RISE

RESEARCH ON IMPROVING
SYSTEMS OF EDUCATION

Creating Education Systems Coherent for Learning Outcomes: Making the Transition from Schooling to Learning

Lant Pritchett

Creating Education Systems Coherent for Learning Outcomes: Making the Transition from Schooling to Learning

Lant Pritchett

Harvard Kennedy School and Center for Global Development

Abstract

Existing systems of education have some elements promoting learning as an objective, but are mainly coherent as systems only around enrollment targets. This paper builds an accountability framework of actors and the four design elements of accountability (delegation, financing, information and motivation) to emphasize that effectiveness in promoting learning requires systems of education that are coherent, in two ways. First, each accountability relationship has to be coherent across its elements, that is, the delegation of what agents are asked to do has to be coherent with the financing, information, and motivation, rather than "pay for one thing and expect another." Second, the relationships have to be coherent across relationships of accountability. That is, if teachers are accountable both to their employer and indirectly to parents/students/communities, then if these two have very different objectives the accountability of teachers will be made incoherent. Such incoherence can explain why small changes in the "right" direction (towards that of high performing systems or demonstrated in other contexts) might consistently fail even where a directed and coherent reform could have major impact.

Introduction and Objectives of this Review

International and national assessments around the world demonstrate that there is a global learning crisis of children already in school (UNESCO 2014). In response, an important transition from schooling to learning is in progress. While the Millennium Development Goal on education emphasized enrollment, the new Sustainable Development Goal (SDG 4.1) shifts focus: "By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes."

Yet current understanding of what leads to learning leaves key questions unanswered. This paper is part of the RISE series that explores three overlapping but distinct sets of questions about education, and in particular learning outcomes[1].

Questions about proximate determinants. The first set of questions concern the proximate determinants of learning outcomes for a particular child. That is, a child at age 15, 16 or 18, at or near the end of "basic education," will have an array of competencies: social, emotional, academic, athletic, psychic, musical, etc. This stock of accumulated competencies is the result of everything about the history of the child from her genetic endowment, to pre-natal conditions, to early childhood health and nutrition, to her household conditions, to influence from siblings and peers, and, as one element of that overall nexus, the experience of formal schooling. If we limit ourselves (for convenience, not priority) only to curricular competencies of basic schooling we can ask: *given* or *conditional* on everything else about the child, what are the learning environment exposures produced by schooling that lead to better or worse outcomes on these assessed competencies? It could be that the child was exposed to more total instructional time (e.g. more years in school, more hours per day of instruction); schools with better physical conditions (e.g. a rain proof roof, adequate toilets); schools with better physical inputs (e.g. adequate desks, or teachers with better formal qualifications); more in-service training (e.g. hours per year of training); teachers with better subject knowledge; teachers that applied pedagogical approaches better for the child; or classrooms with better classroom management.

There are literally thousands of academic papers that address questions about proximate determinants of the type "what is the causal learning gain in domain Z (as measured by assessment with instrument Y) for child J of exposure to instructional conditions X?"

One set of papers in the RISE series addresses these questions. Those papers have three takeaways:

> o Our review of the literature shows large differences in learning outcomes not easily explained merely by cross-national differences in socioeconomic status— e.g. students from equivalently poor households in Vietnam learn much more than

---

[1] Although the empirical measures used for studies of "learning outcomes" are often subject matter assessments of reading, mathematics, science, etc., we use the term "learning outcomes" in the broad sense to include all skills, competencies, dispositions, values, etc. that an education intends to convey. In discussing the relationship between system characteristics and say, mathematics or reading competence, we do not intend to suggest that math and reading are the only, or necessarily the priority, elements of education. For present purposes, they are just two of many curricular competencies; but nearly all countries intend their education system to produce reading and mathematics competencies and these are measured widely so we use them for convenience.

their Peruvian counterparts (Sandefur 2015) and there is strong evidence that this emerges during school-aged years (Singh 2014).

- o Our review of the literature shows that easily observable features of the inputs of the schooling system, such as expenditure per student, education of teachers, or average class size—while at times "statistically significant"—explain very little of the observed variation in learning outcomes at any level, and are almost always small relative both to differences across students and to differences across nations/regions (Glewwe and Muralidharan 2015). Such features are "thin" in the Geertzian sense articulated below (Pritchett 2014).

- o Our review of the literature suggests that, consistent with earlier findings (Crouch and Healey 1997), within nearly all systems, even those with on average poor performance, there are teachers and schools that perform very well. This implies the question is not "Why is good performance impossible in system setting C?" but "Why, in system setting C, is the good performance achievable and achieved by teachers (and, less so, schools) not diffusing to become the norm?"

System determinants of the proximate determinants and their contextual efficacy. The second set of questions has roughly two parts: "What are the determinants of the proximate determinants?" and "What determines the overall learning efficacy of the use of resources devoted to education?"

The first part of this question is relatively easy to pose. Suppose exposure to smaller class sizes were an important cumulative determinant of student outcomes in the learning domain "writing persuasive essays" and that student class size varied across places. The question would be "why do some places have lower class sizes than others?" How does the education system work such that smaller class sizes are the *endogenous result* of the operation of the system? Or, if higher formal qualifications were known to lead to better student learning outcomes, then what accounts for why some systems have teachers with more formal qualifications than others? It could be (but likely isn't) that this question also has an easy answer: systems in poor countries, even making optimal choices about resource allocation, just do not have the resources to produce better learning outcomes.

The second part of this question is more difficult both to state and to answer: "What accounts for the (apparent) contextual heterogeneity in the efficacy of the proximate determinants in producing learning outcomes?"

One conclusion of any review of the empirical literature on the impact of the proximate determinants on learning is that the impact varies widely across contexts (Glewwe and Muralidharan 2015). In some places, smaller class sizes lead to better learning outcomes, while in others they do not. In some places, higher teacher salaries lead to better learning outcomes—in other places they do not. In some contexts, more teacher exposure to in-service training leads to better student learning outcomes, but in other contexts it does not. This is a much more conceptually and empirically difficult question than either researching the proximate determinants of learning, or the system-level determinants of these proximate determinants. Therefore, there is no simple arithmetic decomposition of student learning outcomes in one system versus another that relies on a single estimate of the impact of proximate determinants

(e.g. "the" effect of class size, "the" effect of having a "qualified" teacher, "the" effect of having access to textbooks) that will be particularly helpful in explaining poor outcomes. Poor system-wide outcomes are generally the result of both low levels of application of proximate determinants *and* that the efficacy of application (learning gain per increment) is small.

This current review paper addresses these system issues.

Political determinants of the system. The third related set of questions is how systems of education exist in a political context. What are the political situations in which the political system is amenable to creating and sustaining a system of education that produces excellent learning outcomes? This is the subject of a different review paper.

I)   *Overview of this paper's hard question, tentative answer, and possible path forward*

This paper is being written as part of the background to a large new research endeavor that proposes to ask and answer, with solid evidence, questions about how system-level changes can lead to acceleration in progress on learning outcomes across the entire distribution of learning. Therefore, as an overarching conceptual structure with many detailed elements, much of the paper will be hard going (of necessity rather than of simple poor writing we hope). But we don't want to lose sight of the *motivating hard question*, the *tentative answer*, and the *possible path forward*.

The hard question: **"How can learning get, and stay, so bad (in some places and not others) even though access to enrollment and schooling completed is expanding rapidly"?**

The tentative answer (as overarching hypothesis): **"(Some) education systems were created that were both closed and coherent only around schooling (at best), not around learning, and in these systems low performance on all other dimensions besides the expansion of schooling can emerge and persist in spite of efforts to improve learning."**

The possible path forward (as avenue to be explored): **"Undertake transitions from systems coherent exclusively for enrollment to education systems coherent for learning"**

I.A)   *The hard question*

While there have been massive expansions in nearly all schooling systems and many countries have high levels of learning performance, some education systems raise the question: "How can it be this bad?" It is worth presenting just a few illustrative examples.

An NGO in Egypt launched an after school program to focus on remediating deficits in basic Arabic literacy for children in grade 4, 5, and 6. A preliminary assessment was used to screen students, part of which simply asked children to write some letters of the Arabic alphabet. This revealed a shockingly large fraction of students, who, in spite of being enrolled in 5th grade, could not write a single letter. How can it be that a child—a person with her own hopes and dreams—has attended four full years of school and no one, in that entire time, has identified and addressed that simple skill? Tragically, community-based assessments of learning in India, Pakistan, Tanzania, and Uganda, and other data, often show significant fractions of children in

grades 4 and 5 unable to recognize the letters of their mother tongue alphabet, and many more unable to read a simple text (Sandefur and Beatty 2015). An assessment of learning outcomes in Bangladesh showed that the grade a child had completed had almost no predictive power for whether he could do basic math (Asadullah and Chaudhury 2013) .

"How can it be this bad?" applies not just to learning outcomes but to the outputs produced by the education system, such as teachers with mastery of the content they teach who attend school regularly and engage in classroom activity. The results from the Service Delivery Indicators (SDI) project of the World Bank across Sub-Saharan Africa are coming in and provoke the similar reaction of "How can it be this bad?"

For instance, teachers were given an assessment of whether they could correctly answer questions about the content they were teaching to children in primary school. "Minimum" knowledge was defined as teachers' ability to answer 80 percent correct on assessments of the *primary school* curriculum. In Mozambique only .3 percent, that is three out of a *thousand*, teachers reached this standard; in Togo less than 1 percent; in Nigeria 2.4 percent. Even in what are often thought of as "high performing" countries like Uganda and Tanzania, only 10.1 and 15.6 percent of teachers reach this standard (Filmer 2015).

How can it be that countries have created systems of primary education and yet in those systems less than one in ten teachers actually masters the language and mathematics competencies expected of primary school children?

Table 1: There are very low levels of mastery of basics not just by students but also by primary school *teachers* in Africa

	Kenya 2012	Mozambique 2014	Nigeria (only four states) 2013	Tanzania 2014	Togo 2013	Uganda 2013
Percent of teachers who score higher than 80 percent on primary school language and mathematics assessment	34.8	0.3	2.4	15.6	0.9	10.1

Source: Presentation of SDI results at RISE launch conference (Filmer 2015)

The SDI also measured teacher attendance and, if the teacher was present on the school grounds, whether they were in the classroom during the period that was slated for teaching. The result was that in many countries the combination of school and classroom absence implied that the actual potential instruction per day was often far less than half of the planned school day. In Kenya, the scheduled school day for instruction is 5 hours and 31 minutes; in practice the average child was getting less than half of that, only 2 hours and 31 minutes. How could it get so bad that in Mozambique on any given day 45 percent of teachers were absent from school? Again, these results are not limited to Africa, as high levels of school and classroom absence have been documented in states of India, Pakistan, and elsewhere.

Table 2: Actual classroom instructional time is often less than half the officially scheduled time

	Kenya 2012	Moz-ambique 2014	Nigeria* (four states), 2013	Senegal, 2011	Tanzania 2011	Togo. 2013	Uganda, 2013
School absence rate (percent)	15.2	44.8	16.9	18	23	22.6	29.9
Classroom absence rate (percent)	47.3	56.2	22.8	29	53	39.3	56.9
Scheduled teaching time (minutes)	331	257	284	276	312	328	433
Time actually spent teaching per day (minutes)	150	61	190	195	124	195	176
Ratio average actual teaching minutes per day to minutes per day planned teaching	45.3%	23.7%	66.9%	70.7%	39.7%	59.5%	40.6%

Source: Presentation of SDI results at RISE launch conference (Filmer 2015)

Once systems and organizations reach a certain level of incoherence and dysfunction, empirical findings emerge that invert the conventional wisdom and again provoke the question, "How can this be?" That is, once education systems reach a certain level of dysfunction the impacts of policy actions will be unpredictable using the "conventional" approach. Five examples:

1) Economists expect that reducing prices will lead to increased usage. But Bold, Kimenyi, Mwabu and Sandefur (2013) show that the introduction of free primary school in Kenya in 2003 actually led to no increase in net enrollment in government primary schools between 1997 and 2006—while enrollment in private primary schools doubled. How can it be that making something free does not increase use? Bold et al. argue the most plausible interpretation was that parents anticipated that free schools would be worse schools.

2) Many believe that higher pay for teachers will lead to better outcomes for students. Indonesia more than doubled teacher pay and a recent impact evaluation of that massive

increase, which leads to additional salary costs of billions of dollars per year, found zero impact on student learning (De Ree, Muralidharan et al. 2015). How can it be that paying teachers more—lots more—doesn't lead to better student outcomes?

3) Nearly every discussion of increasing the quality of schools suggests reducing class size as one element of improved quality. Yet, in one district of Kenya, Duflo, Dupas and Kremer (2009) found that reducing class size by adding a civil service teacher had no impact on student learning. In contrast, adding a contract teacher did have substantial impacts on learning. When the policy of reducing class sizes through additional contract teachers was scaled throughout Kenya, Bold, Kimenyi, Mwabu, Ng'ang'a, and Sandefur (2013) found that that the program had the expected impact on learning only when it was implemented by an NGO. How can it be that reducing class sizes through adding regular civil service teachers—or even adding contract teachers through government—doesn't improve student learning?

4) Nearly everyone agrees that providing textbooks to students who lack them will improve learning outcomes. Yet, as Glewwe and Muralidharan (2015) point out, there are now four different rigorous impact evaluations of textbook provision in four different places that do not find a significant increase in the learning of the typical student. And all four studies can provide, with evidence, a *different* reason why this is so. How can it be that providing textbooks to students without them doesn't increase learning in these contexts?

5) India undertook a massive federal program to improve enrollments and quality of its basic education which increased central government expenditure eleven-fold (in nominal terms) over the course of a decade (Dongre, Kapur et al. 2014). During that same period the only assessments of learning that are comparable over time show massive *declines* in student mastery of reading and arithmetic. Moreover, the proportion of students in government schools is on a steady decline. In some states, like Tamil Nadu, there has been a massive *absolute* decline in enrollment in government elementary schools, and there are a million *fewer* children per year enrolled. Some of this is explained by demographic changes, but most by a shift to the private sector (Pritchett and Aiyar 2014). How can it be that massive fiscal efforts to spend more on schools are associated with deteriorating learning outcomes?

I.B) The tentative answer

The hard question is: "How is it that some countries are 50-60 years into pursuing universal primary education as a goal, yet have learning outcomes that are so awful?"

Worse, all the easy answers to the hard question of "How can it be this bad?" are ruled out as general explanations. Pretty much everything everyone believes is the key element of better schools has, by now, been rigorously disproved to have an impact on student learning somewhere. Of course, many of these same notions have also been rigorously proven to have an impact on student learning somewhere else, and are characteristics of well-performing education systems. That is, if one observed that learning outcomes of students were awful and that they lacked textbooks, then the easy answer suggests itself: "Learning is bad because kids lack textbooks." But if studies show that, in places where learning outcomes are bad, adding textbooks doesn't make things better, this obviously cannot be the answer as to why things are so bad. The same logic applies to better teacher pay, small class sizes, teachers with more formal

qualifications, larger block grants, etc. Even if it is the case that these same factors (class size, textbooks, teacher pay, and so on) are proven to matter in some, often well-functioning, systems, this doesn't mean the causal explanation of poor performance in a given country or region is a lack of these proximate determinants or simple policy elements.

An analogy is a car. We all have the empirically well-honed intuition that a car can go further with more gas than with less gas. This is because for cars that work, it is true. But if I have a car whose transmission has failed then adding more gas will not add to miles travelled—even if it allows the motor to run for longer. If the system does not add up to a functional whole, the causal impact of augmenting individual elements is completely unpredictable.

The tentative (long) answer to the hard question "How can it be so bad?" is: "Systems of education were built up in many countries, primarily within governments, that were never actually designed (or emerged) as systems coherent to the purpose of producing uniformly high learning outcomes." These systems had other, often desirable, objectives, like expansion of access. They often had learning as one objective of at least some actors in the system, but the system was never *coherent* for learning.

Suppose one observed a tool being used incredibly ineffectively. Imagine, for instance, seeing a person use an electric drill to cut through lumber by drilling a sequence of holes across the board that added up to a cut through the board. One would likely conclude two things.

One, that the electric drill, as the result of purposive human design and ingenuity probably had some better purpose—it was likely really good at doing something, but cutting lumber was probably not that purpose. There was a *mismatch* between design and use, but one could work backwards from the tool (drill) to its optimal purpose (making holes). After seeing a drill used as a saw, when one later observed a drill used for its purpose of making holes one would say "I see, it is not actually a terrible saw; it is really good at drilling holes, it is a drill."

Two, that there must be a better tool (or a better tool can be invented) to accomplish the purpose of sawing lumber. That is, while one could cut a board into two pieces using an electric drill, there was almost certainly a tool that accomplished the same purpose more effectively—wasting less time and lumber and achieving a better result.

Existing developing country education systems have been spectacularly, fantastically, successful at their purpose. But that purpose wasn't learning (or learning wasn't high among its the priorities). Hence, seeing existing educational systems aim at the purpose of learning is like seeing a drill used to cut lumber.

If one accepts that the purpose of existing education systems was to expand *schooling,* that is to increase the exposure of children to a formal institution called a "school" (both through expanding the fraction of all children exposed and increasing the duration of exposure per child), then system design and purpose seem very coherent.

The most obvious manifestation of the fit between education systems and a coherent goal of increasing exposure to school is that this objective is all that most education systems have ever measured with any reliability. Nearly every schooling system in the world can track enrollment over time, across space, and by grade with high frequency (annually, at worst) and with some

precision (less that it appears, but some) (Pritchett 2014). Yet if one asks "Are children, at a given grade, or on exit, mastering more or less of the competencies the school systems claim to seek to impart?" almost no country in the world can answer this question with reliability and precision over any appreciable length of time. Until strikingly recently, almost no developing country measured and tracked learning performance over time on a universal basis (except for "high stakes for the student" tests such as leaving or entrance examinations, and even these were not comparable over time).

The other key inferred purpose of most education systems is that they behaved in such a way that it appeared an explicit goal of the system was to operate in the way a "modern" bureaucracy does, independently of this furthering another purpose. That is, an apparent *objective* of education systems was to operate based on process compliance with policies that put an emphasis on form over function (Pritchett, Woolcock et al. 2010). Again, the most obvious manifestation of this is that process compliance is what is measured and on which basis, particularly but not exclusively, governmental schooling organizations are explicitly operated.

For instance, every government has procedures for how teachers are hired which depend nearly exclusively on formal qualifications (the completion of certain types of schooling) and perhaps on some score on a pen and paper examination. Nearly all governments hire teachers into something like a "modern" "civil service" employment status. The explanation for this seems to be that it is part of the internalized objectives of the system (e.g. that teachers be hired in this way, on these criteria and into this status) rather than that this is demonstrably instrumental to some other objective. It has been shown again and again, in country after country, that the formal qualifications used in hiring have little or nothing to do with student performance. In fact, in the worst systems these seem to be perversely related; teachers with formal qualifications do worse. Moreover, recent evidence from India (Atherton and Kingdon 2010) and Kenya (Bold, Kimenyi et al. 2013) seems to suggest that "civil service" employment status *per se* worsens teacher performance in learning. It is impossible to rationalize many common features of governmental schooling systems instrumental to an objective of student learning. Their ubiquity must be explained in other ways—perhaps that these approaches are, somehow, themselves the objective.

The assertion that education systems have these two objectives—exposure (enrollment/grade attainment) and bureaucratic (appearing) process—begs the question as to why particularly government systems have these as objectives. That remains for the third in the set of questions above, to be addressed in a separate paper about the political determinants of education systems.

The tentative answer to the questions "How can it be so bad?" and "Why don't things that work elsewhere, work here?" is "The existing system is coherent around other objectives, that do not produce a system in which universal attainment of high levels of learning becomes the driving force of key actors' (organizations and individuals) behavior."

I.C) *The possible path forward*

If the tentative answer is correct this implies an entirely different answer to questions about how to make education outcomes better than the conventional wisdom.

The conventional wisdom acts as if governmental education systems do already have learning as a priority objective and assumes that bureaucratic processes are capable of implementing whatever "policy" is adopted. If these two assumptions were so, then there would be no need for "systemic" reform as (a) more resources into the existing system would produce better learning outcomes, as the working assumption is that lack of resources is an important and binding constraint to better performance, and (b) adopting better policies, particularly ones based on "evidence" (mostly from other places), would lead to better learning outcomes.

This leads to advocacy around an agenda of (a) reallocating resources from other places towards spending on education and (b) promoting the adoption of policies and practices "demonstrated to work" and/or the conventional wisdom of sector experts.

If, however, the problem is that the existing system simply does not have universal high levels of learning as an important objective (in the odd ontological sense, detailed below, in which "systems" have "objectives") then neither of these approaches above should expect to be successful—except of course at perpetuating the system as it roughly is. Therefore, in these cases, a substantial acceleration in the pace of progress in improving student learning will likely require substantial systematic change.

II)   *Systems as an (non) object of study*

So far (and I acknowledge we are far, about 5,000 words far, into this essay) I have bracketed the meaning of the word "system." Let us turn to that.

There are two dominant modes of "explanation." One, particularly relevant to the natural sciences, is explaining phenomena as the properties of objects. The other, particularly relevant to the human sciences, is to explain outcomes as the result of deliberate actions of agents to achieve teleological purposes.

But if we want to understand why the average 15-year-old in Korea scores 554 on the PISA assessment of mathematics and the average in Vietnam is 511, versus an average of only 375 in Indonesia and only 338 in Himachal Pradesh India (OECD 2010) neither of these modes of explanation seems particularly promising. These gaps imply even a good student in Himachal Pradesh has less capability (in this domain, on this assessment) than poorly performing students in Korea. These differences are not well explained by mechanical "proximate determinants" like features of the physical infrastructure or class sizes or access in learning inputs. Although these differ, they do not differ by enough for any decomposition analysis to begin to account for the massive gaps in outcomes.

Explaining differences of this magnitude in terms of the teleological motivations of agents (e.g. "students in Korea/Vietnam study hard" "teachers in Korea/Vietnam strive to do a good job") just begs the question. Students in classroom A in Korea might do better than students in classroom B in Korea because the teacher in classroom A is particularly intrinsically motivated. But to explain the differences between Korea and Indonesia in terms of *average* differences of teacher motivation or of *average* student effort simply begs the question of why teachers in one country are so much more motivated than in another or students put in more effort. That is, it may well be correct that "motivation" is the proximate answer to why learning

10

outcomes are better, but these motivations are the *endogenous* result of the operation of a system that determines agent motivations, not an intrinsic feature of the agents. It is not that Vietnamese teachers are just born more motivated to do a good job teaching than Indian or Indonesian teachers are. Something about the systematic way teachers are recruited, selected, trained, or motivated must account for the average difference.

The answer to at least some important questions about the level and variability in learning outcomes across countries must lie in some things called "systems" which serve to structure the actions of agents. There are at least four big empirical questions that a systems approach seeks to answer:

- What explains the geographic (across countries, states/provinces) variation in the average level of learning outcomes in specified domains?

- What explains the *changes over time* in the average level of learning outcomes? Why are some countries progressing and others not?

- What explains the *magnitude* of the variability in student learning outcomes? Why are there larger inequalities in learning outcomes across students in some places than others?

- What explains the *structure* of inequality in learning outcomes across places? Why in some places are student outcomes more strongly associated with child (e.g. gender) or household (e.g. socio-economic status) than in others?

The goal of answering these questions is *action*: are there ways in which achievable changes to the features of the education system can produce more desirable outcomes—high levels, faster progress, less inequality, less exclusion?

II.A) What is a "system"?

A system is a collection of elements or actors, each of which has its own objectives, and a collection of feedback loops connecting the elements/actors. The feedback loops provide information to elements/actors on the basis of which their actions/behavior can change, and, conditional on their actions, information on their success relative to their objectives, which also can change their actions. A dynamic system also has feedback loops in which elements/actors may disappear from the system.

An obvious example of a system is an *ecosystem*. An ecosystem's elements/actors are living organisms that respond to conditions (like amount of sunlight, moisture in the soil, the availability of food) with actions. These actions in turn affect the conditions for other elements/actors; for example, growth of one plant may block the sunlight available to another, the successful reproduction of one species may reduce the amount of food available for it and other species. Dynamically, changes in conditions and responses from elements/actors can produce changes in the ecosystem both in the short, medium and long run. Ecosystems evolve over time in ways that are determined by the features of the system and not the teleological objectives or purposes of any agent or actor, or anything inherent in the properties of the plants or animals alone. While there is a common tendency to anthropomorphize and turn an ecosystem

into a type of agent (e.g. "Mother Nature") the important point is that the *ecosystem has no purpose*. It cannot have a purpose, as the ecosystem isn't itself an agent.

A second obvious example of a system is a *market.* A market is a system whose elements are agents (people and organizations) who have purposes and objectives and who interact with each other in structured ways, like trading labor to firms for wages or firms selling goods for money. There are informational and resource feedback loops that inform the actions of agents and allow them to change their behavior dynamically, and there are processes whereby agents enter and exit markets.

Ecosystems and markets illustrate several important points about systems.

First, complex dynamic systems can have *emergent* properties. An emergent property is something that emerges as a result of the operation of the system that is not an intention of any agent in the system. Adam Smith of course famously observed that the butcher and baker do not give you meat and bread out of altruism, but because it is in their own best interest. Yet, a system of each actor within a system operating on the basis of self-interest can nevertheless produce a property of "Pareto optimality," which is that no agent in the system can be made better off without making another agent worse off. This is a (potential) emergent property. No agent in the system intended Pareto optimality. The system (market) itself, which can be described as the actors, their intentions, and relationships and feedback loops among the agents, isn't the kind of ontological object that has intentions. The common metaphor of an "invisible hand" is itself misleading because it makes it seem as if "the market" is an agent that has purposes it achieves with actions of its "hand." Nevertheless, systems can have emergent properties that are desirable or undesirable to the agents within the system even without intentionality.

Second, complex dynamic systems can exhibit progress even without any individual agent changing their behavior. This is called *ecological learning*. The process of evolution is the most obvious example as random variation in genetic coding, plus selection so that favorable mutations survive, can lead to organisms with better and better functioning. Vision, for instance, appears to have evolved from differential light sensitivity to the complexity of human (or eagle or other animal) vision, with the ability to perceive color, depth, texture, etc. This wasn't the product of any individual animal "learning" to see or of the "system" (or any anthropomorphism) "wanting" better vision, but just that incremental improvements provided survival value to organisms and hence more of those survived than others. The same is true in markets. The productivity in an industry can improve not because any existing firm increases its productivity but because firms with higher productivity are more likely to survive over time than firms with lower productivity.

Third, over time differences in the characteristics of systems can lead to massive differences in outcomes. Modern human beings are almost identical in their biological potential to human beings 10,000 years ago. Yet if one looks at the total productive potential of the human species it has an "L" curve shape; in that overall populations and output per person were roughly stagnant for thousands of years and then, suddenly, took off and have continued on a growth trajectory. The only explanation for this is that the systems whereby human beings interacted changed to unleash potentials for increasing output that have allowed human beings to increase their populations and living standards by *orders of magnitude* (Seabright 2010) (North, Wallis et al. 2009). The global inequality in productivity across people today is primarily a product of the

fact that some people are embedded in productive systems and other people are not—as is manifest by the fact that migration of a person from one place to another can increase her economic productivity (wages) ten-fold with no change in "human capital" of any kind (Clemens, Montenegro et al. 2008).

A human system is a collection of actors (individuals and organizations). Each actor has objective(s) or goal(s); the actors have possible actions; and the actors have direct and indirect relationships with each other in ways that provide feedback loops both informational and consequential for their objectives. As all human systems are dynamic, all of these components may change over time—new actors may enter into the system, technology may change the possible actions, the feedback loops may get stronger or weaker.

Systems have emergent properties, may exhibit ecological learning, and, while systems do not have objectives, characteristics of the system may have enormous impacts on the extent to which actors are able to achieve their objectives.

II.B) What is a system of basic education?

> *An "official" educational enterprise presumably cultivates beliefs, skills, and feelings in order to transmit and explicate its sponsoring culture's ways of interpreting the natural and social worlds.*
>
> *Jerome Bruner, The Culture of Education, 1996*
>
> *In administration of all schools, it must be kept in mind, what is to be done is not for the sake of the pupils, but for the sake of the country.*
>
> *Mori Anori, Japanese Minister of Education, 1886-1889*
>
> *Let it be understood that the first duty of a democratic government is to exercise control over public education.*
>
> *Jules Ferry, French Minister of Education, 1879-83*

Before describing a system of basic education it is worth placing "basic education" or "schooling" into the broader class of "instruction" and hence the systems (and markets) for "instructional services."

Each of us has acquired an array of skills/competencies/beliefs. Many people play a musical instrument. Many people are adept at a sport or exercise. Many people learned a second language. Many people acquired expertise in a complex game like chess or bridge. Many people have received religious instruction. Many people acquired occupational skills like farming or tax preparation or policing or hair styling.

Mastery of any of these skills/competencies requires instructional services, a sequence of instructional episodes with either paid instructors, volunteer instructors, training classes, or self-instruction (often with a book or on-line material as a guide but sometimes just learning-by-doing).

There are two key points about "instructional services."

First, in most countries the system for the acquisition of instructional services is an ordinary, garden-variety market for a "private good" (one which is economically both rival and excludable) and does not invite any extra governmental attention, regulation, or support. That is, the market for private music lessons looks like the market for any other personal service (say, a massage or housecleaning). There are no special subsidies nor is there any special regulation.[2]

Second, instructional services are instrumental to the maintenance of the broader endeavor. If a religious denomination or practice is to survive is must be able to attract and retain people into sufficient instruction. If classical Western orchestra is to survive as an endeavor, sufficient people must both be attracted into sufficient instruction to produce people with mastery of the array of instruments to produce the music. There must be a sufficient base of people with a taste to consume the music to support it. So instructional services are embedded in a larger endeavor and the success of the broader endeavor depends on the reach and efficacy of the instructional services.

On one level a "school" can be understood as a co-located, bundled, sequenced, set of instructional services. That is, what historically proceeded "school" for the elites was a child receiving lessons from private tutors at home in a variety of subjects.[3] The modern "school" was largely a disruptive innovation (in the sense of Christensen (1997), a low-cost, inferior substitute that met the needs of lower-income users) to the previous world of a collection of tutors or specialized trainings, that allowed expansion of education to proceed.

As we begin to describe a country's system for basic education the first and most obvious fact is the deep and pervasive engagement of the government (central, state or local). This is striking as it is so completely at odds with nearly any other kind of system of instructional services. Even before describing the system it is worth asking why this is so and, as emphasized early on by Blaug (1976), standard approaches of economics are of little use in this.

What is the endeavor or "Project" (in the grand meaning of the word) to which "school" or "basic education" is the integral instructional service support? In analogy test question format: "school" is to X as "violin lessons" are to "classical music" or "Sunday School" is to "Protestant denomination"? I believe the answer is "modern nation state."

As Moore (2015) emphasizes, and as I detail in the companion paper on the political economy of basic education, "schooling" is not a "public good" in the way in which economists define the term—a good that is non-rival and non-excludable, like public defense, clean air, or disease surveillance, as opposed to garden-variety "private goods" like shoes, an apple, or a house, that are both rival and excludable. As we noted above "instructional services" are in fact garden-variety private goods for which there are typically thick and (reasonably) well-functioning markets.

Schooling is therefore a "public good" not in the economic sense but in the sense that it is the collection of instructional services that define and reproduce the "imagined community"

---

[2] The exception is instructional services associated with licensure of various types in which governments may designate only certain bodies as eligible to provide instruction, but this is more a consequence of the regulation of the activity itself than of instruction per se.

[3] In the opening of his famous "Meditations" Marcus Aurelius thanks Catilius Severus for encouraging him to avoid public schools.

(Anderson 1983) of the "public" itself. As Jerome Bruner (1996) and many others have noted, "education" is about transmitting the values, beliefs, skills that support and maintain a society or culture or common sense of a "public."

With that as preface, what is an education system? Since any human "system" is composed of "actors" and "relationships" among actors a brief description of a system of education must have at least four sets of actors:

Citizens/Parents/Students. We can and will parse out the different roles and interests of citizens, parents and students later. For now we will treat this group as those who (1) ultimately control the sovereign, (2) are the direct participants and intended beneficiaries of instructional services via schooling, and (3) those under whose objectives the outcomes of the system are normatively evaluated (both individually and collectively).

The Executive Apparatus of the State. These are the actors who control, via executive decisions like the adoption of laws, regulations, policies and the allocation of budgets, the apparatus of the state. Later we will complicate this by considering tiers of the state, but for now this can be either a nation-state or, in a federal system, the tier of government with primary responsibility for schooling. This includes both politicians and executive policy makers.

Organizational providers of schooling. These are schools and organizations (like Ministries of Education) that control schools. Some schools (particularly in the private sector) are single "mom and pop" operations that are embedded in no larger organization, other schools are embedded in a larger organization, whereas many governments control nearly all aspects of the operation of schools through a large bureaucratic structure.

Teachers. These are the individuals who are the "front-line service providers" who are the direct producers and providers of instructional services.

To complete any description of a system one has to specify how the actors are connected through relationships that provide feedback loops. The "accountability triangle" of the World Development Report 2004[4] provides a useful starting point for such an analysis. It assumes that there is a relationship between each of the actors called a relationship of "accountability."

A relationship of accountability is like an economist's "principal-agent" relationship in which there are "design elements." Design elements can, at least in principle, be structured by the "principal" and then the performance of the agents is endogenous, that is, is chosen by the agent(s) in response to the design elements and to how these elements themselves over time affect other determinants of individual behavior, like intrinsic motivations, professional norms and standards, and organizational practices.

Principal agent relationships are ubiquitous in markets and human relationships more generally and can involve a single principal and a single agent (e.g. you hiring a car mechanic), many principals attempting to control a single agent (e.g. diffused shareholders as owners of a firm attempting to control the CEO), or a single principal attempting to influence many agents (e.g. a company with many employees).

---

[4] I was a co-author of this WDR and the triangle itself was an outgrowth of Moore's (1995) "strategic triangle".

The four "design elements" of a relationship of accountability are:

Delegation. The principal specifies what it is he/she/they want done. This specifies the principal's objectives or goals.

Finance. There is usually some amount of finance provided up front and/or promised so that the agent can decide whether this is adequate relative to the objective to be achieved.

Information. The principal can specify the information on which their decisions about the adequacy of the performance of the agent will be based.

Motivation.[5] The agreement between the principal and agent specifies what will happen for the agent if the outcomes are good (relative to the delegation specified and based on the information available) versus if the outcomes are bad. These can be positive or negative and intrinsic or extrinsic (pecuniary) motivators.

If one combines the four (classes of) actors (citizens/parents/students, the state, organizational providers from ministries to schools, and teachers/principals/headmasters) and the four "design elements" of each relationship of accountability this leads to an (adapted) accountability triangle (see Figure 1). Each arrow in this diagram has analytic content as the red arrows represent a relationship of accountability which embeds within the elements of design, finance, information, and motivation.

"Citizens/parents/students" as multiple "principals" with the state as an agent includes both the privileged and the marginalized along any number of possible economic and social stratifications. The coalitions and inclusiveness in the social and political processes determine whose voice is heard in the political process. This is not a relationship of accountability but is an important determinant of outcomes. In the accountability triangle framework, improvement in education for all can happen when the demands of the citizens as heard in the system are inclusive and equitable.

---

[5] In the original WDR framework this was called "enforceability" to capture the ability of the principal to act on the information available. I feel this is both too narrow a conception of the array of actions available and biases thinking towards "punishing" bad agents rather than a more symmetric conception of rewards and punishments.

The State
(Executive)

Politicians

Policy Makers

Politics
(D,F,I,M)

Compact
(D,F,I,M)

Citizens/Parents/Students

Coalitions/Inclusion

Rich/Privileged

Poor/Marginalized

Voice/Client Power
(D,F,I,M)

Management
(D.F.I.M)

Frontline providers (teachers, principals, headmasters)

Organizations (ministries, schools)

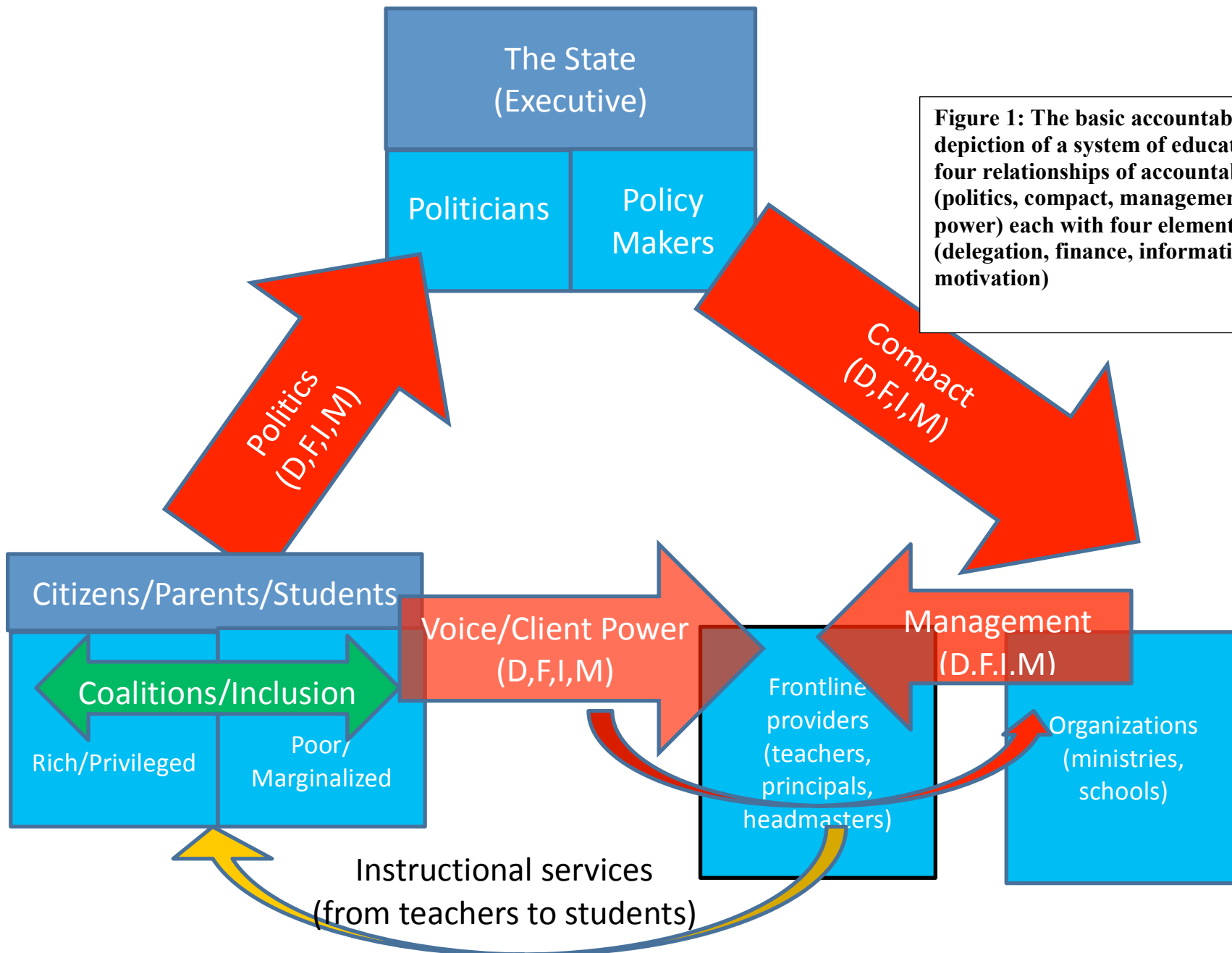Instructional services
(from teachers to students)

Figure 1: The basic accountability triangle depiction of a system of education with of four relationships of accountability (politics, compact, management, and client power) each with four elements (delegation, finance, information and motivation)

This accountability figure, like a wiring diagram for an electrical system, helps set up a vocabulary for a diagnostic for systems of basic education. If, in order to have electricity to flow from point A to point B requires several open connections for electrical flow then a failure to have power at any given socket could be a result of the wiring inside the socket, could be a failure in the wires to the socket, could be a tripped breaker, or could be because there is no power flowing to the house.

My grand hypothesis is that governmental (sub) systems of education work when there is an adequate flow of accountability in the system. In the best possible case this is because politics work well so that citizen/parental demand for education is made a salient issue for politicians and their top policy makers, the executive apparatus of the state is able to convey these effectively to organizations, organizations are able to create the conditions in which teachers can work well, citizens and parents can also participate locally and directly, and high quality instructional services of the types that are contextually relevant are the result.

But weaknesses in one domain can be compensated for. For instance, a strong authoritarian state can promote effective education even if disconnected from a strong relationship of accountability via politics.

This approach provides a useful vocabulary for discussing weaknesses in specific dimensions of the existing system design elements that provide alternative conjectured explanations for performance of agents in the system. For example: "There is weak learning performance, because learning isn't clearly articulated as an objective in the *delegation* element of the *compact* relationship (from the "state" to "organization," e.g. Ministry of Education) versus other clear targets (e.g. enrollments, buildings, inputs)"; "There is weak learning performance because there is weak *information* gathered by the organization (Ministry of Education) about learning, relevant to the *management* relationship with schools or teachers"; or "There is weak learning performance because parents, even when they know their school is bad, have little scope for *motivation* over schools or teachers in the *client power* relationship."

I will use this tool to make the basic points of this essay.

Table 3: Four by four diagnostic for systems of basic education				
Four design elements of each relationship of accountability (Principal (P) to Agent (A))	Principal-agent relationships			
	Politics: Citizens to "the state"/politicians (many P to one A)	*Compact:* "The state" to organizations (one P to one A or one P to many A with non-state providers)	*Management:* Organizations to front-line providers (one P to many A)	*Voice/ Client power:* Service recipients (parents/children) direct to FLP/Organizations (many P to one A)
Delegation: Specification of what P wants from A				
Finance: Resources that P provides to A (either in advance or contingent)				
Information: P collects information on performance of A				
Motivation: How is A's well-being contingent on performance? Change to motivation? - Intrinsic - Extrinsic - Exit (force out)				
Performance of agent (endogenous)				

II.C) One-word descriptors of education systems are useless as functional detail needed

Given the argument that the important determinants of cross-national learning outcomes are systems, the unsuspecting reader might think that we are going to discuss "centralized" versus "federal" education systems, "public" versus "private" systems, "local" versus "top-down" systems, or some other binary contrast. But we are not because, by and large, one-word system descriptions are useless.

Imagine four countries. One has historically had a "money follows the student" system so that private providers account for over two-thirds of students. One has historically had a very centralized national system premised on the notion that the Minister of Education could look at his watch and know what every child in the country was doing. One has always had a strongly federal system so that states could resist nearly any central government initiative in schooling. One historically has had a localized system with 15,000 largely autonomous school districts, most very small.

If "systems" as simple one-word descriptions were an important part of the determinant of differences in learning outcomes then one might expect these four countries to have very different outcomes, perhaps in levels, perhaps in inequality, perhaps in structural inequality.

This description crudely captures the differences between the Netherlands (privatized, primarily to religious schools, as public money follows the student), France (national and centralized), Germany (federal), and the USA (local).

But if one looks at the PISA 2012 outcomes (which is just one assessment tool and operates on just three domains of learning) the differences across these four systems are modest. The average PISA score is 519 in the Netherlands, 518 (for whites)[6] in the USA and 515 in Germany. The difference between Netherlands and Germany is only 4 points, which is only 6 percent of the European range (between Finland and Greece) and only 2 percent of the global range (between Singapore and Peru). France, on average, lags considerably behind the Netherlands, at 500, but even that difference is only 10 percent of the global range (OECD 2012).

At the very least we have to conclude that the seemingly very different "systems" of *privatized, centralized, federal,* and *localized* systems are *capable* of producing adequate results (at least for their dominant race/ethnicity when there isn't "inclusion" in the aggregation of citizens).

---

[6] A main feature of learning outcomes in the USA is stratification by race/ethnicity and hence a failure of social inclusion and an example of where the system of accountability works well for one group but not for another. Any discussion of how a "local" system of education works compared to others has to tackle the question of inequality. If one compares results for just "whites" in the USA to other countries one sees the results are remarkably similar to the Netherlands. In contrast, the gap between whites and blacks in the USA is 84 points, almost half the global range of outcomes. South Africa, another country with a prominent racial divide, has even larger gaps as the top quintile of learners have roughly OECD outcomes while the bottom quintiles have outcomes worse than other African countries.

Country (or subset)	Reading	Science	Math	Average	Difference from Netherlands on average across all three domains	High to Country Difference /European range	High to Country Difference /Global Range
Netherlands	511	522	523	519	0	0.00	0.00
USA (White)	519	528	506	518	-1	-0.02	-0.01
Germany	508	524	514	515	-4	-0.06	-0.02
France	505	499	495	500	-19	-0.30	-0.10
USA (average)	498	497	481	492	-27	-0.43	-0.15
USA (African American)	443	439	421	434	-85	-1.35	-0.47

Table 4: Differences in results across four generic types of systems in OECD countries—privatized, centralized, federal and local—are small.

Source: OECD 2014 and OECD (2012) for USA breakout by race.

The converse is also true as there are big differences in performance between countries that are "federal" or "private" or "decentralized." At the crude characterization level, there are highly functional and (near) dysfunctional systems of each type. For example, "top-down" systems include high performers and low performers, and countries with near-identical inherited "transplanted" colonial systems have widely differing results.

So in both Germany and India, the basic education system is "federal" as the constitution makes basic education a state subject; but India's outcomes and Germany's outcomes are wildly different.

A striking example is Chile. In 1981, Chile adopted a reform that both municipalized the system and allowed public money to follow the student into private schools that chose to receive the government monies. There was a substantial shift of students from public into private schools. And, while there was a substantial debate about whether students learned more in the private or public system, it was clear the overall general effects on average learning were, at best, modest (Hsieh and Urquiola 2006). The first year of Chilean participation in PISA in 2000, almost 20 years into their reform their reading score was 410 whereas comparable countries

without reform were performing about the same: Argentina's reading score was 418 and Mexico's 422. So being an increasingly "privatized" system *per se* had clearly not led to the transformational gains that some might have hoped (this is not to say it was "better" or "worse" than a counter-factual without reform, just that twenty years into the reform it was similar to neighbors with similar or worse socio-economic conditions).

No research program on systems of education can usefully begin hoping to find useful lessons about whether "centralized" or "federal" systems, "public" or "private" systems, or even "top-down" versus "bottom-up" systems work better.

III) *Coherence as a key concept for analyzing systems*

The description of an education system as a set of relationships of accountability can be used to articulate three types of *coherence* in a system of education. Differences in the coherence of accountability relationships across systems can potentially account for the heterogeneity in outcomes.

III.A) *Coherence as a key concept for evaluating education systems*

III.A.1) *Coherence within a relationship of accountability*. Each relationship of accountability has four design elements: delegation, finance, information and motivation. One conjecture is that in order to be effective there has to be at least some minimal coherence across the four. There are many possible examples of incoherence.

Delegation-Finance. There may be incoherence between delegation and finance, in a variety of ways.

One is that the *magnitude* of finance made available (both in terms of wages and resources for inputs) is just incompatible with what teachers are being asked to accomplish in the compact or management relationship. That is, it may well be that "the state" declares a large number of very lofty and desirable goals that it wants its education system to achieve but then makes insufficient resources available to the Ministry or other organizational providers to achieve those goals. This makes any powerful *compact* accountability relationship (between state and organizational providers) impossible, as no Ministry can be punished for not achieving targets in the delegation as they were impossible. Insufficiency of resources relative to goals means that there will be gaps between the actual accomplishments and the delegated objectives and this allows the Ministry to prioritize which of the many stated goals it will actually pursue and/or not accomplish any objectives (including some that were achievable), blaming the lack of resources as the cause of failure.

A second potential *delegation-finance* failure is that the finance is tied into categories that are independent of the delegation, so that "the state" may delegate that ministries achieve various goals but all, or nearly all, of the finance made available goes for salaries, leaving the Ministry with little or no scope to actually undertake programs or actions to pursue any goal other than hiring and assigning teachers. This is a very common problem as often funds are tied into very tight and restrictive budget headings that leave little or no autonomy to the Ministry (or organizational provider more generally) to act to achieve goals.

22

Delegation-Information. It is a commonplace that "what gets measured gets done" which I believe is less true than the converse: "what doesn't get measured doesn't get done" (as lots of things that do get measured do not get done anyway). So in the *management* relationship, for instance, between Ministries and public sector teachers, there may be a very long list of things that are stated as desirable goals for the education system but very few which are measured regularly and reliably. The obvious point (returned to below) is the disjunction between measures of enrollment—which nearly every education system has measures of at least every year at very detailed levels—and of learning, which are often completely absent. This is not to endorse "high stakes" for school or teacher use of examinations/tests/assessments; but if student progress on competencies is going to be an important part of the delegation in any of the accountability relationships *(politics, compact, management, client power)* then there must be some information about learning outcomes available or the accountability relationship is incoherent.

Delegation-Motivation. Another possible lack of coherence in an accountability relationship is between delegation (the declared objectives and goals) and motivation (how agents are rewarded). It is very common in public sector bureaucracies (not just in schooling) that organizations have declared goals but that the structures of rewards such as promotions are strongly structured by agent characteristics, such as seniority, that are largely disconnected from actual performance. So the *management* relationship might declare in the statement of *delegation* that they want the organization/school to achieve a variety of desirable teacher behaviors, but then fail entirely to act to motivate teachers—in any way, intrinsic or extrinsic, on the basis of their undertaking the desired behaviors. This leads to obvious incoherence internal to the *management* accountability relationship.

Table 5: Illustration of potential incoherence *within* a single relationship of accountability, illustrated with *compact* (between executive apparatus of the state and organizational providers, e.g. between a Ministry of Finance and Ministry of Education)

Four design elements of each relationship of accountability (Principal (P) to Agent (A))	*Compact:* "The state" to "organizational providers (e,g. one Principal (e.g. Ministry of Finance) to one Agent (e.g. Ministry of Education) or one Principal to many Agents with non-state providers (e.g. state resources follows the student to schools)		
	Delegation to Finance incoherence	Delegation to information incoherence	Delegation to motivation incoherence
Delegation: Specification of what P wants from A	Delegation lists many ambitious objectives	Delegation lists ambitious learning goals for provider	Delegation lists ambitious learning goals for provider
Finance: Resources that P provides to A (either in advance or contingent)	Provides insufficient or inflexible finance		
Information: P collects information on performance of A		Only enrollment information collected, no systematic information on learning collected on a regular and reliable basis	
Motivation: How is A's well-being contingent on performance? Change to motivation? - Intrinsic - Extrinsic - Exit (force out)			Outcomes for the Ministry (and/or Minister) the same whether learning goals are achieved or not. Outcomes depend on budget utilization and process compliance.
Performance of agent (endogenous)	Cannot perform as delegation specifies. Weak *compact* accountability.	Performance of agent cannot be reliably assessed. Weak *compact* accountability.	No motivation for agent to perform well. Weak *compact* accountability.

III.A.2)   *Coherence for the same element (e.g. information) across different relationships of accountability.*

A second type of incoherence results when an element of the accountability relationship (such as *delegation* or *information*) is wildly different or even contradictory between relationships of accountability.

Perhaps the most obvious example is *information* as each of the four relationships of accountability has different information that is available to and actionable by the principal with respect to the agents.

For instance, in the *compact* relationship between the state and organizational providers, particularly inside the public sector with respect to Ministries of Education, the main information on which the executive apparatus of the state can act is *budget* information. The principal element of the relationship and the one on which regular, reliable reporting is made is the structure and utilization (across expenditure categories and programmatic designations) of the budget. Often these budgets are devoid of any performance information except on inputs—if the budget was used for the inputs for which it was allocated then this constitutes "success" in the accountability relationship. Some places and countries have "performance" budgets but often "performance" is itself recorded as inputs (e.g. classrooms constructed) or at best enrollments/grade attainment.

This is incoherent with the information in the *client power* accountability relationship as citizens/parents/students have extremely "thick" information (that is not codified into bits/bytes and cannot be reduced to easily adjudicated "objective" or "official" information) because the students experience the school from day to day. Students (at least those who attend) know whether their teacher is present or absent on any given day, students know how much "time on task" the teacher is spending, students know whether the classroom feels to them purposeful and ordered or chaotic, students know if the teacher treats them kindly or rudely, students know whether they are understanding the material and able to do the lessons or not and have some idea about their progress. However, students (and by extension their parents) know these things in an "intuitive" way that is not easily incorporated into ways in which students/parents can act as principals that can influence teachers' behavior. Moreover, the *information* collected and acted upon in the *management* relationship particularly between public sector providers and teachers and headmasters is of a completely different content and type. The information collected on teachers by Ministries of education tends to be, at best, narrow and objective information that is "thin" (in principle, easily adjudicated). Examples of this include attendance, participation in various in-service trainings, perhaps some reports of "supervision" or implementation of specific program activities, and in general process compliance (submitting other reports to the Ministry such as enrollment). The information available that students/parents may wish to act upon to influence teacher behavior in the *client power* relationship differs from that collected by the Ministry/school as the organization that is the direct employer of the teacher/headmaster.

Similar analysis and observation could be made about the other elements of accountability relationships: *delegation, finance,* and *motivation.* That is, with authoritarian states and/or weak democracies there might be little coherence between the *delegation* of what education outcomes citizens may desire via *politics* from the state and what the state may delegate via a *compact* relationship to organizational providers.

Or, as another example, the *motivation* that the executive apparatus of the state may use to attempt to elicit better performance from the Ministry (as organizational provider) as part of the *compact* relationship may be incoherent with the instruments available to the Ministry in their *motivation* of front-line providers (teachers and headmasters) as part of their *management.* For instance, Ministries of Finance may attempt to create high stakes performance-related budgets to pressure Ministries of Education to deliver, but the Ministry may be bound by existing civil service laws, regulations, contracts and practicality in the relationship with teachers and teacher's organizations from acting in ways in the *management* relationship that are coherent with the *compact* expectations.

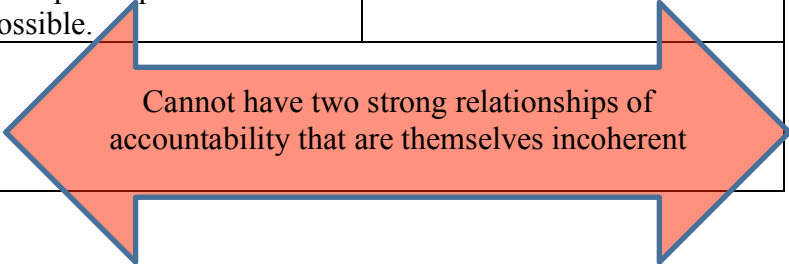| Table 6: Illustration of incoherence in the same element of accountability across different relationships: Example of *information*				
Four design elements of each relationship of accountability (Principal (P) to Agent (A))	*Principal-agent relationships*			
	Politics: Citizens to "the state"/politicians (many P to one A)	*Compact:* "The state" to organizations (one P to one A or one P to many A with non-state providers)	*Management:* Organizations to front-line providers (one P to many A)	*Voice/ Client power:* Service recipients (parents/children) direct to FLP/Organizations (many P to one A)
Information: P collects information on performance of A	Citizens know their own child's experience, but there is typically only aggregate (national/state/locality) information about enrollments, budgets, and inputs, not learning or learning progress. This often channels citizen pressure for "better" schools into these measured characteristics as politically salient.	Between the executive apparatus of the state and the organizational providers (typically Ministries of Education) the main apparatus and hence information is the budget allocation (both aggregate and across categories of expenditure (e.g. wages versus other) and program. To the extent "performance" elements are measured they tend to be measures of inputs or outputs, rarely outcomes and more rarely still learning outcomes.	Particularly in public sector organizational providers the information collected on teachers is mostly bureaucratic process compliance based on official internal systems. Teacher attendance is perhaps measured (though often not well), teacher participation in trainings, teacher reports on compliance with programmatic activities, and reports on measures of enrollment and (perhaps) student attendance. Very little information on teacher performance of any kind.	Parents/students know their daily experience with schooling and hence have "thick" information on aspects of teaching (e.g. is the teacher present, is class time boring, is the teacher kind or mean) and at least intuitive information on progress (e.g. does the child understand the lessons, is she/he able to do homework). But parents/students often lack any concrete, comparable, objective metric of their own child's position or progress or of the child's school versus others (particularly on a "value added" basis that adjusts learning outcomes).

III.A.3) *Coherence across all elements of two different accountability relationships.* The third type of incoherence occurs when two different relationships of accountability potentially affect the same agents from different principals but, even where both of those relationships are internally coherent there is incoherence between them.

The most obvious situations are potential incoherence between *management* and *client power* relationships of accountability for teachers and headmasters and between the *compact* and *client power* relationships of accountability for schools (as organizational providers).

Table 7 illustrates a hypothetical *management* relationship of accountability between a large public sector organizational provider (e.g. a Ministry of Education) and teachers/headmasters and a hypothetical *voice* or *client power* relationship between citizens/parents/students and teachers/headmasters.

Suppose each of these two were coherent as the principal(s)—Ministry or parents—would like to structure them. But if the two relationships are mutually incoherent—in that they disagree on each of the elements: different *delegation*, different *finance*, different *information*, and different modes of *motivation,* then it cannot be the case that teachers/headmasters have strong accountability relationships with both. As an early first century Jewish teacher said: "*No one can serve two masters. Either you will hate the one and love the other, or you will be devoted to the one and despise the other.*"

Table 7: Illustrating incoherence between two different relationships of accountability affecting the same "agents" (teachers and headmasters)		
Four design elements of each relationship of accountability (Principal (P) to Agent (A))	**Principal-agent relationships**	
	Management: Organizational providers (e.g. Ministry of Education) to public sector teachers/headmasters	*Voice/ Client power:* Parents/students to teachers/headmasters
Delegation: Specification of what P wants from A	Desired objectives for teachers designated as process compliance, e.g. teach in this school in this classroom these materials, not outputs or outcome performance	Parents want teachers to advance the interests of their children and to treat their children well.
Finance: Resources that P provides to A (either in advance or contingent)	Wages of teachers are fixed by teacher characteristics (whether related to learning or not)	Parents often provide little or no direct finance to teachers or school.
Information: P collects information on performance of A	Information on teacher performance based on official reports (e.g. attendance), process compliance and (perhaps) some supervision and (weak) performance assessments	Students (hence parents via students) have access to daily experiential observation on teacher behaviors and some knowledge about their own progress.
Motivation: How is A's well-being contingent on performance? Change to motivation? - Intrinsic - Extrinsic - Exit (force out)	Outcomes for teachers/headmasters based almost exclusively on seniority, cannot be fired, disciplined only with great difficulty, little extra reward for superior performance possible.	Parents/students would like to have to have only teachers who do well by their assessment of teacher performance.
Performance of agent (endogenous)	Cannot have two strong relationships of accountability that are themselves incoherent	

Summary of the potential sources of incoherence

Table 8 uses the four-by-four framework of the four relationships of accountability and four design elements of each relationship to illustrate the three potential types of coherence or incoherence in systems of basic education.

The design elements of a single relationship of accountability can be *internally* incoherent such that delegation is too overambitious relative to amount or structure of finance, or delegation is too vague for regular reliable and relevant information on performance to be created, or information created about performance is not consistent with structures of motivation. This type of incoherence can exist independent of any larger system context. A single organization can be *management* wrong with incoherent accountability.

A *system* can be incoherent when there are unconnected or contradictory conditions across relationships of accountability *within a single design element*. A system works well when there are forces working in the same direction so that the *politics* is consistent with the *compact* is consistent with the *management* is consistent with the exercise of *client power*. However, in many systems the delegation relationship—what the principal(s) have as their objective(s)—is completely different across the various relationships as the goals or aspirations of the citizens for the state may have little to do with the goals the state delegates to organizational providers and that itself may be weakly related to how large bureaucratic organizations like Ministries of Education manage teachers.

Finally, a system may be incoherent when there are contradictions between two powerful relationships of accountability that affect the same agents.

Table 8: Three types of potential incoherence in the relationships of accountability within a system of basic education: ***internal*** (mismatch within a relationship between design elements (e.g. *delegation* to *finance* within *compact*), ***within a single element*** *across relationships* (e.g. different *information* or *delegation* used in *politics, compact, management* and *client power*), ***between relationships of accountability***

Four design elements of each relationship of accountability (Principal (P) to Agent (A))	Principal-agent relationships			
	Politics: Citizens to "the state"/politicians (many P to one A)	*Compact:* "The state" to organizations (one P to one A or one P to many A with non-state providers)	*Management:* Organizations to front-line providers (one P to many A)	*Voice/ Client power:* Service recipients (parents/children) direct to FLP/Organizations (many P to one A)
Delegation: Specification of what P wants from A		Internal		⬌
Finance: Resources that P provides to A (either in advance or contingent)				⬌
Information: P collects information on performance of A	Incoherence within a single element across relationships			
Motivation: How is A's well-being contingent on performance? Change to motivation? - Intrinsic - Extrinsic - Exit (force out)			⬌	
Performance of agent (endogenous)			Across all elements of two relationships	

31

IV) **Systems coherent for "schooling only" versus systems coherent for schools with learning**

IV.A) **Coherence for purpose: schooling only versus learning driven systems**

Return to the analogy of an electric drill. An electric drill being used to cut lumber is not fit for purpose and the design of the drill as a mechanism (a small system) is incoherent to the purpose of cutting. But there is a purpose for a drill—to make holes through lumber (and other materials). One can imagine taking the electric drill and then doing experiments (even randomized controlled experiments) with small incremental changes in the drill. One could see whether lumber was cut more effectively with small bits or larger bits (and observe the trade-off between number of holes to cut the lumber versus waste, the amount of lumber turned into sawdust in the process of cutting). But no amount of variations in the drill (smaller versus larger bits, more versus less electrical power) will make the drill as effective at cutting lumber as a saw.

When one sees education systems that are ineffective and inefficient at the purpose of learning (even adjusting for the resources deployed and the socio-economic backgrounds of students) a likely conclusion is that the system was never in fact designed with the purpose of universally high levels of learning. That leaves the challenge of inferring the purpose from the design.

With many developing country systems of education this purpose is, even when not explicit, clear. The purpose of the design is to provide universal publicly controlled schooling, defined as the enrollment and grade persistence to at least some level (e.g. universal primary, elementary or basic) plus some amount of ability of the system to identify and screen for those that are adept at schooling (either by family background or native ability or drive) for further opportunities.

One way of looking at this is to take the two relationships of accountability that do not involve citizens/parents/students at all—that is *compact* between state and organizational provider and *management* between organizational providers and front-line (teachers/headmasters). The basic structure of what I elsewhere characterize as "spider" systems (Pritchett 2013) is that the executive apparatus of the state (which could be either a nation-state or a state/province) provides a budget to a Ministry of Education (and typically no other organizations) in which the *delegation* is strictly circular. That is, the *purpose* of providing wages to pay teachers is defined as paying teachers. Hence if budget is received for paying teachers and teachers are paid then the MoE can provide an adequate account of its behavior—it paid the teachers. Similarly, there might be *delegation* goals for building schools which again are circular (or complete without reference to any other desired outcomes): if schools are built then accountability is complete.

The second half of the "spider" system is that the large organization which is the (mostly) monopoly recipient of government funding for basic education engages in typical civil service bureaucratic management in which teachers/headmasters' accountability is basically for enrollments and the operation of "schooling" without reference to learning outcomes that have any traction. So while the *delegation* to teachers may refer vaguely to some goals for children actually acquiring needed competencies it is clear from the *finance* and *information* and *motivation* structures that *process compliance* is, in and of itself and with no reference to outcomes, completely adequate for discharging accountability. That is, in a typical "spider" system a teacher is a "good" teacher of third grade if the third grade class happens with sufficient regularity—and nothing else. There is often no *delegation, finance, information* or *motivation* connected with student learning at all. This is not to recommend or endorse a high-stakes for the teacher approach based on narrow assessments of learning in limited domains (e.g. standardized tests of mathematics) but the absence of any feedback loops into teacher careers based on student learning *at all* suggests, by reverse engineering from design to purpose, that student learning is not in fact a priority.

These "spider" systems have, for the most part, been fantastically successful at their purpose of expanding schooling. The number of children enrolled in school, the net enrollment rate of children, and the cumulative grade attainment have all expanded rapidly not just in a few, but in nearly all countries of the world. This means that even countries that have managed to achieve little else have managed to expand the lifetime exposure of children to a modern school. Haiti, for instance, consistently ranked among the weakest governments in the world, has seen completed years of schooling expand from .6 in 1950 to 5.2 in 2010. So "spider" systems are fit for some purpose.

IV.B) **Why are empirical results on impacts so heterogeneous?**

One of the questions we need to answer is why there is so much heterogeneity across countries and contexts in the impact of various attempts to raise learning. Sometimes strict monitoring of attendance works to improve outcomes (in NGO-run schools in rural Rajasthan for instance (Duflo, Hanna et al. 2012)) and other times it doesn't (in government health clinics in rural Rajasthan for instance (Banerjee, Duflo et al. 2008) or in health clinics in Karnataka (Dhaliwal and Hanna 2013)). Sometimes information campaigns to make parents aware of the poor learning results in the schools their children attend work and other times they don't. Or, sometimes interventions work like we expect and other times they fail to work for reasons no one anticipated, like the evaluation of providing textbooks in Sierra Leone that didn't work because teachers didn't open and use the textbooks they received (Sabarwal, Evans et al. 2014).

This is particularly pressing because one of the recent fashions, if not fads, in development circles has been the promotion of relying on "rigorous" evidence in the formulation of policy. The idea being that the application of known techniques of randomization across a control and treatment population can definitively establish the causal impact of a given "intervention" (policy or program or action) and that based on this rigorously estimated causal impacts someone (governments?) can (will?) formulate more effective policy.

The problem with this approach is that we now have enough evidence to know that this approach will not work, for precisely the reasons it was known it would not work. Without a reasonably complete analytical specification of the systemic context in which the experiments are done such that one can know that context A is sufficiently similar to context B, there is no "external validity" and hence rigorous evidence isn't. That is, rigorous evidence from context A about the causal impact on learning of program Z has no implications at all for the expected impact of program Z on learning in context B (much less is it "rigorous" evidence for context B) unless there is a theory or framework that specifies the conditions under which the A and B would be expected to have similar impacts. What is worse is that context A and context B could be the same place just different organizations; so that evidence about the learning impact of a policy of contract teachers implemented with an NGO in Kenya is not rigorous evidence about the impact of doing the same policy with the government (Bold, Kimenyi et al. 2013). What is even worse is that context A and context B could be the same place and same organization but at different times when some other interacting or intervening variable shifted. So empirically rigorous evidence about impact, in the absence of a theoretical/conceptual/analytical frame that could provide an adequate description of what is similar, isn't rigorous about anything of policy concern (as it is not rigorous, not only about other places and organizations, but even about the future).

If one were to take this approach of "rigorous evidence" at face value then there is rigorous evidence that nothing in the conventional wisdom actually works. There is rigorous evidence that giving out textbooks doesn't matter, there is rigorous evidence pay for performance doesn't matter, there is rigorous evidence that class size doesn't matter. Of course there is also rigorous evidence that all these elements of the conventional wisdom do matter. The usual approach of doing a "systematic review" of the literature that simply counts studies (in a quality-weighted basis) is not at all helpful. Suppose that context A is a system coherent for learning—so that teachers know what students should learn, that learning is measured on a regular and reliable basis and teachers are motivated to achieve high student learning—and class size is reduced. Let's assume that learning improves (as there is RCT evidence from the USA, for instance, that this is true). Context B is a system coherent for schooling only. Class size is reduced. Let's assume learning doesn't improve (as there is RCT evidence from Kenya, for instance, that this is true). Suppose the only two studies in the systematic review were USA and Kenya. Then the conclusion would be that "class size improves student learning in 50 percent of the studies." Now suppose that 8 more rigorous studies were done in the USA so that a systematic review would conclude "class size improves student learning in 90 percent of studies." Suppose, in contrast, 8 more studies were done in Kenya. Then a systematic review of the rigorous evidence would conclude "class size improves student learning in 10 percent of the studies." All three statements are equally worthless. The (assumed) truth is that "class size improves performance in context A but not in context B" and hence unless one knows whether the relevant context is A or B the systematic review finding of impact in 50 percent, 90 percent or 10 percent of the studied cases is irrelevant.

My working hypotheses are that:

a) the heterogeneity of empirical results is driven by doing the "same" programmatic/policy intervention in systems of very different coherence.

b) Interventions will often work to improve learning results when

  b.1) there is sufficient system coherence to produce a drive for better results, and

  b.2) the intervention is consistent with the existing coherence (e.g. relaxes constraints on agents achieving results they are endogenously motivated by the existing system to pursue)

c) Conversely, even otherwise rigorously proven interventions in other contexts (like higher teacher pay, greater autonomy of teachers over classroom practices, adequate textbooks, smaller class sizes) will not work to produce better learning outcomes if the system is sufficiently incoherent (in any of the three senses of incoherence detailed above).

d) Discrete interventions are likely to work if they resolve a weak design element of accountability in an otherwise coherent relationship (e.g. interventions that improve "motivation" are likely to work if delegation, finance, and information are already aligned and the key constraint is that these are not effectively translated into motivation; interventions that improve "information" are likely to work if delegation, finance, and motivation are aligned and the key constraint is better information on performance). And the converse (e.g. more "finance" is unlikely to work if delegation, information, and motivation are incoherent for learning).

e) If a system suffers from incoherence(s) in learning as an objective then success will require multiple changes of the relationships of accountability to be effective.

While there is not "proof" of any of these conjectures, let me provide interpretations of existing studies that suggest that systems that are coherent for schooling but not coherent around learning objectives are an important obstacle, even for "desirable" reforms to have an impact.

Contract and civil service teachers. Atherton and Kingdon (2010) compare the results of teachers being hired either into regular civil service positions or as contract teachers. Contract teachers made less than a third the salary of civil service teachers. Nevertheless, children learned half as much in Uttar Pradesh when in a class with a civil service teacher as with a contract teacher, even when controlling for the observable characteristics of teachers (e.g. education, age, gender). This suggests that, all else equal, having a teacher in a contractual status such that the only effective relationship of accountability is *management* via the public sector civil service—because *client power* is reduced for a civil service versus a contract teacher—reduces their absolute effectiveness in promoting learning. This is in spite of the fact that civil service teachers are paid much more—so, at least in principle, one might expect their *motivation* to be higher. But, as the *management* relationship of accountability in Uttar Pradesh is not learning-driven, nothing about delegation, finance, information, or motivation depends on student learning.

The system of accountability also potentially explains the findings about contract teachers in Kenya. Duflo, Kremer, and Dupas (2009) studied reducing class sizes by either adding a locally hired contract teacher (meeting the same qualifications are required for a civil service or Teaching Service Commission hire) but at one quarter the cost of a regular teacher via the local Parent Teacher Association or by adding a regular civil service teacher. The program was implemented by an NGO in the Western Province of Kenya. The results showed that a child

who was in a smaller class with a contract teacher had substantially higher scores on math and literacy than those who were in a smaller class due to the addition of a civil service teacher. Even though the addition of a civil service teacher reduced average class size from 82 to 44, scores increased only by .04 of a standard deviation (a magnitude not statistically significant). In contrast, the addition of a contract teacher increased student end-line test scores by .25 standard deviations. This means a contract teacher produced six times as much student learning at one quarter of the cost. These learning differentials are in part because the civil service teachers in schools in which an extra teacher was added actually *reduced* their effort and were 16 percent *less* likely to be in a classroom teaching on a spot visit than even civil service teachers in a control school. The likelihood a civil service teacher in a school that received an extra teacher was in the classroom teacher was only 50 percent.

These results are consistent with an interpretation that the *management* relationship of accountability is not coherent for student learning and also the management relationship actually *reduces* the degree to which there is any effective *client power* accountability. In contrast, when contract teachers are subject to contract renewal at the discretion of the local school committee they both put in more effort (are 28 percent more likely to be in the classroom teaching on spot checks) and produce better learning results which is consistent with stronger *client power* accountability that is coherent around learning objectives.

This interpretation is confirmed by the test of "scaling up" the findings of this study in Western Province to the rest of Kenya. When the program was scaled up nationally using an NGO as the implementer the study also found that reducing class size with contract teachers has large learning impacts (similar to the gains in the original study). However, when the program was scaled up through the government, even though the program was designed to be the same, reducing class sizes through adding contract teachers did not raise student scores at all. This suggests that when the government was implementing the program the contract-hired teachers, even though told they were subject to approval by local committees, assumed they were embedded primarily in the *management* relationship of accountability.

This incoherence between the *management* and *client power* accountability that can explain the contract teacher results can also explain the puzzling results of Bold, Kimenyi, Mwabu and Sandefur (2013): moving to free primary schools in Kenya actually reduced demand for enrollment in government schools. This result can be understood as shift in the balance between incoherent *management* and *client power* relationships of accountability. That is, if parents felt that their *delegation* of what they wanted from a school had some traction because they also provided *finance* then the reduction in the salience of that *finance* element of accountability (even if they thought the government would make good the lost resources) would lead to the front-line agents (headmaster and teachers) being less aligned with their interests.

Let me stress that my point from the India and Kenya experiences *does not* show that "contract teachers are better than civil service teachers." Few of the world's top-performing systems of education have contract teachers and most have some form of civil service teachers. In a system *coherent for learning* civil service status is no impediment. The point is the *heterogeneity* of findings across countries, and the salience of an analytical framework that can explain *why* contract teachers outperform civil service teachers in some contexts and not others and why reductions in class size produce learning gains in some contexts and not others.

Textbooks. As Glewwe and Muralidharan (2015) point out there are now four different rigorous evaluations of providing additional textbooks in four different contexts: Kenya (Glewwe, Kremer et al. 2009), Tanzania (Mbiti and Muralidharan forthcoming), Sierra Leone (Sabarwal, Evans et al. 2014), and India (Das, Dercon et al. 2013). All four find that the direct causal impact on the typical student is very near zero.

This is an important case because it illustrates what these kinds of hypotheses these impact evaluations are testing. By analogy, think of experiments with growing wheat. One could examine an RCT that looked at the response of yields to nitrogenous fertilizer or to Mozart music. In one case one knows that nitrogen is a necessary input into wheat and hence one is not testing "is nitrogenous fertilizer an input" one is testing "at existing rates of application, in existing soil conditions, in combination with other fertilizers and watering levels is an incremental application effective?" On the other hand one is testing whether playing Mozart music in a wheat field is an input at all, in the sense that one could entertain the hypothesis that wheat yields are unresponsive to Mozart music at all volumes, over all types (sonatas, operas, piano concertos), and at all times. So, everyone knows that textbooks are an input into education. So the finding that the impact of additional textbooks doesn't matter in a given context implies that some interactive effect is offsetting the potential gain.

In each of the four cases the authors assert a different interactive effect. In the Kenyan case Glewwe et al. ultimately conclude the textbooks were too hard for the typical student and find that they did affect the learning of the children who in the baseline were in the top 20 percent, but not those below that level (Glewwe 2009). In Sierra Leone, Sabarwal et al. find that teachers receiving textbooks when they had not regularly received them before tended to "save" the textbooks by not using them (Sabarwal, Evans et al. 2014). In Tanzania, Mbiti and Muralidharan (forthcoming) provisionally find that there is a strong interaction between teacher performance pay and textbooks—so that textbooks without added teacher incentives have no effect but when there are both textbooks and incentives learning does increase. Das et al. find that grants for textbooks are effective in the first year when unexpected but in the second the impact is fully offset by reductions in spending by parents so net textbook usage doesn't increase, funding just shifts between parents and the state (Das, Dercon et al. 2013).

Many of these interactive effects illustrate how a systemic approach can help encompass and understand the heterogeneity in empirical results. The Mbiti and Muralidharan (forthcoming) result illustrates that if there is no endogenous motivation of teachers for learning, which is the result of a "schooling only" versus "learning" system of accountability, then "more" won't work but "more and system change" might. Similarly, the fact that the textbooks in Kenya were far harder than actual fourth grader competence reveals that the overall system had not been coherent around universal high levels of learning (and Kenya does much better than its East African counterparts in measures of inputs and learning outcomes) as this mismatch, which itself reduced student learning, was tolerated. Similarly, teachers in Sierra Leone were perhaps more concerned with "process compliance" due to the typical *management* accountability than with student learning, which is not a routine part of their accountability.

Monitoring of attendance. Another example of heterogeneity of findings that illustrates the importance of coherence is the empirical research into addressing the problem of absence. In an influential paper Duflo, Hanna and Ryan showed that putting cameras into classrooms and requiring teachers to provide a date/time stamped image of their attendance both increased the

attendance of teachers and also thereby substantially increased student learning in Rajasthan India (2012). However, this experiment was carried out in schools operated by NGOs that did not receive government support and hence the *management* relationship was entirely that of the NGO. When the same researchers tried similar techniques of better monitoring of attendance of front-line service providers in the same country (India) but a different sector (health) that were *government* workers they observed very different results. In one case there was an attempt to motivate auxiliary nurse midwives through better monitoring of attendance combined with reducing pay of those whose attendance was too low. This policy actually *reduced* attendance of the ANMs (Banerjee, Duflo et al. 2008). A more recent attempt to implement biometric monitoring of primary health care clinic workers, including doctors, nurses, pharmacists found no impact on doctor attendance, a modest increase in the attendance of other workers—but the increased attendance was associated with *lower* utilization and *lower* patient perception of quality (Dhaliwal and Hanna 2013). Again, the lesson is neither that "better information on attendance combined with extrinsic incentives" works or doesn't work. It is that unless one can create policy coherence in implementation it may make things better or worse.

Teacher pay. Having a structure of compensation designed to attract, retain and motivate teachers is clearly an essential part of any education system. Yet whether or not "higher pay" for teachers does or does not improve learning depends on how it fits into an overall system. The recent experience of Indonesia is instructive. It was thought that teacher pay was too low to attract and motivate quality teaching so a reform proposal was designed that changed several elements of teacher *management*, including how teachers were to be evaluated, and that these changes would be part of a package that would include a massive increase that nearly doubled teacher wages. However at the last minute, as part of a political process, many of the key elements of the package of reform were dropped due to opposition. The increase in pay was carried out with little or no accompanying systemic reform. Since the reform was staggered over time to accommodate the massive fiscal cost this allowed some randomization of timing of which teacher got the raise which allowed a comparison of the performance of otherwise statistically identical teachers who differed only in whether they got the raise. The finding was that after two and three years there was no discernible impact on either teacher effort or student learning. So schooling costs increased by 5 billion dollars a year and the title of the evaluation suggest this was "Double for Nothing" (De Ree, Muralidharan et al. 2015).

This is not to suggest that teacher pay cannot be *part* of an overall reform that makes a system coherent for a learning objective, but if there is a relationship of accountability that is not coherent for learning one should expect just changing the level of pay to have little or no impact. India illustrates this in "perfect storm" proportions in which civil service teachers are paid five to eleven times as much as the going wage for private sector teachers or contract teachers and produce no better (and perhaps worse) learning outcomes (Murgai and Pritchett 2006).

If rigorous experiments were the hallmark of science then universities today would have departments of Alchemy. Experiments alone without an analytical frame (like a Periodic Table that classifies elements into "similar" categories in ways that predict behavior) cannot, and as we have seen in the case of education, do not, add up to anything useful to policy prescription. The systematic approach using multiple relationships of accountability is one such possible frame to help encompass and reconcile existing empirical results which can otherwise seem contradictory, paradoxical, and puzzling.

IV.C) **How can schooling have expanded so much and yet (in some cases) learning results have gotten and stayed so bad?**

We return to the tentative answer to the big question. The question is not "why have systems of education failed?" but rather a much harder question of "why is it that (some) education systems have succeeded so impressively at some aspects and yet failed so dismally at others?" That is, if governments had simply failed to pay any attention to education as a priority or devote resources to education and both the quantity and quality of education had stagnated, that would lend itself to one kind of analysis. But, by and large, that situation is rare (though there are countries that still lack primary completion and have large inequalities in schooling access and grade attainment). Rather more common is the case of many countries that have had massive, historically rapid expansions in school exposure and which thereby spend large fractions of the resources deployed by government on schooling, yet have much lower learning outcomes than are demonstrably achievable for similarly situated countries.

The tentative answer is that this can be explained by systems of education that, although proclaiming the goal of achieving uniformly high levels of learning (competencies, skills, values, knowledge), became, for a variety of reasons, coherent only to the purpose of schooling. That is, the system of relationships of accountability were adequately coherent to produce continued progress in enrollment and grade attainment expansion (and concomitant expansion of teacher and non-teacher inputs both in total and, frequently, per child) but were incoherent for learning both within, between, and across elements of accountability.

This coherence for schooling but incoherence for learning allowed low levels of learning to persist in three ways.

IV.C.1) Closed non-performance oriented systems produce isomorphic mimicry as a optimal organizational strategy

In an ecosystem some animals survive by having specialized traits of either offence or defense. One strategy that is common is mimicry. Animals survive or enhance their reproductive success by effectively pretending to be something they are not. Sometimes this is camouflage, blending into the background, sometimes this is looking bigger or stronger by puffing up, sometimes this is looking like another animal that really is poisonous. For instance, the Eastern Coral snake has red, black and yellow bands and is venomous. The Scarlet Kingsnake also has red, black and yellow bands but has no poison. But the Scarlet Kingsnake can gain nearly all of the survival value of being poisonous if it just looks enough like the Eastern Coral snake. And, since the mimic doesn't actually have to devote any biological resources to producing venom, being a mimic could be even better survival value than being the real thing.

Organizational theorists (e.g. DiMaggio and Powell 1983) postulate that organizations often adopt mimicry or "isomorphism" as a legitimacy and survival strategy. One strategy for organizational legitimacy could be to produce consistently excellent results. But this is actually a risky strategy if key stakeholders or principals do not actually measure results or if there is ambiguity or even conflict over what "excellent results" are across different principals. In this environment isomorphic mimicry, that is adopting the forms of other organizations known or agreed to be successful, can produce organizational legitimacy even in the absence of producing the actual function of those organizations.

The question becomes a question about systems. What are the characteristics of systems in which isomorphic mimicry is a successful strategy? For instance, in a natural ecosystem mimicry only works if (a) there is something to mimic (pretending to be a poisonous snake only works if there is a poisonous snake that others recognize as such) and (b) the others that need to be fooled can only rarely, or with difficulty, tell the difference. Pritchett, Woolcock and Andrews (2010) argue that for many public sector organizations isomorphism is a viable organizational strategy because two conditions exist. One, systems are "closed" rather than "open" to novelty in that it is hard for new entrants and/or new ideas to emerge. Two, novelty is judged on "agenda conformity" rather than "enhanced functionality." Think of police forces. Police forces around the world tend to look quite similar around the world. Similar uniforms, similar hierarchical rank structures, similar divisional structures, similar processes. Even in the most corrupt and dysfunctional police forces in the world most policemen wear uniforms. If there can be only one official police force (not counting private security firms) and if there is no clear functional measurement of the efficacy of the police force, only measures of process compliance (at best) then rotten police forces can persist year after year after year. Innovations that threaten the status quo, particularly innovations that are "disruptive" in the sense of Christensen (1997), can be easily thwarted as there is not an agreed-upon and reliably measured standard of "enhanced functionality" with which these innovations could prove themselves and hence acquire legitimacy and survival value.

Figure 2: When do ecosystems produce isomorphic mimicry as an optimal organizational strategy?
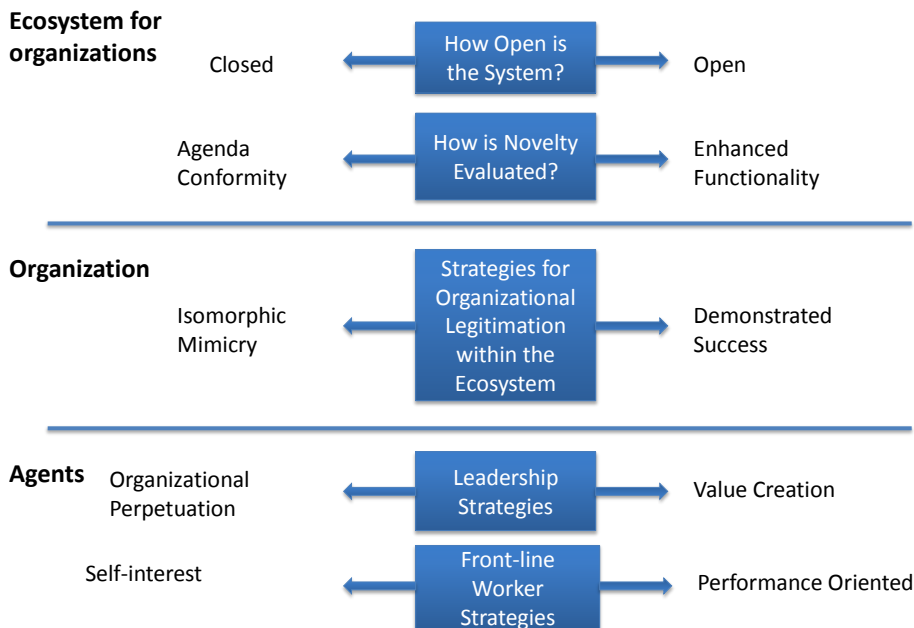
Source: Pritchett, Woolcock, Andrews 2010.

From snakes to police to schooling. The sociologist John Meyers (with others) has proposed that the best explanation for the near universality of the pace and magnitude of the expansion of schooling is a form of global mimicry (Boli, Ramirez et al. 1985), (Meyer, Ramirez et al. 1992). This mimicry took the form that the proof of a nation-state as a "modern" entity was an expansion of public schooling, whether or not this actually fulfilled any local functional needs. If the drive for expansion of schooling was itself mimicry then no measurement of the functionality of schooling as an expansion of learning of students was necessary. However, in the absence of any clear specification or relationship of accountability that strongly related to learning then all other decisions about the structures of the education system were, in a sense, up in the air. That is, if "what leads to enhanced functionality?" could not be used to resolve questions about features of the schooling system, such as, what kind of training do teachers need? All questions could be answered by mimicry: "what makes us look like a modern system?" If highly functional education systems have teachers with bachelor's or master's degrees then expansion of teachers with higher degrees becomes a mark of a "better" system, quite independently of any functional role this might play. With the heterogeneity point discussed above, it is clear I am not saying that in a system coherent for learning better-qualified teachers would not lead to better learning results; they may. But adding more gas to a car that runs will demonstrably allow the car to go more miles. Adding more gas to car with no transmission or wheels will have no impact.

Once the system has become predominantly isomorphic both around expansion of exposure (e.g. more "modern" or "developed" or "good" countries have children with more schooling) and around the actual organization of schooling itself (e.g. "better" school systems have smaller classes and trained teachers), while lacking any system of articulation of clear goals and standards on which enhanced functionality can be regularly and reliably measured, then learning can get completely bad (at least in pockets). This occurs as learning becomes strangely disarticulated from the internal (inside the *compact* and *management* relationships of accountability) legitimation of the system itself.

This, in particular, explains the persistence of good and bad schools inside the same system (often within the same locality). As Crouch and Healey (1997) pointed out two decades ago, part of the deep puzzle to be explained is why in the same city and even in the same neighborhood very good schools and very bad schools exist, and yet there appears to be little pressure for the latter to become the former (or exit). The point is that even in circumstances in which government schools are generally quite weak, not all the government schools are weak. As the LEAPS study in Pakistan (Andrabi, Das et al. 2007) demonstrates, the best government schools are often quite similar in learning quality to the best of private schools (in some places with "selective" public schools (particularly secondary) the "heritage" public schools are still among the best) but that the worst public schools are far worse than the worst private. Pritchett and Viarengo (2013) find a similar pattern among some (but by no means all) developing countries.

IV.C.2) Purpose is an organization's immune system

The wonder is that human beings don't die. That is, often survival is treated as the default and people investigate the cause of death. But what keeps humans alive in a threat-rich environment is an immune system that is dynamic, flexible, and self-regulating. Without an immune system, a person dies of some specific disease but that is only a superficially proximate explanation of the cause. The cause was a weak immune system, and if the end hadn't of been from that disease it would have been from another.

Similarly, there are many, many ways in which organizations—private or public—can become dysfunctional. Public sector organizations can be used by politicians for a variety of purposes from patronage hiring to corruption in contracting. The employees of the organization can use the organization in a variety of ways for their own purposes again ranging from outright corruption of extracting pay for no work at all ("ghost workers") to payment for little or no effort ("zombie workers") to just an "easy life." What is the equivalent of the dynamic, flexible, and self-regulating adaptive system that keeps organizations functional in the face of the many forms of decrepitude? One conjecture is "purpose," that a major factor in the *motivation* of the front-line workers is that they internalize the goals and objectives of the *delegation* of the organization. That is, they believe in and receive strong benefit from contributing to the "mission" of the organization. This helps maintain strong internalized norms of action for those within an organization.

If an organization loses a sense of coherent purpose, or if the purpose ceases to be deeply internalized as a norm by the vast majority of the members then an organization can slip past a "tipping point" and norms are no longer the norm. For instance, if achievable absence rates are on the order of 5 to 7 percent and the actual absences are, say, 10 percent, then it is clear that some are cheating on a clear norm. But if absence rates are, as they have been documented to be in many countries now, above 20 percent and classroom presence rates are only 50 percent then the "norms" of attendance and effort are likely no longer internalized norms of the typical organizational member.

This is a different explanation from arguing that the techniques of managerial control—such as strict monitoring of behavior and punishment of deviance—have broken down. These techniques can only work either in cases where mere compliance is all that is needed and in which the basics are internalized and deviance is truly deviant.

IV.C.3) Resistance to "effervescent" innovation

That final challenge is to explain not only how things get bad, but how they stay bad. That is, why are there no feedback loops in the system that push a failing or flailing system back towards functionality. Once education systems have settled into incoherence for learning outcomes and coherence only for schooling then the kind of incremental repair or tinkering that works to keep functional systems and organizations going and improving fails. As we saw above

42

there is rigorous proof that many things known to work don't work. A low-level equilibrium trap is a trap because, like a marble at the bottom of bowl, minor pushes may cause the system to change a bit but will ultimately revert to its previous position. Only major or multiple shocks can change the fundamental dynamic.

So strengthening "information" alone in ways that are incoherent with "delegation" and "motivation", for instance, may not change the overall trajectory.

This is not to say that systems won't have many innovations, often they will. But, going back to figure 2, if these innovations are not judged on functionality then the stream of innovations will not produce better results. Without a clear metric of "functionality" more innovation can just add noise.

Often dysfunctional systems will have pockets of excellence, but, like the bubbles that rise off of a glass of soda, these are effervescent. The innovations in NGOs and in parts of the public system do not expand and often wither away when their "champion" moves on. This is important as it explains the combination of apparent dynamism with stagnant or only slowly improving performance.

Conclusion

This essay attempts three things.

One, it articulates an analytic framework for describing an education system. That is, it is not merely descriptive of a system as the actors and potential functions, but also proposes that a key conceptual construct is relationships of accountability and a particular conceptual structure of *design elements* of a relationship of accountability (*delegation, finance, information, motivation)*. This leads to a 4 by 4 education system diagnostic in terms of four relationships of accountability (*politics, compact, management,* and *client power*) each with four elements.

Second, it proposes the concept of system *coherence* with three elements. Coherence *within* relationships of accountability, for instance in the *management* relationship of schools to teachers, is *delegation* coherent with *finance, information,* and *motivation.* Coherence *across* the same relationship of accountability in different relationships, for instance, is the *information* used to assess education performance coherent between *politics, compact, management* and *client power?* Coherence *across the system,* for instance, while *management* and *client power* might be internally coherent they might be incoherent as the fundamental *delegation* of management may be incoherent with the fundamental *delegation* of client power, that is, parents/students/communities might judge performance on outcomes relevant to their needs whereas *management* may entirely centered on input oriented criteria and process compliance (Banerji 2015), (Aiyar, Dongre et al. 2015).

Third, it argues that education systems' incoherence with respect to learning may account for many of the observed features of existing education systems, including the heterogeneity of the impact of specific proximate determinants (e.g. teacher pay, class size) and piecemeal reforms. Moreover, this conceptualization may provide a useful framing for a research agenda into impacts of system reform.

This is a tentative first step in formulating a research program into reforms of education systems that are capable of producing sustained accelerations in the pace of improving learning outcomes. Tentative because the description and hypotheses proposed here are not articulated as known but rather as a path into knowing.

Reference

Aiyar, Y., A. Dongre and V. Davis (2015). "Education Reforms, Bureaucracy and the Puzzles of Implementation: A case study from Bihar."

Anderson, B. (1983). Imagined communities: Reflections on the origin and spread of nationalism, Verso Books.

Andrabi, T., J. Das, A. I. Khwaja, T. Vishwanath, T. Zajonc and T. L. Team (2007). Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate. Washington DC, World Bank.

Asadullah, M. N. and N. Chaudhury (2013). "Primary schooling, student learning, and school quality in rural Bangladesh." Center for Global Development, Working Paper.

Atherton, P. and G. Kingdon (2010). "The relative effectiveness and costs of contract and regular teachers in India." CSAE WPS/2010-15.

Banerjee, A., E. Duflo and R. Glennerster (2008). "Putting a band-aid on a corpse: Incentives for Nurses in the Indian Public Health Care System." Journal of European Economic Association **6**(2-3): 487-500.

Banerji, R. (2015). "HOW DO SYSTEMS RESPOND TO DISRUPTIVE PEDAGOGIC INNOVATIONS? THE CASE OF PRATHAM IN BIHAR."

Blaug, M. (1976). "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey " Journal of Economic Literature **14**(3): 827-855.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a and J. Sandefur (2013). "Scaling up what works: Experimental evidence on external validity in Kenyan education." Center for Global Development Working Paper(321).

Boli, J., F. O. Ramirez and J. Meyer (1985). "Explaining the Origins and Expansion of Mass Education." Comparative Education Review **29**(2).

Bruner, J. S. (1996). The culture of education, Harvard University Press.

Christensen, C. (1997). The Innovator's Dilemma: The Revolutionary Book that Will Change the Way you do Business. Cambidge MA, Harvard Business School Press.

Clemens, M., C. Montenegro and L. Pritchett (2008). "The Place Premium: Wage Differences for Identical Workers." CGD Working Paper Series #148.

Crouch, L. and F. H. Healey (1997). Education Reform Support: Volume One: Overview and Bibliography. Wasington DC, USAID.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan and V. Sundararaman (2013). "School Inputs, Household Substitution, and Test Scores." American Economic Journal: Applied Economics **5**(2): 29-57.

De Ree, J., K. Muralidharan, M. Pradhan and F. Rogers (2015). "Double for nothing? Experimental evidence on the impact of an unconditional teacher salary increase on student performance in Indonesia." Unpublished manuscript. The World Bank, Washington, DC.

Dhaliwal, I. and R. Hanna (2013). Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India.

DiMaggio, P. J. and W. W. Powell (1983). "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields " American Sociological Review **48**(2 (April)): 147-160.

Dongre, A. A., A. Kapur and V. Tewary (2014). "How Much Does India Spend Per Student on Elementary Education?" Available at SSRN.

Duflo, E., P. Dupas and M. Kremer (2009). "Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya."

Duflo, E., R. Hanna and S. Ryan (2012). "Incentives Work: Getting Teachers to Come to School." American Economic Review **102**(4 (June)): 1241-1278.

Filmer, D. (2015). Service Delivery Indicators: Education Quality in Sub-Saharan Africa RISE Launch Conference, Washington DC.

Glewwe, P., M. Kremer and S. Moulin (2009). "Many Children Left Behind? Textbooks and Test Scores in Kenya." American Economic Journal: Applied Economics **1**(1): 112-135.

Glewwe, P. and K. Muralidharan (2015). "Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications."

Hsieh, C.-T. and M. Urquiola (2006). "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." Journal of Public Economics **90**(8): 1477-1503.

Mbiti, I. and K. Muralidharan (forthcoming). "Kiufunza: Improving Early Grade Learning Outcomes in Tanzania."

Meyer, J. W., F. O. Ramirez and Y. N. Soysal (1992). "World expansion of mass education, 1870-1980." Sociology of Education: 128-149.

Moore, M. (1995). Creating Public Value: Strategic Management in Govermnet. Cambridge MA, Harvard University Press.

Moore, M. (2015). "Creating Efficient, Effective, and Just Educational Systems through Multi-Sector Strategies of Reform."

Murgai, R. and L. Pritchett (2006). "Teacher Compensation: Can Decentralization to Local Bodies Take India from the Perfect Storm Through Troubled Waters to Clear Sailing?" India Policy Forum **3**(1): 123-177

North, Douglas, J. Wallis and B. Weingast (2009). Social Order: A Conceptual Framework for Interpreting Recorded Human History, Cambridge University Press.

OECD (2010). PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I). Paris, OECD.

OECD (2012). PISA 2012 Results in Focus.

Pritchett, L. (2014). "The risks to education systems from design mismatch and global isomorphism: Concepts, with examples from India." UNU-WIDER WP/2014/039.

Pritchett, L. and Y. Aiyar (2014). "Value Subtraction in Public Sector Producation: Accounting Cost and Economic Cost." Center for Global Development **Working Paper 391**.

Pritchett, L. and M. Viarengo (2013). "The State, Socialization and Private Schooling: When Will Governments Support Alternative Producers?" Harvard CID Working Paper Number 272.

Pritchett, L., M. Woolcock and M. Andrews (2010). "Capability Traps: Mechanisms of Persistent Implementation Failure." CGD Working Paper Series(234).

Sabarwal, S., D. Evans and A. Marshak (2014). "The permanent input hypothesis: the case of textbooks and (no) student learning in Sierra Leone." World Bank Policy Research Working Paper(7021).

Sandefur, J. (2015). "Inequality in Test Scores Within and Between Nations." RISE working paper.

Sandefur, J. and A. Beatty (2015). "Learning Profiles for South Asia and East Africa." Seabright, P. (2010). The Company of Strangers: A Natural History of Economic Life. Princeton NJ, Princeton University Press.

Singh, A. (2014). "Emergence and Evolution of Learning Gaps across Countries: Panel Evidence from Ethiopia, India, Peru and Vietnam." Young Lives Working Paper #124.

Team, U. G. M. R. (2014). Teaching and Learning: Achieving Quality for All, UNESCO Publishing.