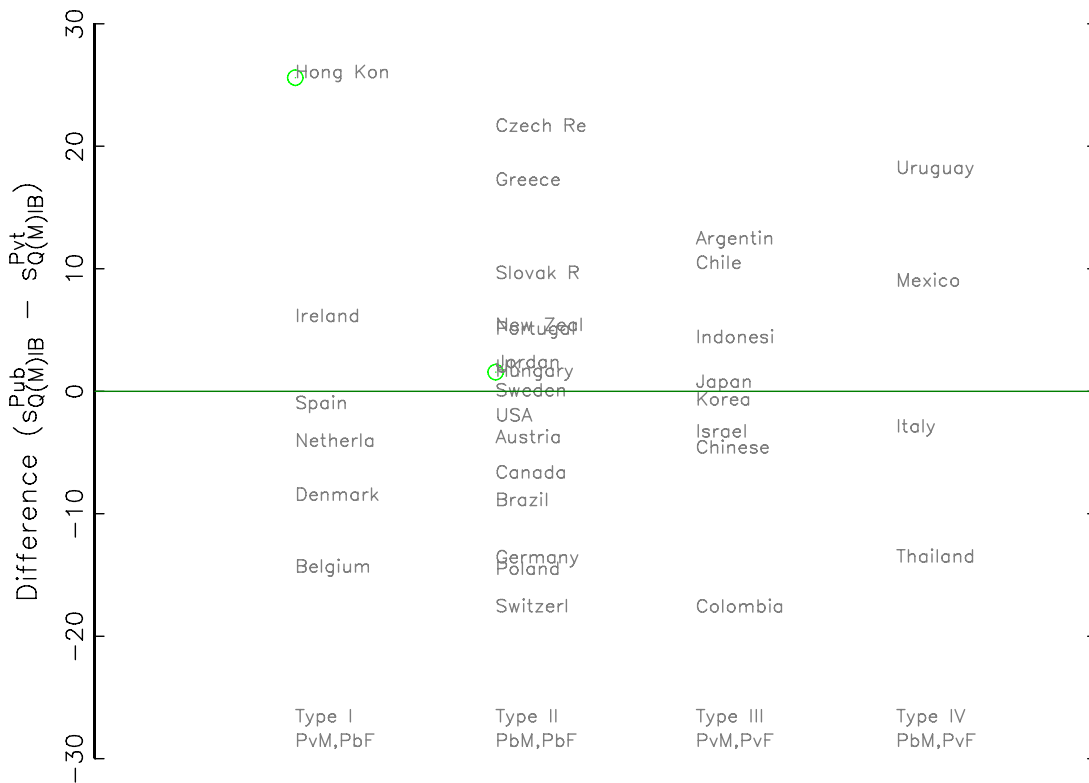
schools, in these systems significant amounts of “private” money flow into publicly managed schools, such as Mexico where schools are almost all publicly managed by public funding is equal to only 55 percent of total school funding.

Figure 6: Share of Public Funding and Public Management of Schools



There is no clear pattern of the gap in the variability of adjusted school quality ($\sigma_{Q(M)|B}^{Pub} - \sigma_{Q(M)|B}^{Pvt}$) with respect to this four-fold division of schooling systems. Each of the four quadrants has countries where the variability is higher in the public sector and countries in which the variability is higher in the private sector. Interestingly, in most of the European “money follows the student systems” (large scale public funding of privately operated schools) the variability is higher in the private sector (Denmark, Netherlands, Belgium) while in Chile (classified as Type III but easily could be a type I “money follows the student” system) the variability is much higher in the public sector.

Figure 7: Excess of public over private variability in adjusted school learning performance in mathematics, by type of school system



Categories of system by public management and funding

III.D) Measures of governmental effectiveness

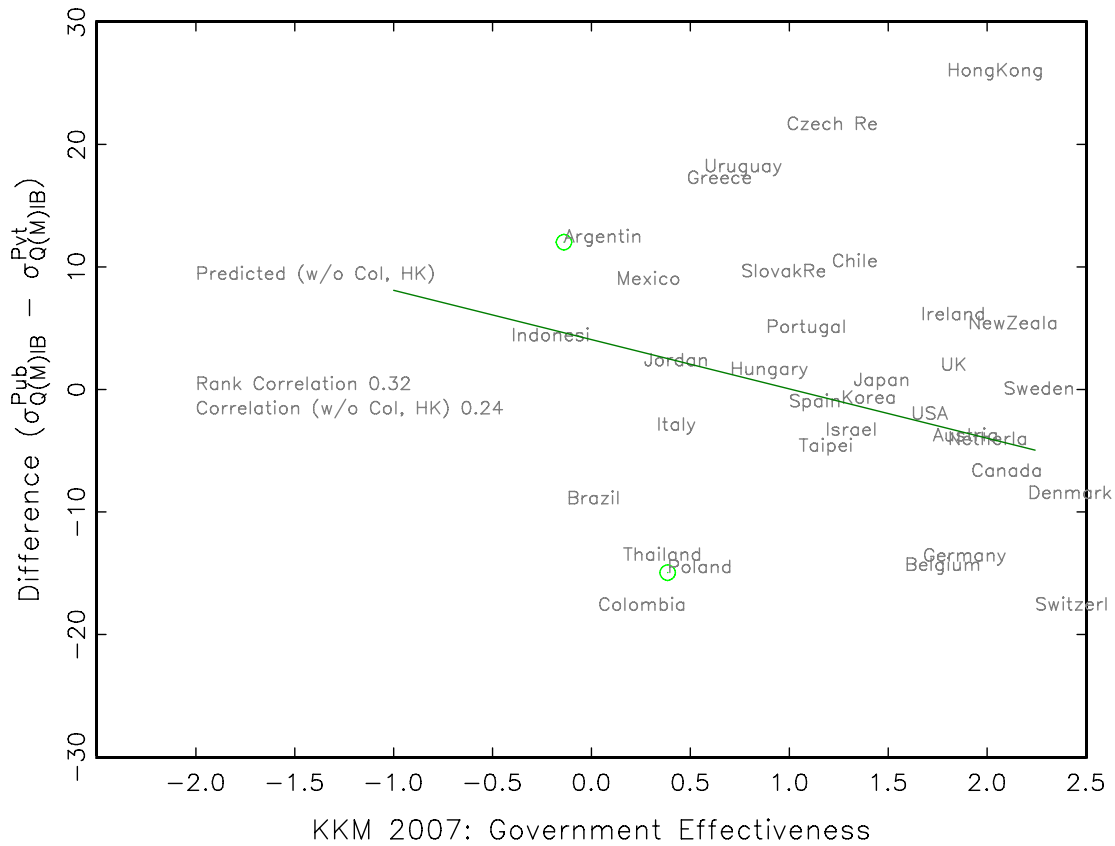
Kaufmann, Kraay, and Mastruzzi (2005) (KKM) create measures of six aspects of governance by combining data drawn from 25 data sources constructed by 18 different organizations⁴⁰. One aspect of governance is “Government Effectiveness” which includes the following variables: (i) quality of public service provision, (ii) the quality of the bureaucracy, (iii) the competence of civil servants, (iv) independence of the civil service from political pressures, and (v) the credibility of the government’s commitment to policies. This KKM “Government Effectiveness” index is a plausible proxy for “weak”

⁴⁰ In spite of many critiques (Oman (2006), Thomas (2007)) they serve our limited purpose of illustration.

versus “strong” state in implementation capacity. Figure 7 plots this variable the difference in variability in adjusted school quality ($\sigma_{Q(M)|B}^{Pub} - \sigma_{Q(M)|B}^{Pvt}$). The first thing to notice is that, not surprisingly as PISA participation itself is a complex and demanding task, not many truly weak states that have participated in PISA. Indonesia has the lowest “government effectiveness” in our sample but is in the 42nd percentile among all countries in the KKM rankings, Mexico is in the 60th percentile, so the “weak” states in this graph are quite strong compared to say, Pakistan (at the 28th percentile).

There is some modest positive correlation between KKM’s measure of “government effectiveness” and the extent to which the private sector’s variability is lower than the public sectors. The rank correlation between the government effectiveness and the difference in variability is .32 (with Hong Kong and Colombia as obvious outliers). But the correlation is modest and countries at both low and high levels of government effectiveness exhibit a similar range in the difference of variability in school quality (e.g., Colombia and Argentina vs. Switzerland and Hong Kong).

Figure 7: Difference in Variability in School Quality between the public and private sectors and “Government Effectiveness”



III.E) Variability in school quality, raw and adjusted

To this point in the paper we have been interested in organizational and systemic question of whether direct hierarchical control of larger numbers of individual schools and teachers through the public sector leads to more uniformity in quality across schools than a completely localized system in which each school (or small groups of schools) makes their own decisions, with parents and students choosing which school to attend. It could be fairly argued that this misses the important point about equality in schooling which is not (just) that schools provide uniform quality but also that there is equal access to schools so that the quality of school a child attends is not determined by their parental economic or social status. Crudely put, it could be the case that the variation in school

quality in the public sector is higher so that there are good and bad schools but that rich and poor children have equal chances of entering the good or bad schools, whereas the private sector has uniformly good schools but the poor children have no access.

Addressing this topic requires (and deserves) a paper in its own right. Of course, the first and most important point is that it is more or less irrelevant to the fundamental design system alternatives, like providing support directly to students or localizing school systems. When considering alternatives the question is not comparing economic/social stratification in an existing public system versus an unsupported localized system. Rather the real question is what would be the distribution of school quality and access to school quality by economic and social characteristics in alternative systems with different instruments of support (Pritchett 2009).

What we can easily do is report the unadjusted and adjusted measures of variability. Imagine there were two schools each in the public and private sector which differed in intrinsic quality by the same amount so that underlying variability in school quality is equal. Suppose that in the public sector students were assigned to one of the two schools irrespective of their socio-economic status or ability. In the private sector everyone wanted to be in the good school, but access to that school was rationed by setting a high price such that enrollment was determined by willingness to pay and that was correlated with student background characteristics associated with high learning outcomes. Then in this simple example the *unadjusted* difference in learning achievement across the two schools in the private sector will be much higher than the adjusted variance as parent/student selectivity will exacerbate the underlying differences in quality. In this case the unadjusted variability in the public sector will be lower than in

the private sector while the adjusted variability will be the same (if the adjustment recovers the true underlying variability which is assumed in this example to be equal)⁴¹.

Table 3 reports the variability in unadjusted and adjusted average school performance⁴². The striking fact is that the *reduction* in variability from making the adjustment for student background characteristics is, on average across these countries, almost exactly the same in the public and private sectors. That is, the unadjusted average standard deviation in the public sectors is 36.4, which the adjustment for background characteristics reduces by 13.9 points to 22.54. In the private sectors the unadjusted average standard deviation is 36.5 and the adjustment reduces this by 13.97 points to 22.55. We are not sure exactly what to make of this striking coincidence. It does imply the findings above that, on average, the variability in learning achievement across schools is the same in the public and private sectors does not depend on the adjustment procedure used, as it holds in raw unadjusted measures as well.

⁴¹ Of course this simple example generalizes to much more general formulations, basically if students sort into schools (by whatever process) such that underlying school quality is correlated with student characteristics positively associated with learning then rankings on unadjusted measures of average student learning performance will not reveal true school quality (and will make schools with good students look like good schools). This of course has generated the massive literature on how to estimate “value added” rankings of schools that adjust for student background.

⁴² In the *adjusted* average school performance we regress students’ achievement on individual characteristics (i.e., age, sex), family background characteristics (i.e., parents’ education, occupation status, and socio-economic status, number of books at home) and the school average of these background characteristics. Then we regress the residual extracted from the regression and we regress it on the school fixed effect and the error term. This allows us to examine the average school performance which eliminates the differences in average performance due to socio-economic status. On the other hand, the *unadjusted* average school performance does not control for background characteristics and provides the raw information. We are going to use these two measures in our analysis.

Country	Public			Private			Private Share	Difference (II-IV) (col I of Table 2)
	I $\sigma_{Q(M)}^{Pub}$	II $\sigma_{Q(M) B}^{Pub}$	Difference (I-II)	III $\sigma_{Q(M)}^{Pvt}$	IV $\sigma_{Q(M) B}^{Pvt}$	Difference (III-IV)		
Hong Kong	56.92	44.72	-12.19	36.62	26.70	-9.92	91.99	18.02
Czech Rep.	54.80	31.44	-23.36	24.76	13.43	-11.34	5.03	18.01
Uruguay	40.13	27.04	-13.09	20.69	12.07	-8.61	21.23	14.96
Greece	35.78	22.29	-13.49	7.67	7.51	-0.16	6.49	14.78
Argentina	47.39	32.16	-15.23	34.03	18.15	-15.89	31.25	14.02
Chile	44.91	26.68	-18.23	37.53	12.74	-24.79	57.73	13.94
Slovak Rep.	46.31	25.66	-20.65	36.67	16.32	-20.36	6.91	9.34
Mexico	43.40	30.08	-13.32	33.97	21.17	-12.80	12.09	8.91
Ireland	18.43	12.78	-5.65	13.12	5.70	-7.42	61.88	7.08
New Zealand	15.77	8.61	-7.15	10.83	3.58	-7.25	5.54	5.04
Portugal	31.58	19.64	-11.94	30.64	15.13	-15.50	11.61	4.50
Indonesia	48.89	33.88	-15.01	37.31	29.69	-7.62	32.67	4.20
Jordan	19.99	13.62	-6.37	17.81	9.90	-7.91	14.08	3.73
UK	22.58	12.75	-9.83	15.77	9.42	-6.35	4.98	3.33
Hungary	55.42	26.18	-29.24	50.23	23.74	-26.49	15.62	2.44
Japan	49.06	32.55	-16.51	48.51	31.12	-17.39	28.76	1.43
Sweden	15.95	10.86	-5.10	10.99	9.73	-1.26	7.78	1.12
Korea	37.08	22.00	-15.08	40.39	21.61	-18.77	47.28	0.38
Spain	13.66	9.50	-4.17	18.29	9.81	-8.48	43.55	-0.31
USA	24.71	13.64	-11.07	23.25	14.85	-8.40	8.92	-1.21
Italy	50.50	36.08	-14.42	58.11	37.64	-20.47	6.35	-1.56
Israel	41.23	29.84	-11.38	47.58	31.76	-15.82	28.54	-1.92
Austria	52.28	30.64	-21.64	64.56	34.61	-29.95	8.69	-3.96
Netherlands	58.75	33.25	-25.49	57.26	37.91	-19.34	68.04	-4.66
Taipei	46.65	24.70	-21.95	55.88	29.66	-26.22	40.35	-4.96
Canada	17.24	13.42	-3.82	20.90	19.11	-1.78	7.21	-5.69
Denmark	11.27	7.83	-3.43	21.13	16.42	-4.71	23.61	-8.58
Brazil	36.64	24.84	-11.80	48.92	34.04	-14.88	15.87	-9.20
Thailand	28.78	13.31	-15.46	38.91	23.34	-15.58	12.63	-10.02
Germany	60.26	28.76	-31.49	70.20	40.77	-29.43	4.99	-12.01
Belgium	39.93	22.79	-17.13	61.34	39.61	-21.73	68.22	-16.82
Poland	13.91	9.37	-4.54	39.33	27.15	-12.18	6.74	-17.78
Switzerland	39.83	24.21	-15.62	52.05	45.43	-6.61	3.76	-21.22
Colombia	18.33	11.20	-7.13	56.47	36.79	-19.68	18.73	-25.59
Average	36.42	22.54	-13.88	36.52	22.55	-13.97	24.39	-0.01
Median	39.88	24.46	-13.40	36.99	21.39	-13.84	14.85	0.75
Std. Dev.	15.23	9.57	7.21	17.39	11.59	8.08	22.82	11.02

III.F) Variability in school quality and variability in measured inputs

We do not want to get into estimating full blown relationships between school specific measures of quality and inputs, but it is at least worth asking the question whether differences in inputs across schools are even potentially capable of explaining either the public-private differences or the differences in the variability in the public sector across countries. Just as a crude first cut on such an analysis we sorted countries by adjusted school quality variability in the public sector ($\sigma_{Q(M)B}^{Pub}$) and in table 4 we compare the 1/3 of countries with the least variability to the 1/3 of countries with the highest variability. There are indeed massive differences in the public sectors across countries, as the 11 lowest and 11 highest countries differ by 21 points.

Table 4: Differences in school quality variability and in measured hard inputs ⁴³			
	11 countries with lowest $\sigma_{Q(M)B}^{Pub}$	11 countries with highest $\sigma_{Q(M)B}^{Pub}$	Difference, low variability countries less high variability countries
Median Variability in School Quality $\sigma_{Q(M)B}^J$			
Public	11.2	32.16	-20.96
Private	9.9	31.12	-21.22
Difference	-0.31	1.43	
Variability across schools in index of inputs			
Public	0.89	0.97	-0.08
Private	0.99	1.02	-0.03
Difference	-0.1	-0.05	
Variability across schools in class size			
Public	2.3	5.02	-2.72
Private	3.71	4.34	-0.63
Difference	-1.41	0.68	

⁴³ We use the following hard input variables: index of inputs (i.e., the index of quality of educational resources extracted from the school principal questionnaire of PISA 2006. This index includes seven items based on the school principal's perceptions of potential factors hindering instruction at school (PISA 2006 Technical Report, p. 340). The following items are included: shortage or inadequacy of science laboratory equipment, shortage or inadequacy of instructional materials, shortage or inadequacy of computers for instruction, lack or inadequacy of internet connectivity, shortage or inadequacy of computer software for instruction, shortage or inadequacy of library materials, shortage or inadequacy of audio-visual resources. All items were inverted for IRT scaling and positive scores indicate better quality of educational resources) and class size (i.e., the student/teacher ratio from the school principal questionnaire of PISA 2006).

However, differences in measured “hard” EMIS-visible inputs are unlikely to explain these differences. For instance, while it is the case that the measured variability in an input index is lower in the low variability than the high variability countries (.89 versus .97) this difference is small relative to the difference in scores. Moreover, the private sector in low variability countries has *higher* variability than the public sector (.99 versus .89) but *lower* variability ($\sigma_{Q(M)B}^{Pvt}=9.9$ versus $\sigma_{Q(M)B}^{Pub}=11.2$). Also, the private sector in the low variability countries has *higher* index of input variability than the public sector in high variability countries (.99 versus .97), but measured variability that is less than a third as large (9.9 versus 32.2). Another feature of table 4 is that the input variability is only slightly larger in the private sector in high variability (in the public sector) countries than in the low variability countries (1.02 versus .99), but the adjusted score variability is more than three times larger (31.1 versus 9.9).

These same comparisons are true for another hard measure, class size. In the low variability countries the class size variability is *larger* across schools in the private than public sector even though adjusted score variability is lower. So, while it is true that larger variability in adjusted scores in the public sector ($\sigma_{Q(M)B}^{Pub}$) countries do appear to have larger variability in class size, it seems unlikely this alone, without other systemic differences, could explain the massive differences in variability.

These comparisons for Mexico mostly clearly illustrate a pattern true in some (but by no means all) other weak implementation countries (Appendix 7). In Mexico the variability in hard inputs across schools is much *smaller* in the public than private sector (.92, roughly the level in the low variability countries) versus 1.17 in the private sector. The variability in class size across schools is also *lower* in the public sector than the

private sector as well (10.9 in public versus 12.2 in public). And yet $\sigma_{Q(M)B}^{Pub}$ is 30.1 versus a $\sigma_{Q(M)B}^{Pvt}=21.2$. So in these cases greater equality in the EMIS visible hard inputs is side by side with much larger variability in measured adjusted learning performance outcomes. The distributions in figure 4a above suggest the answer: Mexico has many schools with roughly equal EMIS-visible inputs and class sizes as other public schools but which are, even controlling for the socio-economic background of their students, really weak schools.

Conclusion

*Yes, my guard stood hard when abstract threats
Too noble to neglect
Deceived me into thinking
I had something to protect
Good and bad, I define these terms
Quite clear, no doubt, somehow.
Ah, but I was so much older then,
I'm younger than that now.*

Bob Dylan, My Back Pages

That a low caste child from a poor family in India (or anywhere else) be able to fully utilize their abilities and effort to enhance their life-chances is essential. Inequality in the opportunities facing children to develop and utilize their talents through their access to quality schooling is indeed a central problem facing all societies, and should be a priority on social, political, and economic (Pritchett and Viarengo 2008a) grounds. This goal however is not furthered by rhetoric that only serves to sustain illusions about the ability of “bureaucratic high modern” schooling systems to deliver on this promise in weak implementation environments. Nor the goal served by external assistance to poor

countries that re-enforces these tendencies to transplant the forms of modern bureaucracies without their functions and sustain illusion of equality by promoting a sub-culture of “institutional isomorphic mimicry” (Andrews, Pritchett and Woolcock 2009).

We are not claiming that more localized systems that allow for much greater autonomy and innovation at the school level in order to break organizational monopolies on the utilization of public funds and in order to create the space for innovation are a panacea. But we are claiming that merely *assuming* that centralized systems deliver on the promise of equality merely because they exercise hierarchical control over the few items they “make EMIS legible” suffers all the defects of driving while blind. A state which is blinded to the complex social realities of the learning process, of day to day interactions of teachers with students, in favor of managing tenuously related inputs, particularly when the state is weak, is the real risk to children, not the “risk” of creating space for real innovation.

If we are correct that the equality maintained by centralized systems is an illusion, this does raise two questions, “if equality is only a *rationale* for centralized systems, what is the real reason?” and related “if localized systems would be superior in providing *both* higher levels of learning and do so with *less* inequality, why have they been so rarely adopted?” As this paper is part of a much larger research agenda, we have answers to both these questions, but the answer cannot fit in a fortune cookie so you will have to read more (Pritchett 2003, Pritchett and Viarengo 2008b, Pritchett 2009).

References

- Aghion P. and J. Tirole (1997), "Formal and Real Authority in Organizations", *Journal of Political Economy*, Vol. 105, No. 1, pp.1-29.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers (2006) "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 20(1), pp 91-116.
- Das J., P. Pandey, and T. Zajonc (2006), "Learning Levels and Gaps in Pakistan", *World Bank Research Working Paper No. 4607*.
- Das J. and T. Zajonc (2008), "India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement", *World Bank Policy Research Working Paper 4644*.
- Desai S., A. Dubey, R. Vanneman, and R. Banerji. (2008) "[Private Schooling in India: A New Educational Landscape](#)." India Policy Forum, Volume 5. Sage Publications. Eds. Suman Berty, Barry Bosworth, and Arvind Panagariya
- Duflo E., P. Dupas, and M. Kremer (2007), "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya", NBER Working Paper No. 14475
- Filmer, D., and L. Pritchett, (1997) "Child Mortality and Public Spending on Health: How Much Does Money Matter?" Policy Research Working Paper No. 1864, World Bank, Washington, DC.
- Goyal S. (2007a), "Learning Achievements in India: A Study of Primary Education in Orissa", *Document of The World Bank*
- Goyal S. (2007b), "Learning Achievements in India: A Study of Primary Education in Rajasthan", *Document of The World Bank*
- Hanushek (2003). "The failure of input-based schooling policies." *Economic Journal*, 113(485) (February): F64-F98.
- Hanushek E.A. (2006), "School Resources" In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 865-908.
- Jensen, M. C. and W. H. Meckling (1992) "Specific and General Knowledge and Organizational Structure," in W. Lars and H. Wijkander (eds.), *Contract Economics*. Oxford, Basil Blackwell: 251-274.
- Koretz D. (2008), "Measuring Up: What Educational Testing Really Tells Us". Cambridge, MA: Harvard University Press.

LEAPS – Andrabi T., D. Jishnu, A. I. Khwaja, T. Vishwanath and T. Zajonc (2007), “Learning and Educational Achievement in Punjab Schools (LEAPS)”

Murgai R. and L. Pritchett (2007) “Teacher Compensation: Can Decentralization to Local Bodies Take India From Perfect Storm Through Troubled Waters to Clear Sailing?” *India Policy Forum* 2006/07.

Pande V. and L. Pritchett (2007) “Making Primary Education Work for India’s Rural Poor: A Proposal for Effective Decentralization.” (with Varad Pande). *Social Development Working Papers #6: South Asia Series*.

Pritchett L. (2004), “Towards a New Consensus for Addressing the Global Challenge of the Lack of Education,” in Bjorn Lomborg, editor. *Global Crises, Global Solutions*. Cambridge University Press.

Pritchett (2009), “The Place Premium: Wage Differences for Identical Workers across the US Border”, *Harvard Kennedy School Working Paper No.09-004*

Pritchett L. and M. Viarengo (2008a), “Producing Superstars for the Economic Mundial: The Mexican Predicament with Quality of Education” *forthcoming as a paper in the “Mexico Competitiveness Report 2009” of the World Economic Forum and Harvard’s PEPG Discussion Paper 09-01*

Pritchett L. and M. Viarengo (2008a), “The State, Socialization and Private Schooling: When will Governments support Alternative Producers?” *Unpublished Manuscript*

Scott J. C. (1998), “Seeing like a State. How Certain Schemes to Improve the Human Condition Have Failed”, New Haven and London: Yale University Press.

Vegas E. (1999), “School Choice, Student Performance, and Teacher and School Characteristics: The Chilean Case”, *World Bank Policy Research Working Paper No. 2833*

Woessmann, L. (2007), "International Evidence on School Competition, Autonomy and Accountability: A Review", *Peabody Journal of Education* 82 (2-3), pp. 473-497

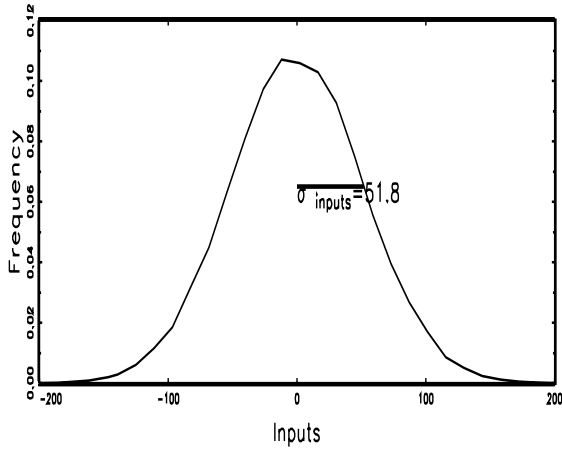
Woessmann L. (2008), "Public-Private Partnerships and Schooling Outcomes across Countries", in: R. Chakrabarti, P.E. Peterson (eds.): *School Choice International*, MIT Press, Cambridge, MA.

Appendix 1: Simulating Reductions in Variance Across Schools from Input Equalization

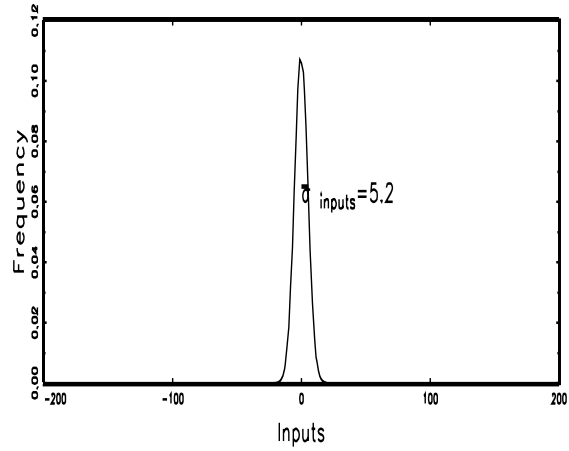
Figures A1.1 and A1.2 illustrates that the equalization in school quality outcomes possible through equalization of inputs depends on the empirical strength of the relationship between inputs and outcomes. Using simple simulations of scenarios that are identical other than in the assumed connection inputs and school equality, we show that for the strengths of visible input/outcome relationships widely observed in empirical studies the scope for equalization through centralized reallocation is very small.

Figure A1.2 shows the implication of equalizing inputs with a tight relationship between the controllable inputs (X) and school quality. All goes according to plan. A radial reduction in inequality of inputs across schools by 90 percent (from a standard deviation of 52 (Figure A1.1a) to a standard deviation of 5.2 (Figure A1.1.b)) produces substantial reductions in variance in school quality (from 54.7 to 18.1). This is because the relationship between inputs and outcomes is assumed to be tight (Figure A1.1c, an R-squared of .9) so collapsing the inequality in the inputs dimension leads to lower inequality in outcomes (see figure A1.1d). Another way of illustrating the strength of the input/outcome relationship is that when the relationship between inputs and outcomes is strong, then low input schools have a very different distribution of outcomes than high input schools (see Figure A1.1e) and measured overall inequality in outcomes is in large measure driven by the differences between students in low input and high input schools. The equalization of inputs leads these two distributions to converge and inequality is reduced (figure A1.1f).

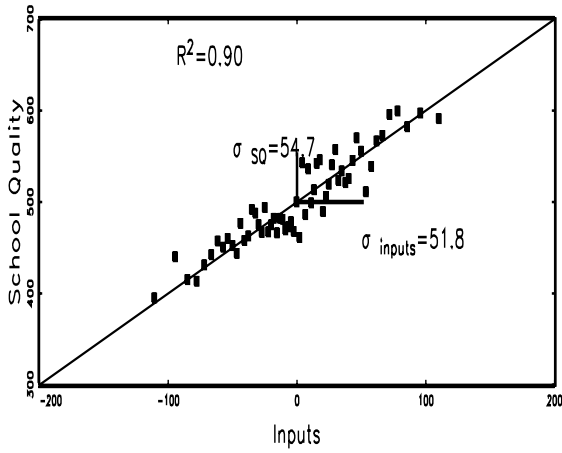
1a: Initial unequal distribution of inputs



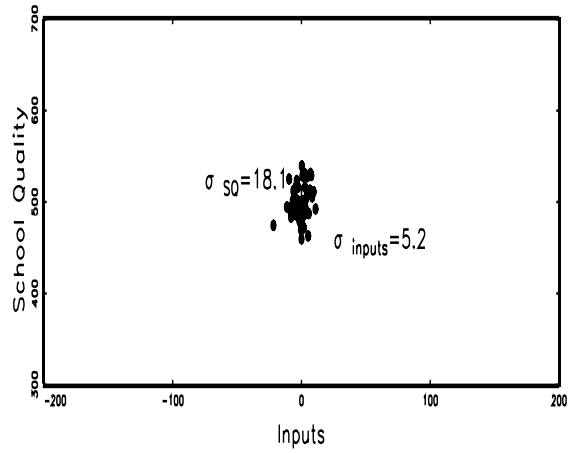
1b: A simulated radial reduction in inequality of inputs by 90 percent



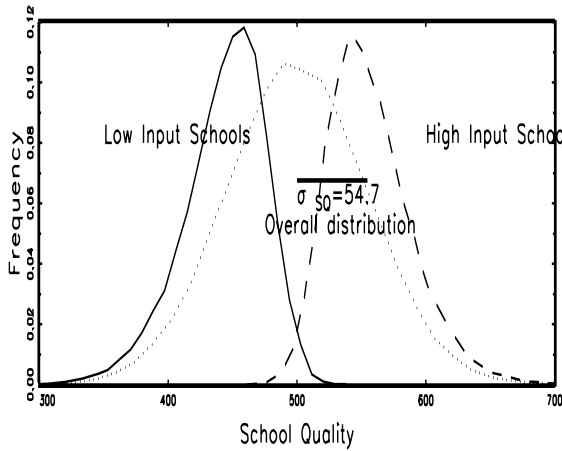
1c: An assumed strong relationship



1d: Equalizing controlled inputs does lead to more equal school quality



1e: Low input schools are low quality



1f: Equalising inputs reduces variability in school quality

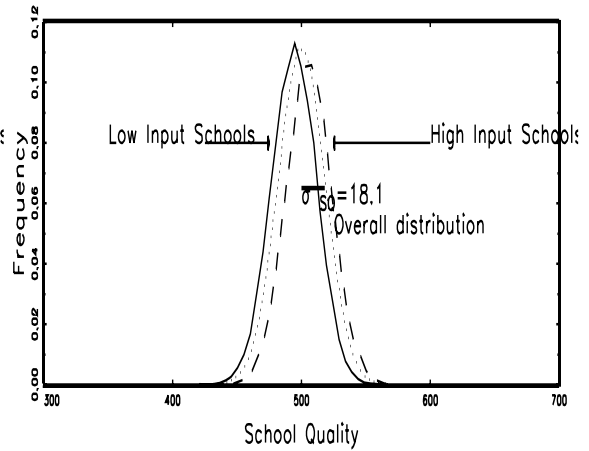
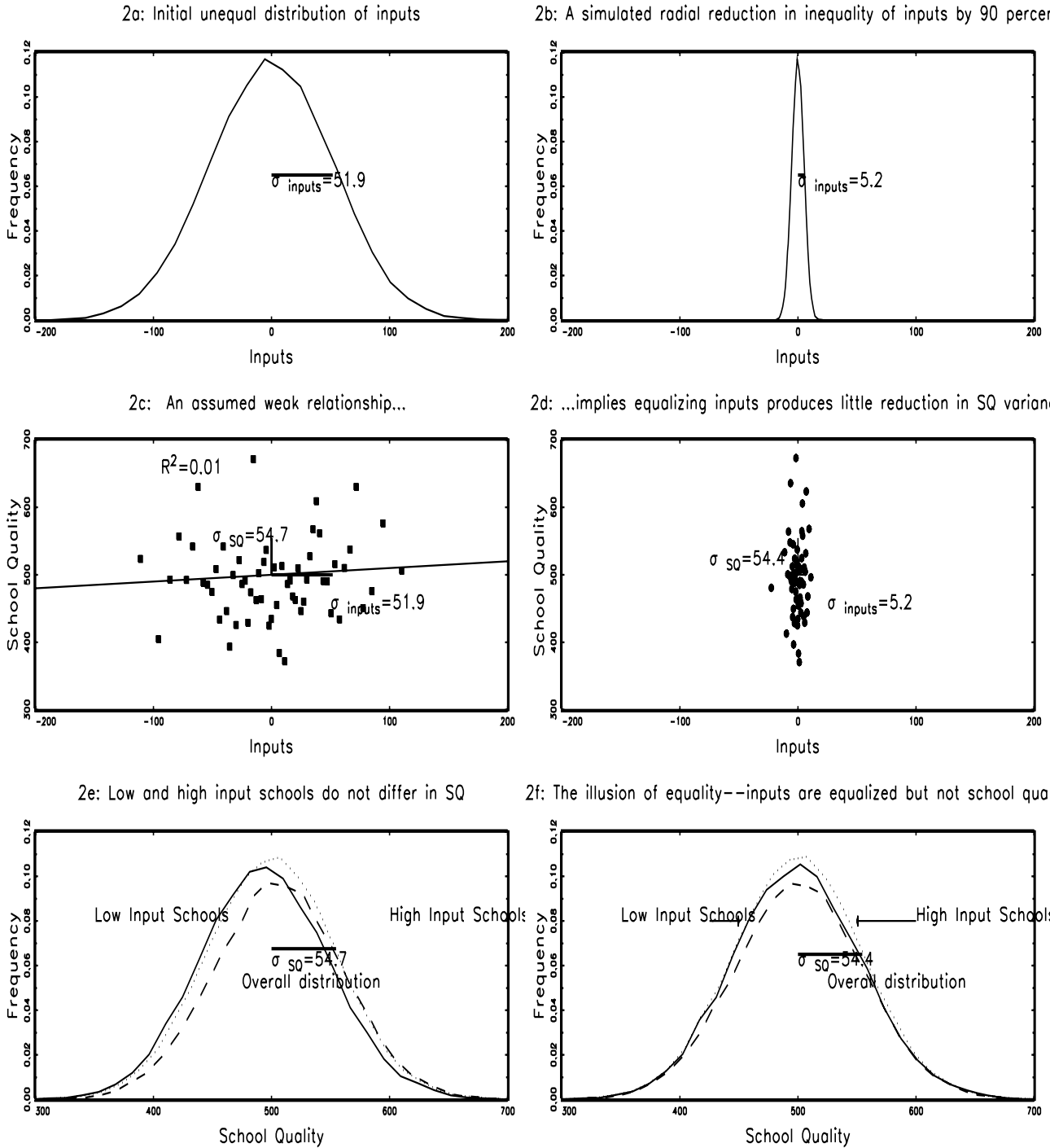


Figure A1.2 illustrates a potential illusion of equality. We start from exactly the same inequality in the distribution of an input (Figure A1.2a) and simulate exactly the same massive radial reduction in inputs across schools (Figure A1.2b). But this scenario assumes there is only a weak empirical relationship between the input and school quality (an R-squared of .01, figure A1.2c). In this case the reduction in the variability of inputs has some, but very little impact in reducing school quality variability. The simulated standard deviation across schools falls from 54.7 to only 54.4 (Figure A1.2d) as the reduction in variance in the inputs dimension does not reduce the variance in outcomes by much given the weak assumed relationship between inputs and outcomes. Figure A1.2e shows why (in a proximate sense)—low input and high input schools did not differ that much in school quality to begin with, so, unlike figure 1e in which there was a large gap in the distribution of school quality of low input and high input schools, in this case, although they differ, it is not by much do reducing the gap between the inputs of the schools does very little to change the overall distribution (Figure A1.2f).

This illustrates one potential illusion of equality. To an educational system tracking only inputs it would appear that schools *are* equal. School quality is at times *defined* by input standards and/or process compliance. This has a compelling internal logic in hierarchical bureaucratic systems as this is all the system can see and control. When there are no independent measures of outcomes the illusion of equality can be officially asserted without contradiction even when the most brutal inequalities in the school quality actually experienced by students or as measured by any reasonable construal of school quality in terms of contribution to educational goals (nothing said so far implies privileging any educational goal or measure over any other) may be massive.

Figure A1.2: The illusion of equality from controlling inputs with only a weak relationship to outcomes...



Appendix 2: How important are “hard” characteristics of schools

We illustrate generic points about the “educational production function” literature with one study of the relationship between observed inputs and learning achievement, a study chosen because it is typical of hundreds of such studies only atypical in being of much *higher* quality than most. This empirical study used the measured learning achievement of grade 4 students in Sri Lanka in the subjects of Mathematics, English, and the child’s first language. In doing so all that is asserted is that these learning objectives are at the least *some* of the legitimate goals of the Sri Lankan educational system, not that these are the only goals, or even the most important goals. This study had measures of not only of student performance but also had very detailed information about the children’s own background (e.g. parent’s education, resources in the home, participation in private tutoring) and very detailed information about the school (more below). The study then used standard econometric techniques to estimate the association between the three learning achievement measures and child, household, and school level variables. There are three findings of interest from this study (as we discuss below, this is a concrete instance of findings that reviews of this literature show are typical in many such studies).

Before moving to the investigation of school impacts, it is worth noting that the magnitude of the association with participating in private tutoring was enormous. A student who spent some time in tutoring versus the student who did not had scores, on average across the three measures, 37.2 points higher⁴⁴. While we are not asserting this is causal nor that encouraging tutoring is good public policy, this empirical result does show that the numbers are likely really measuring learning achievement and also are at least suggestive of the magnitude of possible learning gains from effective teaching.

First, this study has a rich array of variables intended to measure school quality, including infrastructure variables, teacher training, characteristics of teachers, principals, all told 35 variables *plus* variables based on actual observation of teachers in the classroom. Basically, anything one could imagine a typical hierarchical school administration or EMIS could “see” to manage was measured--and more. This examination of the connection with learning of school characteristics results in one telling sentence “variables with no explanatory power in any of the tests are dropped”--meaning if a measured characteristic was *never* statistically significant in *any* specification for *any* learning measure it was not reported in the paper. This filtering of the “ever significant” variables left only seven! (forgive us the exclamation point) variables. So nearly all of what is commonly measured as “school quality” simply failed to have any demonstrable association with learning.

Second, of the seven variables that were included in the final specification (because they were statistically significant in at least one specification for at least one learning outcome) the magnitude of the estimated associations was empirically tiny. For instance, it is commonly asserted that more experienced teachers are superior and that

⁴⁴ In the study they normalize the instrument of have a standard deviation of 1, we are just renormalizing the results for a standard deviation of 100 to be consistent with the scenarios above.

retention of teachers a policy priority and this justifies paying more to teachers with more seniority. The estimates were that an additional year of teacher experience added .36 points to their students learning achievement (on a scale of a student standard deviation of 100, so the “effect size” it is .0036, and is a *hundredth* of the 37.2 “effect size” for two hours a week with a tutor). This implies the typical teacher with 10 years experience has students with scores only 3.6 points higher than a rookie.

Third, the largest impact of any measured school characteristics was a measure of “having desks.” Bringing the others up to that standard would (if the effect were causal) add 8.1 points to student scores. However, this illustrates another common point about the limited scope for equalization of even those characteristics with measurable and substantial associations—which is that they are discrete, either “adequate” or “not adequate” which means the most that can be done is equalize the two groups. But, this is a tiny fraction of the total inequality across schools. So in this example, 60 percent of schools were already adequate so bringing other schools up to this standard would do nothing about the variance among those schools, 40 percent did not have adequate desks and, while bringing those schools up to standard (again on the premise these are causal) would reduce the gap in averages it would not reduce the existing variances among those schools. So the maximal achievable equalization through eliminating even the largest impact input gaps is tiny.

Questions about pedagogical processes are massively complex as there are human beings on both sides of any extended learning experience. Since they are so complex we do not assert this simple example generalizes. For the same reasons, that school learning is such a complex and localized experience it would not be surprising that the overwhelming bulk of the published literature is similar in finding mixed, small, and limited gains to most EMIS “visible” school inputs. In those instances where this is true, the input equalization achievable through centralized control will produce only the illusion of equality.

Appendix 3: Simulation of “qualifications hiring” versus “observation”

We use a second set of simulations to illustrate the consequences for equality in school quality of two different processes for selecting and retaining teachers. Suppose teacher ability can be summarized in a single number, s , and that is distributed randomly (a uniform distribution) among the population. Suppose that a person’s ability as a teacher can only be observed by allowing them to teach. Suppose also that pre-service training can augment teacher ability. Imagine two difference processes.

- *Observation:* Teachers are hired from the general population (without training) and are given one probationary period. Following this probationary period if their observed ability is above a threshold they are retained (tenured) and they teach $N-1$ more periods until retirement.
- *Qualifications:* Teachers are hired from the pool of those with the appropriate qualifications, which augments teacher ability by amount b (and potentially also reduces the variance) and all teachers hired are retained for N periods until retirement.

Which process, ‘retention on observables’ or ‘hiring on qualifications’ will produce the most equality across schools? A simple simulation in can be constructed so that the answer depends on the ratio of the amount by which training augments ability (b) compared to the observable variability of ability among the untrained (which is a combination of the underlying variability and the accuracy with which it is observed).

The “observe and retain” strategy, which draws new hires from a pool of teachers with ability distributed evenly between 0 and 1. The optimal strategy for a school seeking to maximize teacher ability is to retain a teacher if their observed ability is above a cutoff c , where the optimal cutoff is higher when teachers are retained for more periods $c^*(N)$ ⁴⁵. The expected value of teacher ability for any school at any time t with optimal cutoff c given the period N is:

$$E(a(c(N))) = (\text{Fraction new teachers}(c(N)) * (\mu) + (\text{Fraction retained}(c(N)) * ((1 + c(N)) / 2))$$

Suppose there is observation period and then all retained teachers teach for 8 more periods for a total career of 9 periods (one can think of this as a “period” of three years, so the observation period is three years, a total career 27 years). The optimal cutoff with 9 periods is .75 (teachers are retained if their observed ability is in the top quarter) with the result that a third of all teachers at any given point are new and 2/3 are retained (distributed evenly over the remaining periods) which is shown in Figure A2.3c.. The resulting expected value of teacher ability is: $(1/3) * (.5) + (2/3) * (.875) = .75$. Figure A2.3a provides a simple graphical illustration. The standard deviation of teacher ability (across

⁴⁵ For a uniform distribution the formula is $c^*(N) = \frac{\sqrt{N}}{(\sqrt{N} + 1)}$, so that for instance if $N=9$, $c=.75$ while if $N=25$ $c=.8$. Thanks to Andrew Dittmer for these closed form formula.

all schools) is smaller than the distribution of the $U(0,1)$ distribution of $.289=1/\sqrt{12}$ because of the selection and is around .21.

The outcome of the “hire qualified” process depends on the impacts of the training on teacher ability. If it raises the post-training ability of all teachers by a common amount then the uniform distribution is just shifted up by amount b and the new average is $.5+b$ while the standard deviation remains unchanged, which is illustrated in Figure A2.3b. However, one can also imagine that training augments those with natural ability by more than those with less natural ability to that training *increases* the variability of ability of teachers. Alternatively, it may be the case that training both increases the mean and lowers the variability of ability among teachers, which is simulated as a radial reduction in the uniform distribution. Since all hired teachers are retained, there are equal numbers of teachers at all levels of seniority (Figure A2.3c).

The simulation creates individual “schools” each having six teachers, each of which follow either the “observe and retain” or the “hire qualified” strategy and that each teacher teaches 30 students and that student outcomes are the simple sum of their teacher’s ability and another random term which represents the student specific elements. This allows the calculation of the mean and variability of school quality across the two processes. We’ll proxy the relative variability by the coefficient of variation, which is the standard deviation across schools divided by the mean. There are three scenarios illustrated in Figure A2.3d, which shows the relative coefficient of variation of “hire qualified” to “observe and retain.”

The first assumes that observation is perfect and that teacher qualifications only increase the mean. The “qualifications” parameter is scaled in terms of “effect sizes” (as this is how nearly all empirical studies are scaled) which is the number divided by the a student standard deviation, which in this simulation is .4, so in the graph an effect size of 1 implies that $b=.4$. The top line shows that, under these assumptions, qualifications would have to have an implausibly enormous impact on ability to make up for the inability to retain based on observed quality.

A second possibility is that teacher ability is difficult to observe, even after observing a teacher for a number of years. We simulate this by adding “noise” to the observed teacher ability on which schools choose to retain teachers (we add variance of .4, a student standard deviation [check this]). This both reduces the mean and increases the variance of the “observe and retain” strategy so that the critical threshold of the ability increment for the c.v. of the “hire qualified” process is lower.

A third simulation assumes the pool of “qualified” teachers has a lower variance than the distribution of ability among the unqualified. Under the assumption that “qualifications” reduce the variance in teacher ability by half the “hire qualified” process produces lower variance than “observe and retain” even if there is no impact on the average ability.

Figure 3a: Observables: teachers retained if observed first period teaching quality is above cutoff

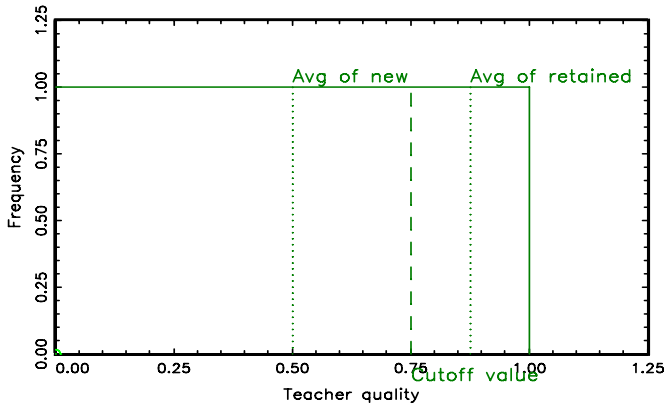


Figure 3b: Qualifications: Teachers only hired with qualifications, all hired are retained

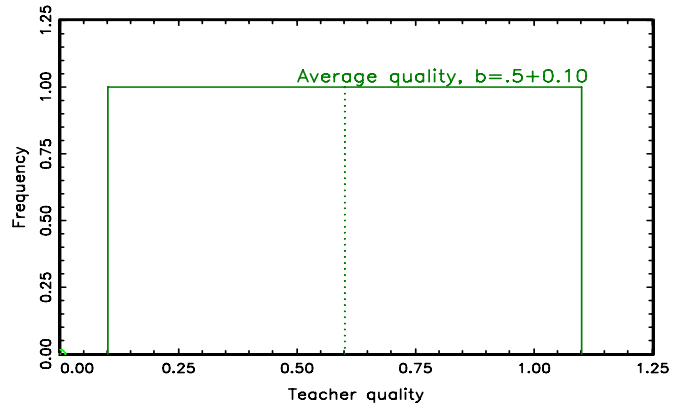


Figure 3c: Fraction of teachers by years of seniority observables versus qualifications process

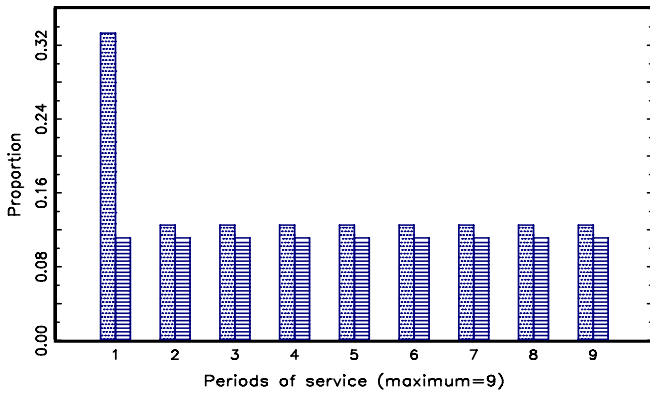
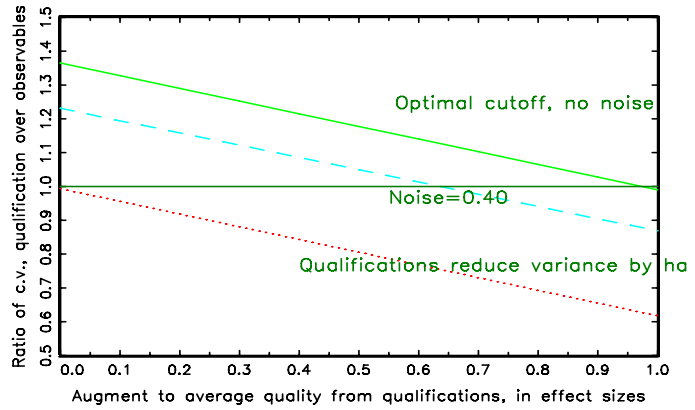


Figure 3d: Ratios of variability of qualifications versus observable by quality augment from qualifications



Appendix 4: Results using PISA Math, Reading or Science Scores				
Country	Math $\sigma_{Q(M) B}^{Pub} - \sigma_{Q(M) B}^{Pvt}$	Reading $\sigma_{Q(M) B}^{Pub} - \sigma_{Q(M) B}^{Pvt}$	Science $\sigma_{Q(M) B}^{Pub} - \sigma_{Q(M) B}^{Pvt}$	Share in private schools (%)
Hong Kong	18.02	7.76	10.59	91.99
Czech Republic	18.01	4.96	16.49	5.03
Uruguay	14.96	11.22	13.90	21.23
Greece	14.78	18.95	14.50	6.49
Argentina	14.02	23.71	18.45	31.25
Chile	13.94	10.79	10.88	57.73
Slovak Republic	9.34	5.74	5.99	6.91
Mexico	8.91	8.37	2.34	12.09
Ireland	7.08	4.86	6.12	61.88
New Zealand	5.04	6.78	4.61	5.54
Portugal	4.50	12.48	8.17	11.61
Indonesia	4.20	0.85	0.52	32.67
Jordan	3.73	-0.05	2.05	14.08
UK	3.33	4.68	2.57	4.98
Hungary	2.44	8.77	-0.01	15.62
Japan	1.43	10.20	5.31	28.76
Sweden	1.12	2.21	0.39	7.78
Korea	0.38	-0.44	-0.72	47.28
Spain	-0.31	-0.86	0.80	43.55
USA	-1.21	NA	-3.24	8.92
Italy	-1.56	7.26	-1.39	6.35
Israel	-1.92	-4.82	-6.42	28.54
Austria	-3.96	-10.29	-7.86	8.69
Netherlands	-4.66	-9.01	-0.47	68.04
Chinese Taipei	-4.96	-5.02	-6.34	40.35
Canada	-5.69	-1.15	1.53	7.21
Denmark	-8.58	-8.34	-4.41	23.61
Brazil	-9.20	-1.59	-5.16	15.87
Thailand	-10.02	-2.19	-9.60	12.63
Germany	-12.01	-2.37	-3.42	4.99
Belgium	-16.82	-6.48	-8.18	68.22
Poland	-17.78	-13.23	-21.60	6.74
Switzerland	-21.22	-12.72	-22.90	3.76
Colombia	-25.59	-30.12	-13.14	18.73
Average	-0.01	1.24	0.30	24.39
Median	0.75	0.85	0.46	14.85
Std. Dev.	11.02	10.30	9.52	22.82
Corr. w/ Math		0.84	0.92	

Appendix 5: Using TIMSS 2007 versus PISA 2006

Table A5.1: Comparison of the results of the difference in variance across schools in quality, public versus private schools, PISA 2006 Versus TIMSS 2007, Mathematics.					
	s.d.	Absolute standard deviation Across schools, Public Sector	Absolute standard deviation Across schools, Private Sector	Difference PUB-PRIV	Share of enrollment in the Private Sector
Colombia					
PISA2006	78.77	11.20	36.79	-25.59	18.73
TIMSS2007	67.24	13.06	23.96	-10.91	17.32
Hong Kong					
PISA2006	87.10	44.72	26.70	18.02	91.99
TIMSS2007	83.12	60.57	43.85	16.72	34.93
Indonesia					
PISA2006	73.83	33.88	29.69	4.20	32.67
TIMSS2007	79.58	29.14	27.47	1.66	34.93
Japan					
PISA2006	81.47	32.55	31.12	1.43	28.76
TIMSS2007	75.17	9.32	11.61	-2.30	9.38
Jordan					
PISA2006	73.05	13.62	9.90	3.73	14.08
TIMSS2007	90.72	22.15	24.25	-2.09	47.19
Thailand					
PISA2006	73.30	13.31	23.34	-10.02	12.63
TIMSS2007	81.67	32.46	27.93	4.53	8.44

Appendix Table A7.1: School Inputs and variability in students' performance across school sector (low and high Public Sector adjusted average mathematics test score)

		School Quality Input		Class Size		Residuals			
	Country	Sd_quality Input PUB	Sd_quality Input PRIV	Class_ Size PUB	Class_ Size PRIV	Std. Dev. of adjusted school average mathematics score-PUB	Std. Dev. of adjusted school average mathematics score-PRIV	Difference PUB-PRIV	Share of enrollment in private schools
LOW	Denmark	0.84	0.68	2.16	3.2	7.83	16.42	-8.58	23.61
	New Zealand	0.89	1.22	2.51	6.75	8.61	3.58	5.04	5.54
	Poland	0.87	1.01	2.15	2.66	9.37	27.15	-17.78	6.74
	Spain	1.00	0.99	1.73	3.71	9.50	9.81	-0.31	43.55
	Sweden	0.92	0.50	2.3	3.42	10.86	9.73	1.12	7.78
	Colombia	0.86	0.90	11.35	9.72	11.20	36.79	-25.59	18.73
	UK	1.01	1.42	2.19	2.01	12.75	9.42	3.33	4.98
	Ireland	0.87	1.01	2.14	2.04	12.78	5.70	7.08	61.88
	Thailand	0.89	0.91	5.38	5.65	13.31	23.34	-10.02	12.63
	Canada	0.97	0.96	2.66	5.67	13.42	19.11	-5.69	7.21
Jordan	0.84	1.20	5.05	7.71	13.62	9.90	3.73	14.08	
HIGH	Germany	0.99	0.95	5.02	3.33	28.76	40.77	-12.01	4.99
	Israel	1.22	1.15	5.98	4.34	29.84	31.76	-1.92	28.54
	Mexico	0.92	1.17	10.92	12.25	30.08	21.17	8.91	12.09
	Austria	1.09	0.98	6.96	3.09	30.64	34.61	-3.96	8.69
	Czech Republic	0.74	1.13	3.81	2.95	31.44	13.43	18.01	5.03
	Argentina	1.24	1.09	6.39	6.65	32.16	18.15	14.02	31.25
	Japan	0.96	1.02	4.11	5.61	32.55	31.12	1.43	28.76
	Netherlands	0.88	0.85	3.52	4.67	33.25	37.91	-4.66	68.04
	Indonesia	1.30	1.33	5.88	9.36	33.88	29.69	4.20	32.67
	Italy	0.97	0.78	2.73	2.54	36.08	37.64	-1.56	6.35
Hong Kong	0.88	0.92	2.25	2.14	44.72	26.70	18.02	91.99	