Working Paper Number 43
June 2004

Towards a New Consensus for Addressing the Global Challenge of the Lack of Education

By Lant Pritchett

Abstract

This paper is part of the Copenhagen Consensus process, which aims to assess and evaluate the opportunities available to address the ten largest challenges facing the world. One of these ten challenges is the "lack of education." This paper will define "lack of education," in terms of enrollments, attainments and learning achievement. It provides an analytical framework to evaluate the various options that can be used to address this issue.

Education can be described as equipping people with the range of competencies necessary to lead productive, fulfilling lives fully integrated into their societies and communities. Many of the international goals are framed exclusively around enrollment, which is merely a means towards creating competencies and learning achievement. This paper discusses the scope and options for improving people's competencies, and describes the conditions for effective policy action.

Towards A New Consensus for Addressing the Global Challenge of the Lack of Education

Lant Pritchett[1]

Kennedy School of Government, Harvard University

And

Center for Global Development

Executive Summary

This paper is part of the Copenhagen Consensus process, which aims to assess and evaluate the opportunities available to address the ten largest challenges facing the world. One of these ten challenges is the "lack of education." As a challenge paper this paper intends to:

- Describe the scope of the global challenge of "lack of education" in enrollments, attainments, and learning achievement.
- Provide an analytical framework for assessing opportunities to address this challenge.
- Review the existing literature to produce estimates of the costs and benefits of five classes of feasible opportunities for addressing this challenge.

As grappling with this Herculean task in the limited space available leads to an extremely dense document, I want to highlight the main conclusions up front, and in particular stress how they differ from (at least some strains of) the current international conventional wisdom.

- *Scope.* Defining the scope of the problem of 'lack of education' must begin with the *objectives* of education—which is to equip people with the range of competencies (which includes both cognitive and non-cognitive skills, knowledge, attitudes) necessary to lead productive, fulfilling lives fully integrated into their societies and communities. Many of the international goals are framed exclusively as targets for universal enrollments or universal completion. But getting and keeping children "in school" is merely a means to the more fundamental objectives and while schooling is necessary, it is by no means sufficient. The challenge of the lack of education must remain focused on creating competencies and learning achievement.
- *Framework.* The analytic framework for evaluating options must consider both the demand for education by parents and children and the supply of educational services.
    - *Demand.* The choices made by parents and children about schooling are based on objectives, constraints and information. Since the vast majority of those not receiving adequate education are children from poor households, it has to be acknowledged that their choices are difficult ones that trade off additional spending on necessities and

[1] With assistance from Michal Shwarzman. I would like to thank, without implication, Barbara Bruns, Michael Clemens, Alex Schlegel, and Miguel Urquiola for comments. Also I would like to thank the Copenhagen opponents, Paul Schultz and Ludger Woessmann for their insightful critiques. I would like to particularly thank Luis Crouch to whom, as usual, I am indebted for seeing clearer and writing clearer than I dared.

additional help and labor from children in needed tasks versus more spending on and time devoted to education today. Their choices are also based on the information they have about the benefits to education generally; the abilities, aptitudes and preferences of their children, the quality of the available schooling options.

- *Supply.* A key element that is missing from nearly all existing analysis of the lack of education is a coherent, general, positive theory of government action. Typically in policy discussions and international forums it is simply assumed that the variety of *normative* reasons why a government *ought* be involved in schooling (e.g. "externalities" or credit constraints or a concern for inequality) are actual the *positive* explanation of government action. This is simply untenable as a theoretically or empirically adequate explanation of what governments actually do. Assuming perfect governments is as unrealistic as assuming perfect markets. Some plausible positive model of government action is necessary to answer the question of what is an "opportunity"—a realistic discussion of opportunities must address not just the question of *what* should be done but also *who* should do it and *why* it is not already being done.
- Teachers are the core of any system of schooling. Learning requires millions upon millions of individual interactions between teachers and students every day. These interactions are effective in generating learning achievement when teachers:
  - Know what it is they are trying to accomplish,
  - Have command over the necessary knowledge themselves,
  - Are competent in at least one (and preferably multiple) ways of teaching,
  - Are motivated to perform,
  - Have adequate facilities and instructional resources with which to teach.

  Any discussion of the supply of education has to have a positive model of how it will do this.

- *Opportunity 1: Physical expansion.* A quantitative expansion of existing school facilities is far from sufficient for addressing the lack of education—and in the majority of country cases will have almost no impact at all. Nearly all of the global lack of education is due to (a) children not attending *available* schools, (b) children dropping out of available schools too early, and (c) low levels of learning achievement while in school. While there are certainly situations in which school availability is a key constraint (e.g. sparsely population rural areas, junior secondary in some countries) system expansion ("building more schools") as an opportunity for addressing the lack of education tends to be vastly over-rated.
- *Opportunity 2: Improve quality*
  - *Radial expansion in budgets.* There is almost no evidence that "more of the same" or "business as usual" expansions in expenditures per pupil will be capable of improving learning achievement. This is not to say that increases in expenditure *cannot* improve quality or that increases in expenditure won't be necessary, but only that there are many ways in which expenditures have increased without improvements in learning achievement. Expanding the budget is not *the* solution—it is only a part of a broader

solution—and more money in isolation in systems with entrenched and pervasive problems more resources may even make matters worse.

- *Expand specific interventions.* While nearly everyone engaged in the debate about schooling acknowledges that more resources is not a panacea—the question is whether budgets can be expanded on specific items in a way that would increase the efficacy of schooling.
  - *Known.* Nearly every empirical study in developing countries that can establish any relationship between schooling inputs and outputs finds that, starting from the existing pattern of expenditures, more spending on elements of infrastructure, recurrent inputs, and learning materials are more cost-effective that raising teacher wages or reducing class sizes. But whether or not sustained changes in the composition of expenditures is feasible in the absence of systemic reforms and within a reasonable total budget is an open question.
- *Techniques to be discovered.* A second possibility is that there exist techniques or interventions that would substantially improve learning but that these are not known because there has been too little rigorous evaluations of education options. Expanding the scope of rigorous evaluation is certainly a key opportunity because the since the current level is so low the marginal returns to additional research are high. However, as with selective budget expansions, the problem is how improved knowledge will translate into action.

- *Opportunity 3: Expand demand for schooling through increased income or raising returns.*
  - Since parental income and education are key determinants of child enrollment, expanding education has a cumulative effect which leads to further improvements. Increases in per capita incomes will typically increase schooling. This is only worth stressing because, while education sector specific solutions are often what is under direct policy consideration, government actions that improve living standards generally are almost certainly the surest route to expanded educational attainment.
  - Even more particularly, government policies that increase the returns to schooling will be key to raising the demand for education. It is plausible that one of the major reasons for the low (and declining) enrollments in many parts of the world is that a stagnant economy has meant low and declining returns to education.
- *Opportunity 4: Reductions in the cost of schooling to increase demand.* There is no question that educational choices respond to costs and three types of actions have been shown to increase enrollments: (a) reductions in fees, (b) either targeted or non-targeted cash or non-cash inducements to attend school (e.g. school lunches) and (c) making cash transfers from poverty reduction programs conditional on school attendance (e.g. PROGRESA in Mexico). The key question is whether these programs are cost-effective, which depends critically on their design characteristics—in particular the extent to which the program can be targeted to avoid "infra-marginal" transfers. If infra-marginal benefits—those going to students who could have enrolled without the transfer--are high then these programs cannot be justified on their educational expansion benefits alone.

- *Creating the conditions for effective policy action: Systemic school reform to increase accountability.* After evaluating the four policy action opportunities I will suggest that in many cases the key constraint is that the problem in many countries with low levels of education is the way in which the production of schooling is organized—the relationships of accountability for performance simply do not exist. Currently schooling systems are entirely *input oriented* and systemic reform is necessary to create a *performance orientation.*

  The key opportunity for remedying the lack of education in the world are actions that simultaneously create:

  - *Clear objectives for the schooling system,*
  - *Adequate financing to achieve those objectives,*
  - *Greater producer autonomy to manage for results,*
  - *Greater producer accountability for results which implies a need for transparent, consistent, measures of progress towards the objectives.*

  There are many ways to do this (in particular, one does not need "markets" or "vouchers"). All of the innovative reform agendas: p*ublic sector reforms: school autonomy (e.g. 'charter' schools)*, *"Community" control*, *Use of non-government providers*, *Market based ("voucher") programs*, *Public sector reforms: political decentralization* are ways of changing the system to produce these objectives.

  A key difference is between those who believe that the lack of education can be addressed with only "policy action"—actions taken within more or less the existing structures for organizing the production of schooling—and those who believe "systemic reform" is *the* one opportunity that rules them all.

- *Current initiatives*: *EFA/FTI.* The EFA/FTI is what there is, and it is the best there is. The key will be to keep the innovative components and its general well designed structure from being overwhelmed by the tendency to revert to an "input oriented" and "nation-state centric" approach. That is, EFA/FTI can be a force for building elements of output orientation, performance measurement, autonomy and accountability into schooling systems. It will take enormous force to resist the usual pressures to lapse into the donor driven "business as usual" of funding more inputs.

A final two explanatory notes about this document. First, this is not a "review of the literature" in the usual academic sense that the document will describe the literature, the results, the differing methodologies. There simply is not space to cover the broad agenda in this depth. Rather after reviewing the literature I try and present tables and figures which best convey the key general findings and results—in many instances one could have chosen any number of examples or papers that have the same or similar findings. So the tables and figures I present are typically not the evidence for the arguments I am making they convey the sense of the evidence for the arguments that I make.

Second, since many of the readers of this document will have no direct experience living or working in the poorer countries of the world in which the lack of education is most pressing I would ask them to abandon now any assumptions based on their personal experience. Citizens of the developing world cannot take for granted what every resident of a rich industrial country

does. In particular, many of the same words are used in discussions—but they convey problems of different orders of magnitude. When "poor pedagogical approaches" are discussed it is usually taken for granted the children are not being hit with by the teacher with a stick—not always so. When "under motivated" teachers it is usually assumed they at least show up—not always so. When "poor facilities" are discussed it is usually assumed there is at least a classroom—not always so. When "political patronage" is discussed it is usually assumed that teaching positions are not openly sold as patronage—not always so. "Low quality teachers" is not assumed to *not* mean "cannot read and write at a fourth grade level"—not always so. The challenge of the lack of education is in the rural areas of India, China, Indonesia, Nigeria, Pakistan, Bangladesh (to name the largest developing countries) and it is the stark reality that parents, children and teachers in these countries grapple with day in and day out that needs to be addressed.

Part I) The Scope of the Challenge and a Framework

The "lack of education" as a global challenge must be understood as a failure to master the many distinct competencies—decoding, cognitive skills, factual information, socialization—necessary to thrive in a modern economy and society[2]. Basic schooling is a means to acquiring the skills and competencies that contribute to a fuller, more humane, and productive human existence[3]. Remedying a "lack of education" implies that each individual have *sufficient exposure to learning opportunities to achieve mastery of the basic competencies needed in his/her society and economy*[4]. What level of schooling is "sufficient" will vary: some countries may consider completion of a five year cycle of primary schooling sufficient, others a longer primary cycle, still others have extended basic to include primary plus a cycle of junior secondary. I will use the term 'basic' schooling to encompass any of those definitions.

This first section examines enrollment, grade attainment, and learning to document the lack of education, with a special emphasis on the low learning achievement of those in school.

I.A) Enrollment

---

[2] The paper will not review the justification or benefits of quality basic education (e.g. growth externalities, fertility reduction benefits, enhanced democracy, reduced crime). It will be assumed that the lack of quality basic education is a major social challenge and a legitimate social objective. In particular, I will not assume that there is or is not any particular "externality" or "public good" aspect to schooling.

[3] Tertiary education will not be covered except insofar as it relates to basic education. While there are legitimate arguments that tertiary education is increasingly important as the scope of the market expands and as the market for talent is increasingly international, there are three justifications for not addressing tertiary education. First, tertiary presupposes quality basic, as unless a sufficient number of high quality prospects emerge the tertiary system cannot function. Second, Heckman and Carneiro (2003) have demonstrated that, in the USA, public sector interventions at early stages of learning have much higher payoffs than either educational interventions at later stages (e.g. college financing) or direct labor market interventions (e.g. job training). Third, pragmatism—I could either do a challenge paper on tertiary education or on basic—but the issues are sufficiently distinct that attempting to cover both in the space and time available is impossible.

[4] For instance, the Millennium Development Goals (MDG) target for the universal completion of a primary cycle cannot be taken literally as a goal that children sit in a building called a "school" for a given number of years. The goals of Education For All include a stipulation that each child complete a primary cycle of "good quality" (UNESCO 2002).

Official data on *enrollment* rates in primary schooling show a varying magnitude of "out of school" children across countries and regions of the world. Sub-Saharan Africa lags all other regions of the world with only 56 percent of primary school-aged children enrolled in school. South Asia and the Middle East/North Africa also have well less than 100 percent net enrollment ratios.

The nearly exclusive use of enrollment data in discussions of education is unfortunate as these data mask three key features—(a) the difference between gaps in *attainment* that are due to "never enrollment" versus "drop-out" (b) the *gaps* in educational attainment by categories other than gender, and (c) the actual *learning achievement.* The next three sub-sections address these in turn.

Table 1: While gross enrollments rates are high nearly everywhere, universal* net *primary enrollment is more rare

(Estimates of Primary School Enrollment and Repetition by Regions)

Area	*Gross enrollment 2000*	*Net enrollment 2000*	*Repetition 2000*	*On-time enrollment 2000*
Low-income	102	85	4	55
Middle-income	110	88	10	61
High-income	102	95	2[a]	73[b]
Region				
Sub-Saharan Africa	77	56	13	30
South Asia	98	83	5	-
Middle East/North Africa	97	84	8	64
East Asia	111	93	2	56
Latin America	127	97	12	74
East Europe/FSU	100	88	1	67[a]
OECD	102	97	2[a]	91[a]

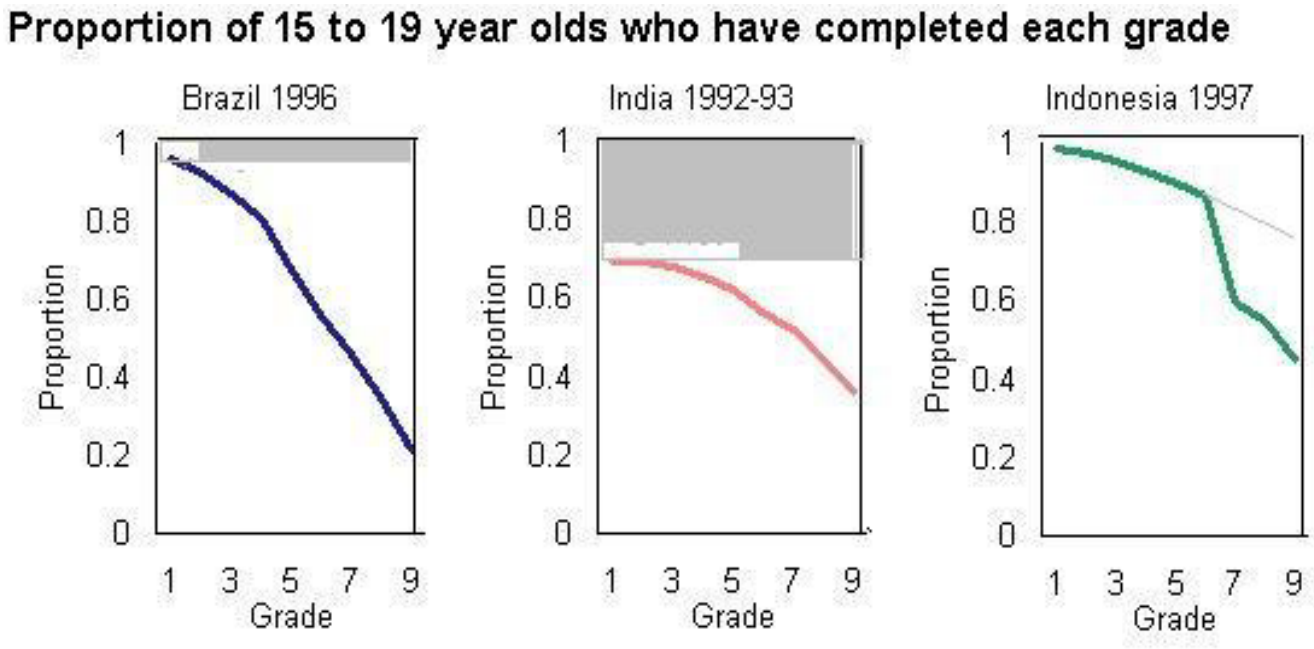
Notes: Countries with populations of less than 1 million are excluded.
a. Data are based on between 25 percent and 50 percent of the total population of the country group or region.
b. Data are based on between 10 percent and 25 percent of the total population of the country group or region.
Source: Glewwe and Kremer (2004). "Schools, Teachers, and Education Outcomes in Developing Countries" forthcoming in Handbook on the Economics of Education.

I.B) Attainment profiles and their patterns

The Demographic and Health Surveys (DHS) are a comparable set of over 50 nationally representative household surveys that have been carried out in a large number of developing countries. The advantage of these household surveys over administrative data on enrollments is three-fold. First, they have self-reported enrollment which does not rely on administrative data which can often over-state actual enrollment. Second, they have data on grade attainment—the highest grade completed—which is the result of enrollment, repetition, etc. Third, they can be linked with household characteristics such as wealth and residence to give a clearer picture of socio-economic differences in education.

Patterns in attainment. The data on attainment from the DHS can be used to generate "attainment profiles" as in Figure 1, which show the fraction of the cohort aged 15-19[5] which has completed a given grade or higher[6]. The fraction of children who did not complete even grade 1—those who "never enrolled"—is shown as the gap between 1 and the "intercept" of the graph. The slope of the attainment profile shows the fraction of children who do not progress across any given year. This includes both "drop-out" within a level of schooling (e.g. primary), including drop-out after repeating a grade, and not making a transition to the next level of schooling (e.g. primary to junior secondary). If the goal is universal completion through some specified grade (say, grade 9), then the total deficit in school years attended relative to the goal is the trapezoidal area.

Figure 1: Attainment profiles reveal different patterns of enrollment and progress as differing amounts of the deficit from universal completion are due to children who never enroll (e.g. India) children who abandon schooling within a cycle (e.g. Brazil) and students who do not progress across a cycle (e.g. transition to junior secondary in Indonesia).



The attainment profiles allow a decomposition of the problem of low *grade attainment* into that due to "never enrollment" and that due to "failing to complete a minimal level conditional on enrollment" and show at least three distinct patterns represented by the three countries shown:

- Brazil (cohort aged 15-19 in 1996) shows the pattern of high initial enrollment with high drop-out both within and across levels of schooling. Nearly all of the deficit from grade 9 completion is due to those who once enrolled but did not persist.

[5] The ages 15-19 are chosen to balance the relevance of the information to current conditions (the older the cohort the longer ago educational decisions were taken) versus the censoring problems of those still in school. One can produce estimates adjusted for the censoring—all of the basic patterns observed are the same.

[6] All of the attainment profile figures come directly from a World Bank web site maintained by Deon Filmer that contains over 90 http://www.worldbank.org/research/projects/edattain/edattain.htm

- India (1992-1993) shows the pattern of low ever enrollment. Of this cohort, only about 65 percent ever enrolled in schooling—but of those that did enroll, most persisted. Astoundingly, grade 9 completion was *higher* in this period in India than in Brazil.
- Indonesia shows that pattern of high initial enrollment, persistence through primary school, which a large drop-off in enrollment across the transition from primary to junior secondary.

Table 2 shows the relative role of "drop-out" versus "never enrollment" in explaining the deficit from universal completion of basic schooling (defined as either grade 5 or grade 9) across various regions of the world. The very poorest countries in West Africa and South Asia show a substantial fraction of children who never enroll in school and hence, a large fraction of the deficit from even grade 5 completion is due to "never enrollment." In contrast, even in quite poor regions, like Central America, East and Southern Africa, and East Asia (this sample does *not* include the richer "tigers") almost half of the deficit from completion of grade 5 is due to children who begin schooling but then drop-out before completing even grade 5.

Table 2: "Drop-out" within the primary cycle (prior to grade 5) typically accounts for more of the deficit from universal primary completion that does "never enrollment"—and even more so for "basic" (grade 9) completion

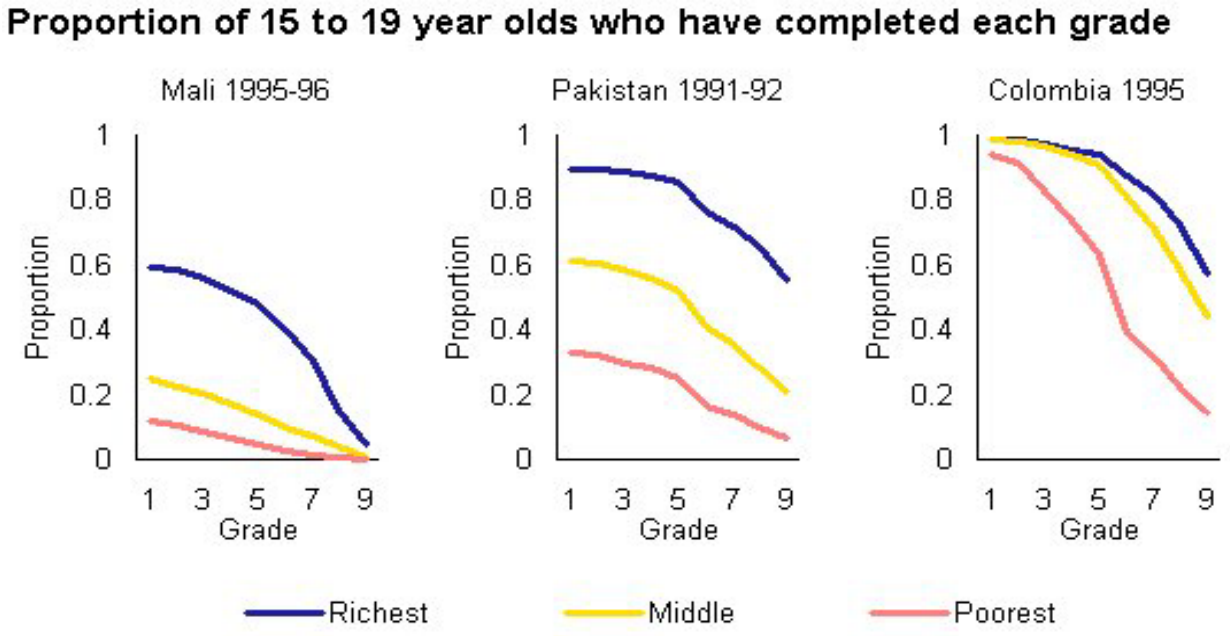
(Decomposition of the total years of schooling attainment deficit from either universal grade 5 or grade 9 completion into the fraction due to children "never enrolled" and those who "drop-out")

Region (unweighted average of available countries)	Never Enrolled	Completing Grade 5	Deficit from universal Grade 5 completion due to "drop out"	Completing Grade 9	Deficit from universal Grade 9 completion due to "drop out"
West and Central Africa	43.7%	41.6%	15.3%	13.3%	33.4%
East and South Africa	17.1%	55.7%	46.8%	12.9%	69.7%
South Asia	27.9%	59.7%	16.4%	27.0%	39.5%
Central America	9.7%	65.6%	52.9%	25.3%	74.7%
MENA	17.8%	69.8%	21.8%	30.8%	57.1%
East Asia and the Pacific	9.1%	73.5%	53.2%	30.5%	78.5%
South America	1.9%	83.4%	76.3%	42.9%	92.0%
Europe and Central Asia	0.5%	99.2%	24.1%	83.3%	79.6%
Representative countries from the above regions					
Nigeria	21.5%	73.1%	8.5%	37.8%	37.2%
Kenya	3.5%	84.3%	54.5%	18.9%	87.4%
India	21.6%	69.6%	13.7%	39.0%	38.6%
Guatemala	14.1%	57.4%	48.5%	23.6%	68.0%
Turkey	4.4%	92.2%	25.7%	40.0%	82.8%
Phillipines	1.4%	92.2%	63.0%	52.4%	90.6%
Brazil	3.8%	67.8%	75.2%	21.6%	89.2%
Kazakhstan	0.5%	99.1%	32.4%	92.7%	72.6%
Source: Based on Filmer (2004) and Filmer and Pritchett (1998).					

Socio-economic characteristics and attainment. The enrollment data can reveal differences in enrollment by residence and by sex. Without links to the household, however, the enrollment data cannot provide evidence on the attainment differentials by the socio-economic condition of the households. Filmer and Pritchett (2000) have demonstrated that questions in the DHS on asset ownership (e.g. does the HH own a radio, bicycle) and housing characteristics (e.g. does the household have a separate kitchen, toilet facilities) are adequate for the construction of an asset index as a proxy for long-run household wealth. As Figure 2 indicates, there are large gaps in attainment between children from richer and poorer households.

Figure 2: The gaps in attainment between richer and poorer households are substantial in every country of the world, sometimes driven by low ever enrollment of poorer children (e.g. Mali, Pakistan) and sometimes by much higher drop-out rates (e.g. Colombia)

(Attainment profiles of the poorest 40%, middle 40%, and richest 20% ranked by an index of household wealth based on asset ownership and housing characteristics)

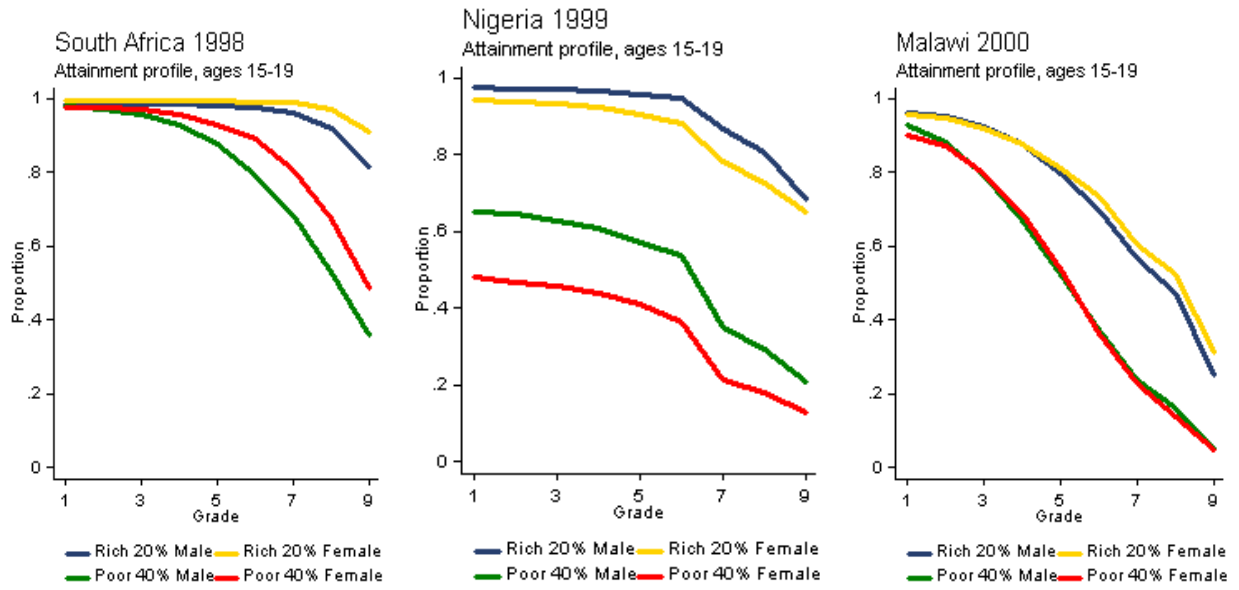


In addition to the gaps by household wealth it is also the case that, conditional on wealth, children with more educated parents are more likely to be enrolled and that rural children are less likely to be enrolled.

In some (though not all) regions of the world, girls are also significantly less likely than boys to attend school and to complete basic education[7]. Figure 3 shows the attainment profiles from three countries in Sub-Saharan Africa which show how varied the experiences are as all three show different attainment profiles, different wealth gaps, and different gender patterns. In South Africa it is boys who drop-out much earlier than girls, in Nigeria both poor and rich girls are less likely to enroll but once enrolled their attainment profiles are similar, in Malawi the patterns of boys and girls are nearly identical. In spite of the tremendous importance of gender issues in society in general and within educational systems in particular, in this challenge paper I will not address gender issues in any depth (except in the section on demand side transfers) because I have very tight space constraints and no value to add to the already enormous literature on this issue.

[7] It is worth noting that this appears to be more a *regional* that a *country* issue—that is the gender gap in education within India varies as much across states as the gender gap varies across countries in the world and within regions within other countries (e.g. Sahelian countries) the gender gap varies enormously. It is also worth nothing that there are many countries and regions of the world in which there either is no gender gap or girls receive more schooling than boys.

Figure 3: Three countries from Sub-Saharan Africa illustrating the different combinations of enrollment profiles, wealth gaps, and gender differences in attainment (Attainment profile by wealth and gender)



I.C) Low levels of learning achievement

The completion of primary schooling or higher in itself, however, does not guarantee that a child has mastered the needed skills and competencies. In fact, all of the available evidence suggests that in nearly all developing countries the levels of learning achievement are strikingly, abysmally, low.

Figure 4 shows the results of all the available recent (since 1984) internationally comparable assessments of achievement in reading, mathematics, and science. The examinations have often used widely different approaches to testing and measuring learning achievement and so the figure displays the both the regional distribution and the results separately for each participating country on each of the nine examinations. The figure shows each country's average score, for each assessment, scaled by calculating the standard deviation of scores across all participating OECD countries. Consequently, the figure reports how many "OECD standard deviations" each country was above or below the OECD mean. In addition a box-plot shows the regional average, 25th, 75th, 90th and 10th percentile by that same metric. These examinations have all been schooling population-based and are cover populations aged 9-10 or 13-15. Obviously, by testing only children in school these results *understate both* the differences in *population* achievement – between countries where schooling is universal at these ages, and countries where not all children are still enrolled – and the differences in school quality.

Except for two "regions"—Eastern Europe and FSU, and the four "Tigers" in East Asia—the developing countries lag far behind the OECD in learning achievement. Latin American countries are 4.8 OECD standard deviations below the OECD median, countries in Sub-Saharan Africa 6.7, countries in the Middle East and North Africa 4.7, and countries in East Asia 3.1. These astoundingly large differences imply that the developing countries are not just

the lower tail of the OECD, but have performance far below the *poorest* performing OECD countries. For instance, on the recent PISA Greece, which was the poorest performing major OECD country, with an average score of 447, scored roughly two OECD std. deviations (of 31 points) below the OECD mean of 505. Most of the non-Tiger non-FSU/EE developing countries are well below the *worst* OECD country. Thailand at 433 may not be so far behind, but Argentina (388), Mexico (387) and Chile (384) are roughly 60 points (two OECD std. deviations) behind, Indonesia at 367 is 80 points behind, and Brazil at 334 is 113 points behind. Peru at 292 is an astounding 155 points—**5** OECD std. deviations — below Greece, the worst performing OECD country, and 265 points behind Japan[8].

These large differences in country average performance translate into large differences in performance at the top and bottom end of the spectrum. For each of the three subject areas in the recent PISA—readings, mathematics, and science—Figure 4 compares the distribution of student scores between the typical (median) OECD country and a developing country. In assessments of reading competence only 3.1% of Indonesian students scored above the *average* French student. Conversely, the average Indonesian student performed at the 7th percentile of French performance. In Mathematics only 3.2% of Brazilian students could outperform the average Danish student while the average Brazilian student would find himself at the same level of only 2% of Danes.

In Science, the poor and lagging performance of US students has been a major concern of educators and policy makers. The fact that US students lag behind Japanese and Korean students has raised serious concerns about the nation's competitive position and ability to maintain a technological lead and has mobilized efforts at increasing expenditures and a variety of education reforms. But according to the PISA results, the gap between US (499) and Peruvian (333) students, is more than *three times as large* as the gap between US students and their Japanese counter-parts. Only 3.3% of Peruvian students can perform at or above even the average the performance of US students, which is generally regarded as unimpressive, at best.

[8] The low level of achievement is perhaps illustrated by a concrete question. In the TIMSS, the question was asked: "Three-fifths of the students in a class are girls. If 5 girls and 5 boys are added to the class, which statement is true of the class?" (a) There are more girls than boys, (b) There are the same number of girls are there are boys, (c) There are more boys than girls, (d) You cannot tell whether there are more girls or boys from the information given." The correct answer is (a). One would be tempted to say "obviously". But while 82 percent of Japanese eighth graders and 62 percent of USA eighth grades answered this question correctly, only 31 percent of South African and only 30 percent of Colombian students answered this question correctly—when random guessing would produce 25 percent correct on average!

Figure 4: Performance on internationally comparable assessments of learning achievement in math, reading, and science show most developing countries strikingly behind the OECD *(Performance on all existing international comparisons (IEA, SISS, TIMSS, PISA, PIRL) scaled as "OECD cross national standard deviations below the OECD mean")*

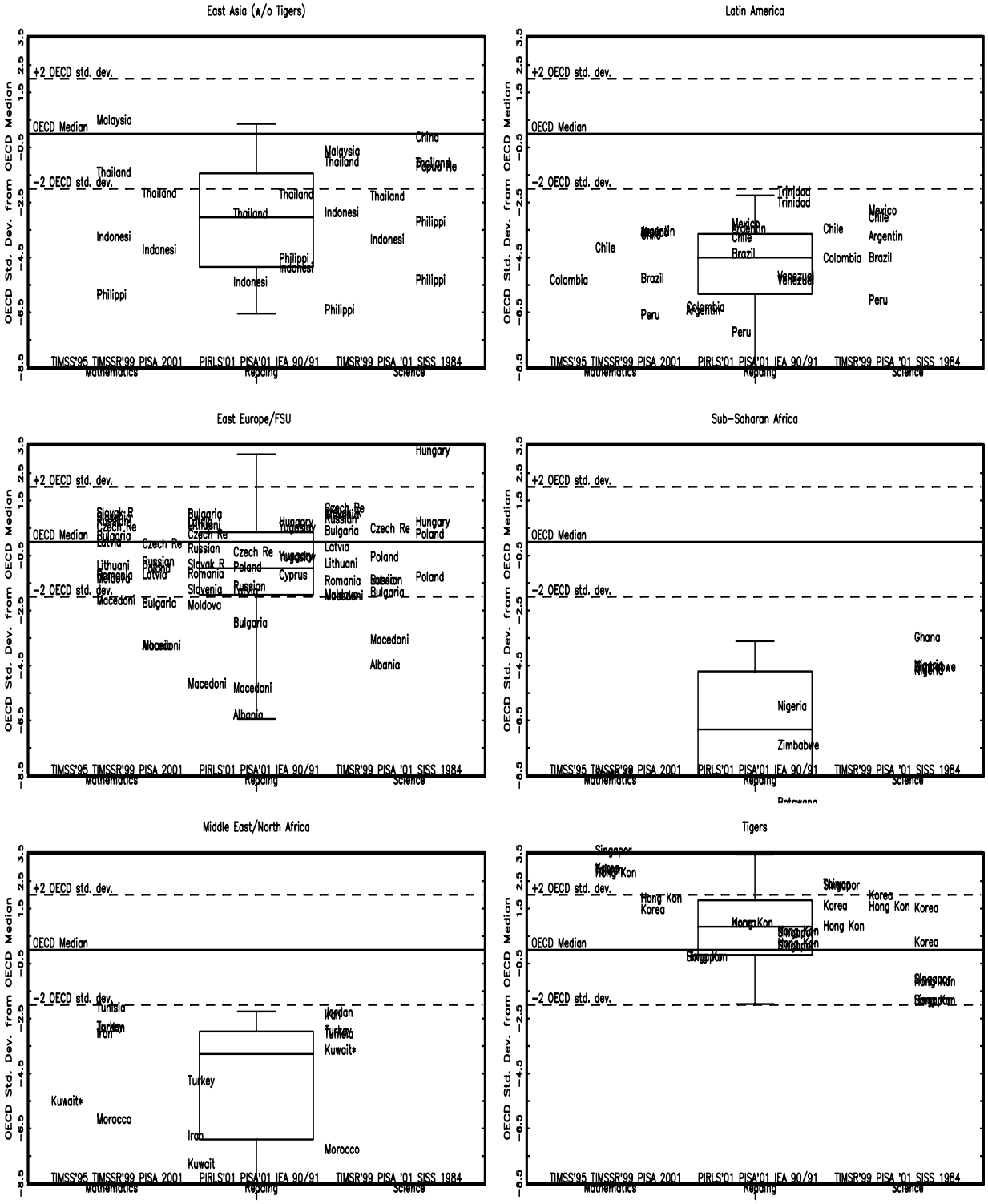
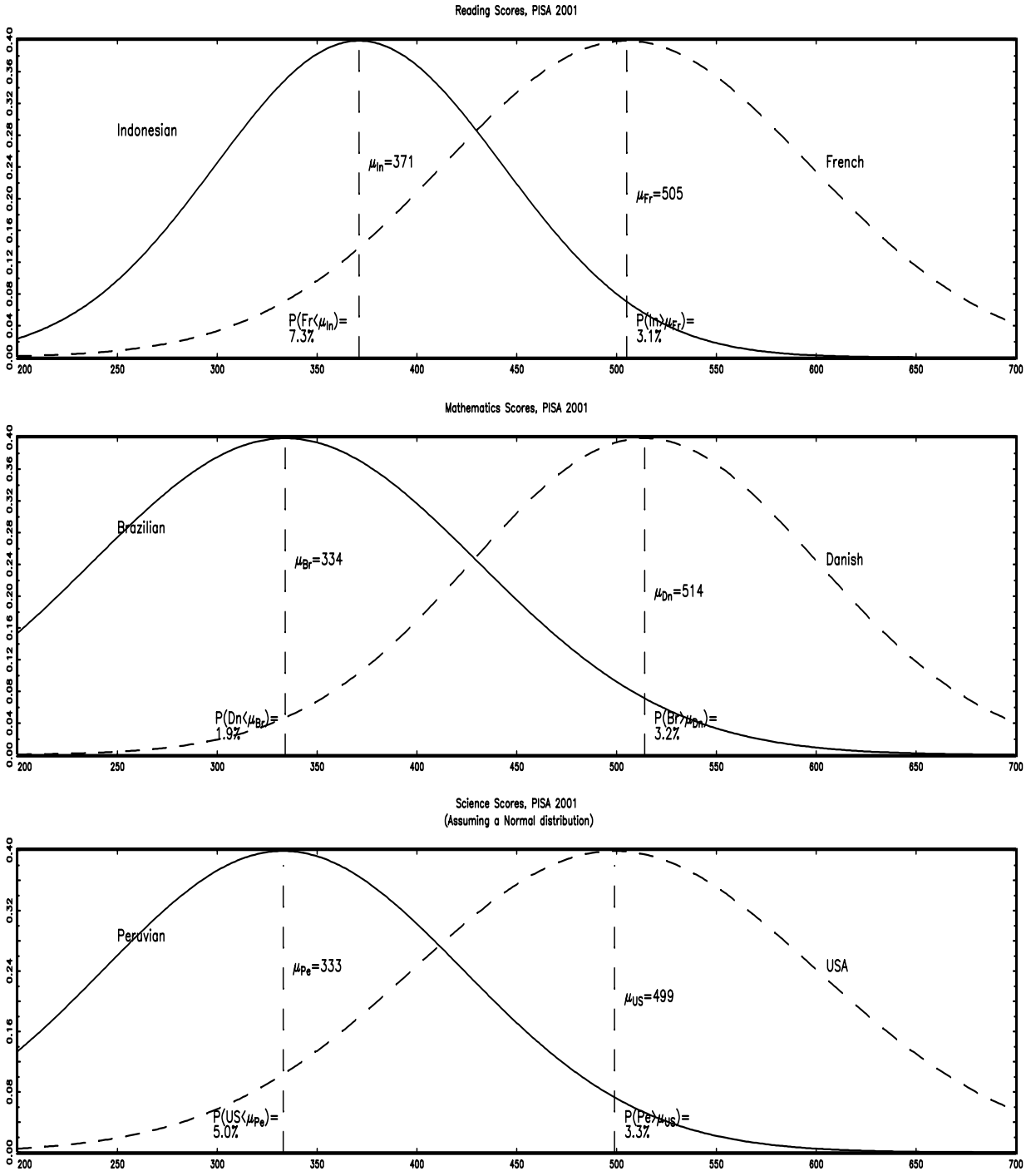


Figure 5: Low average country scores imply that the average developing country student ranks very low in a typical OECD country distribution and that even the best students rank only at about the average OECD student

(Comparisons of two countries distributions of scores for math, reading, and science showing the fraction of students from the OECD below the developing country and the fraction of students from the developing country above the OECD country mean)



There are many reasons not to rely exclusively on the international examinations. One might believe that these paint an unfairly negative picture of developing country student performance because the examinations are not based on the curriculum covered, or because "Western" students are more adept at these type of examinations, or because the exams are "biased" in one way or another. However, since participation in international assessments is voluntary (based on interest on the part of the country) and requires substantial technical capacity to implement, these results may *overstate* typical developing country performance if the learning achievement in countries that do participate is higher than in those that do not.

The available evidence from other types of assessments of student learning—school leaver examinations, individual assessments, and tests done for research—is patchy and hard to compare, but is consistent with very low levels of learning. In an assessment of competencies in Bangladesh, Greaney et al. (1998) found that two-thirds of those who had completed primary school failed to achieve the minimal level. School leaver examinations in many African countries (which rarely appear in international assessments) suggest that relative to the curriculum objectives learning achievement is low. Finally, a common problem in research into the determinants of learning outcomes done by economists is that the original tests designed to test student competence are designed based on the curriculum. However, when representative samples of students are given these "age and grade appropriate" examinations, the results cluster around scores that are consistent with random guessing (making research difficult)[9].

Table 3: Examples of low test performance of children — even after years of schooling

Study	Country/ Year	Age/ Grade	Description of test	Findings
Greaney et al.	Bangladesh 1998	Age 11+ in rural areas	A test of basic learning skills in reading, writing, written mathematics and oral. A professional panel specified minimal acceptable levels of performance for each areas based on minimum skills considered necessary to function in the market place.	About two-thirds of those who had *completed primary school* failed to achieve the minimum competency level in all four basic skill areas.
Banerjee, Cole, Duflo, and Linden	India 2000	Sample of schools in Vadodara and in standards 3 and 4	Pretest and posttest were administered for schools with and without remedial education assistance. The test included a math section and language section testing competencies prescribed by the Vadodara Municipal Corporation.	The raw scores on the exam exhibited a basic lack of knowledge with only 5.4% (Vadodara) and 14% (Mumbai) of *third standard* children able to pass the minimum competencies for math scoring.
National Exam. Council of Tanzania	Tanzania 1998-2000	Grade 7	Primary School Leaving Certificate (PSLC) testing skills in language, math and general knowledge.	In 1998-2000 only 21.3% (language), 19.32% (math) and 21.95% (GK) of students who sat the exam passed the PSLC.
Glewwe and Jacoby (in Lavy 1996)	Ghana 1994	Primary School	Raven's Progressive Matrices test, which measures abstract thinking ability, reading and mathematics	The mean scores on the simple reading test show scores equivalent to random guessing after *six years of primary schooling.*

[9] This fact however rarely makes it into the published literature because the test results cannot be used econometrically. This is based on conversations with economists who have encountered this problem in Ghana, Kenya, and Pakistan.

No one who has spent much time in and around classrooms in developing countries would be surprised by the findings of low learning achievement since in most, though not all, countries in the world nearly every aspect of the schooling system is seriously deficient:

- Infrastructure is lacking—schools lack sufficient classrooms and necessary facilities.
- Even the most basic instructional materials (pencil, paper, chalk) and textbooks are often in short supply, and more advanced materials (libraries, science supplies) that would allow hands-on learning almost entirely absent.
- The entire system of recruiting, training, assigning, supervising, and monitoring teachers is "badly designed" in nearly every regard[10].

Part of all of these problems is that within the public sector there is no system at all for managing for performance. The system is entirely *input* driven. To the extent that the management of the education system works at all, it works to deliver exclusively *logistical* targets—Was money spent? Were schools constructed? Are bodies in place? Often, even this logistical information is not actively used in management. The result of this is that, even though nearly all teachers are sincere and well meaning—and many are heroic in their efforts—the front-line providers do not get any support or incentives to produce high learning achievement.

I.D) Scope of the challenge of "lack of education"

The scope of the challenge of the lack of basic education should be conceived as the *failure of children to achieve mastery of the basic cognitive and non-cognitive competencies necessary to thrive in a modern economy*. Figure 6 illustrates the connections between competencies and schooling. The vertical axis represents any measure of a basic competency (e.g. literacy, numeracy, abstract reasoning skills). A horizontal line indicates the threshold minimal target for that competency. The slope of the line indicates the speed of learning (gain in competency per unit time). While most of the discussions of the "lack of education" suggest that "years of schooling" are the target, it is useful to stress that the true goal is achievement and that "years of schooling" are only a proxy. The true targets are competencies.

Children come into school with very different levels of school readiness (children A and B begin lower than children D and E) and figure 6 illustrates the four dimensions of the challenge of basic education.

- Some children, represented by child A never enroll in school.
- Some children, represented by child B, enroll in school but drop-out before completing all of the grades in the basic cycle.

[10] Actually "badly designed" is in scare quotes because this is assuming the design features must be a technical "mistake" because a benevolent and rational government would never do anything like this. But one view is that the system does exactly what it was designed to do: maximize bureaucratic power over communities, maximize union and bureaucratic leader control over teachers and communities. It is very hard to avoid the habit of assuming that the fault is a mistake in technical rationality.

- Some children, represented by child C, complete the entire basic cycle but the combination of their initial level and the low learning achievement in school leaves them below the threshold even after completing basic schooling[11].
- Other children do achieve the threshold, either because they begin the schooling cycle with high levels of preparation (children E and F) or because they are in a high quality school (School II for child D).

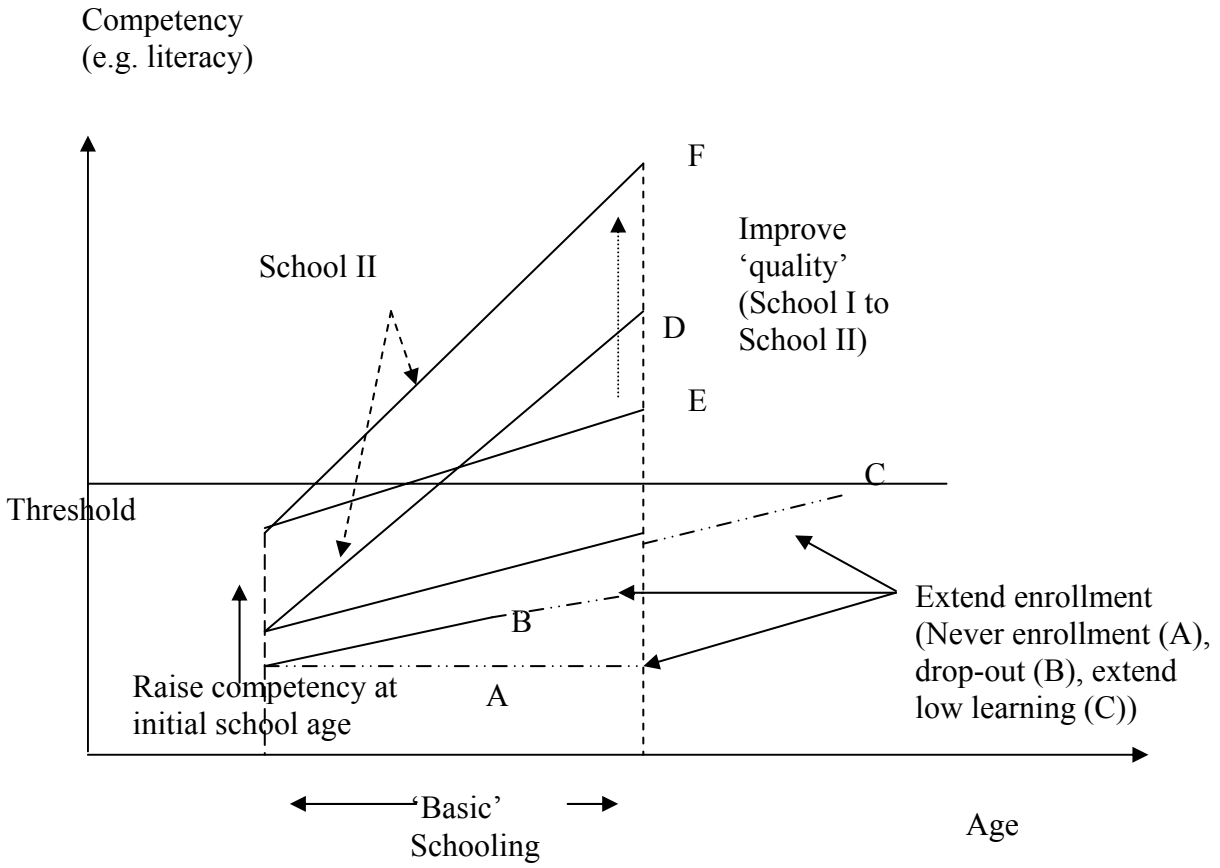


Figure 6: How to raise the levels of competency through schooling: The proximate determinants of learning achievement

Table 4 shows estimates of the lack of education decomposed into those never enrolled, those who enrolled but did not complete grade 5, enrolled and completed grade 5 but did not complete grade 9, and finally an estimate of those who were enrolled at age 15 (roughly grade 9)

[11] The same figure can help interpret the apparent deficiencies in learning achievement documented in the international comparisons above. For example, the TIMSS examinations in 1994/95 were given in seventh and eighth grade class-rooms. The median gain across this grade for the OECD and Tigers was 33 points in Mathematics and 41 points in Science. For the average Colombian eighth grader at 385 to reach the OECD median of 500 at the OECD rate of progress would take 3.5 years (=115/33) of additional schooling. But the gain from seventh to eighth grade in Colombia only gives a gain of 16 points (369 to 385) so to reach the OECD median performance at the Colombian rate of progress would take more than *seven additional years of schooling* (7.2=115/16).

but have not achieved a "minimal" level of learning achievement. Of course, establishing a "minimal" level of skills or competence will always be arbitrary, but no more arbitrary than setting up completion of "primary" school as a target—"primary" school of five years in some countries, six in others, eight in others. Presumably the basis of setting up any given level of grade attainment as the target is the presumption that the threshold level of competencies are acquired in that span (on average? for all who complete?) but judging progress just by grade attainment with no indicator of learning achievement or mastery of competencies is hollow. To construct estimates of the lack of learning achievement I will assume that students more than one standard deviation below the performance in the median OECD country "lack education."[12]

Table 4: In lower middle and middle income countries there is an enormous lack of education—almost none of which is due to children not enrolling in school…

(Estimates of the "lack of education" among a recent cohort of 15-19 year olds in various lower-middle and middle income developing countries)

Country	Cohort Aged 15-19 in year:	Never Enroll	Did not complete grade 5	Did Not Complete Grade 9	Percent completing grade 9 with "inadequate" learning achievement[1] (more than one std. dev. below the OECD mean)	Total fraction of cohort that "lacks education" (either did not complete grade 9 or inadequate learning achievement)
Colombia	2000	1.5%	12.5%	50.4%	73.4%	86.8%
Indonesia	1997	1.6%	10.3%	55.2%	48.8%	77.1%
Morocco	1992	34.8%	49.5%	83.1%	75.6%	95.9%
Philippines	1998	1.4%	7.8%	47.6%	71.5%	85.1%
Turkey	1998	4.4%	7.8%	60.0%	36.8%	74.7%

Notes: 1) This uses the TIMSS-R scores for eighth graders on Mathematics in 1999. To calculate the fraction of students with scores below 400 (one standard deviation (100 points) below the OECD median of 500) we use the country mean and standard deviation and assume a normal distribution. This makes a number of heroic assumptions: (a) that scores are roughly constant over time so the 1999 test represents the cohort 15-19 in the survey year, (b) that eighth and ninth grade competencies would be roughly similar.

The framework in figure 6 illustrates the *proximate* determinants of increased mastery of basic competencies.

- Increase levels of competency on enrolling in school—"early childhood" interventions.

[12] If scores are normally distributed, then about 16 percent of students are one standard deviation below the median. I think few would argue that those in the 16th percentile of the achievement distribution among 14-15 year olds in American schools (which are usually near the median on international comparisons) are not seriously deficient in basic skills and at serious disadvantage in a modern economy.

- Increase learning achievement by lengthening the time (and/or grade progression) in school by:
  - Reducing the number of children who never enroll in school, or
  - Reducing late enrollment (to the extent this extends years completed), or
  - Raise grade attainment by increasing the number of years a child who does enroll completes.
- Increase the competency gain per year of schooling.

Part II) Analytical Framework for Evaluating Opportunities

What opportunities are available for addressing the lack of education? None of the *proximate determinants* are under the *direct* control of any given international or national or regional policy maker. Rather, the levels of these proximate determinants of competency—early childhood preparation, enrollment, persistence, learning achievement—are all the outcome of the decisions taken by millions of households (parents and children) and by millions of educators.

An "opportunity" is a *feasible "intervention"* that affects outcomes. An "intervention" just means some change—somebody does something differently—which could be anything from greater expenditures, to increased teacher training, to adopting a new textbook, to lower school fees, to systemic changes in the "rules of the game" in the way schools are run. Any assessment of the "opportunities" must invoke a plausible model in which the "intervention"— changes the previous decisions made by parents and teachers in a way that leads to more learning (through any of the channels in figure 6—new enrollment, less repetition, longer persistence, or greater learning) in the post-intervention situation.

To evaluate an "opportunity" one has to specify a *complete, coherent, causal chain* from the action to the desired outcome. That complete coherent causal chain has to include a *plausible positive behavioral* model for each agent (parent, child, teacher, headmaster, politicians, analysts) whose actions are in the critical path from policy action to outcome. Simply put, one needs to be able to answer the question "why will they do that?" for everyone involved. If one claims that building more schools will increase enrollment, one needs to be able to say *why* the children will show up. If one claims that expanding the budgets will raise learning, one needs be able to say *why* those using the additional resources will use them to increase learning. If one claims that rigorous evaluation will improve outcomes, one needs to be able to say *why* the new knowledge will be adopted. And, if one claims teacher training will improve learning, one needs to be able to say *why* the training will change teacher behavior.

The analytical framework used to evaluate opportunities will be *demand*—to understand the education choices of parents and children—and *supply*—to understand the decisions of those engaged in the provision and production of education:

II.A) Demand

The schooling a child receives depends on decisions taken by parents and the child. This decision balances at the margin all of the anticipated private benefits against all of the costs. The benefits of schooling include pecuniary benefits from higher expected levels of income and non-pecuniary benefits which include improved well-being (e.g. health) and may include other psychological benefits (e.g. a greater sense of personal worth). The benefits depend on the

quality of the available schooling options as well as the current academic inclinations and capabilities of the student—which are affected by previous decisions.

The marginal benefits of additional schooling are balanced against the marginal costs of additional schooling options. The costs include both opportunity costs of the student time devoted to travel—to and from school, in class time, homework—and all of the incremental money costs of enrollment in school—tuition, fees, books, supplies, and uniforms. Credit constraints may affect schooling decisions such that investment choices depend not just on increment to discounted life-time benefits versus costs, but also on current income.

II.B) Supply

The supply of education depends on both schooling and non-schooling inputs into a child's education (education is *not* synonymous with formal schooling). The *supply* of formal schooling could come from public, private-religious, private non-religious non-profit (e.g. NGOs), or private for profit firms.

It is perhaps provocative, but not unreasonable, to think of "basic schooling" as one segment of a much larger market for "instructional services." There is a market for instruction for both children and adults in a wide array of activities—foreign languages, sports, music, dance, religion, and trade skills (such as computer training). It is impossible to walk down the street of any even medium sized city in a growing economy and not be struck by just how many of the small businesses are related to instruction of one kind or another. The typical "primary school" produces a package of instructional services as a graded, sequenced curriculum delivered through some type class-room guided instruction involving lectures, exercises, readings from texts, homework, feedback, etc.

It is assumed, as a first approximation, that instructional services are a "perfectly competitive" industry in the economists' sense and that the industry supply curve of primary schooling is roughly infinitely elastic. Schooling is not particularly capital intensive, the economies of scale are not particularly large and costs of entry and exit are small.[13] The markets for all other child related instructional and care services are dominated by small providers, with large firms almost completely absent[14]. Moreover, instructional services do not rely on factors that are in fixed supply in the long-run, so the long-run elasticity is likely quite large.[15] There is often a great deal of confusion about schooling being "supply constrained". What is "constrained" is the supply of schooling services at less than their full cost. At a price equal to

---

[13] The costs per student obviously rise as students per teacher rise. So when total students is very small then costs are high. But primary schools with a few hundred students are cost effective, and there are some arguments that schools above a few thousand students are arguably "too large" so economies of scale will not make any single school large as a fraction of the market. While there are economies of scale in some inputs into schooling—e.g. textbooks, instructional materials, test development—these can be provided by separate firms.

[14] In the USA the market for child care and other child oriented instructional services—e.g. music, dance, religion, sports—and for tutoring are dominated by small suppliers. For example, no large firm has been able to make significant inroads into the substantial market for piano lessons. Even in universities, where one might suspect some substantial economies of scale, in the USA where there is a competitive market. There are roughly 2million [check] students enrolled in higher education in the USA while the highest quality schools have remained quite small.

[15] The major input into instruction is people with sufficient education to become teachers. In very low-income environments teaching may absorb a substantial fraction of the educated population, and hence make the supply of people into teaching modestly upward sloping.

the long-run marginal cost (equal to long-run average cost since the cost curve is flat) the supply of schooling is almost certainly enormously (if not infinitely) elastic and analogy and historical experience suggest that in the absence of any government intervention primary schooling would be provided by thousands of small enterprises. Schooling is unlikely to limited by a lack of suppliers of educational services.

Of course, in practice nearly all primary schooling is carried out by producers who provide these services at less than their full cost, including governments and religious groups. The principal lacunae in the economics of education is that, while the bulk of basic education around the world is carried out by the government directly, *there is no empirically plausible, general, positive behavioral model of the public sector production of schooling*[16]. That is, there are a variety of perfectly coherent "market failure" or "equity" justifications that *normatively* justify *some* public action in education. However, assuming "normative is positive"—that the *reason* governments do what they do is *because* of the market failures or equity *rationales* for public sector intervention—does not have even surface plausibility as a *general* model.

One needs a positive model of schooling, not just a general model of political economy because the features of schooling span across a variety of types of governments. I have developed this argument at length elsewhere (Pritchett 2004a) but there are four obvious empirical points that argue against "normative as positive" (NAP) as even a useful first approximation to a general positive model. First, as has been pointed out by economists from at least Milton Friedman to Caroline Hoxby—"anything public production can do a voucher can do better." That is, while NAP can justify some public action, it cannot explain why the chosen instrument of public sector intervention has, nearly everywhere and always, been the direct production of schooling services by a public sector agency. Second, even if NAP could explain some production, it cannot explain why this is—nearly everywhere and always—the only support for education. Third, NAP to explain public schooling is not a coherent general model of public sector action, as governments, especially in developing countries, produce schooling even when they are not democracies and do a range of very nasty, corrupt, abusive, or just grossly indifferent to human welfare actions as well. Invoking one model to explain why governments do "good" things and another to explain why they do "bad" things is more than just a little *ad hoc*. Fourth, NAP has a hard time explaining the actual behaviors of government in the allocation of expenditures, either across levels of schooling, or across inputs.

Therefore it cannot be simply assumed—because it is frequently plainly counter-factual—that (a) governments produce schooling because they are intrinsically benign, (b) governments are other than rhetorically committed to universal enrollment, (c) governments are deeply interested in quality schooling, or that (d) the public sector produces schooling services efficiently. As with any other actors, the actions, efficiency, and efficacy of public sector officials (politicians and policy makers) and producers in the public sector—as suppliers of schooling—will depend at least, in part, on the incentives they face and on ideologies that have little to do with education as an enterprise for the delivery of cognitive development services, but rather to the control of education as a key for the transmission of socialization.

---

[16] A key qualification here is "general," while there are many theories of why particular governments produce education (e.g. economists who work on "median voter" models to explain government production (Kremer and Sachryvev, 1998, Gradstein and Justman 2002) these suffer from the defect that (a) median voter models are not particularly robust even in the best of cases and (b) it is not obvious median voter models are at all relevant in non-democratic settings.

II.C) What are the opportunities and how can they be assessed?

With the field prepared—delineating the problem as a lack of learning (for which enrollment in school is the means) and delineating a framework of the determinants—we are ready to discuss the specific opportunities often proposed in various literatures in a piecemeal way within an overall coherent scheme. In assessing opportunities there are two fundamentally different alternatives, depending on whether one endogenizes the behavior of the government or not: *Policy Action* or *Systemic Reform.*

Four policy actions. The approach which is nearly always taken is to *assume* a government that is motivated to remedy the lack of education and examine the question "What policy interventions should a government, which is willing and politically able to act, take to address the lack of education?" This approach assumes that governments *produce* schooling and hence control the supply side directly as a matter of policy and can decide to build more schools or train more teachers or raise class sizes or lower class sizes, etc. Table 5 delineates the four classes of policy actions (with alternatives in each class) that will be considered.

- Supply side, expand the quantity: Building or expanding schools.
- Supply side, improving quality: (a) increase expenditures radially, (b) increase expenditures on specific interventions, (c) expand rigorous evaluation of interventions.
- Demand side: Direct (targeted) support to households that lowers the cost of schooling.
- Demand side: Raising the benefits of schooling.

Table 5: Proposed policy actions to be considered as opportunities to address the lack of education

Supply		Demand	
Quantity	Quality	Cost reductions	Raising Returns
Physical expansion	▪ Expansion in spending per student ▪ Expansion in specific, known, pedagogical interventions (e.g. increased instructional materials, textbooks, teacher training) ▪ Expansion in the evaluation of pedagogical innovations, which are then taken to scale.	▪ Vouchers ▪ Gender ▪ Conditional transfer ▪ School feeding/health ▪ Late enrollment ▪ School Fees	Policy reform *Interactive growth*

Systemic Reform: the Fifth Element. The approach of evaluating supply side policy actions that assume direct government production is unusual for economists. In nearly all other markets economists consider government interventions as taking the form of setting the rules of the game and of affecting the relative prices by taxes and subsidies or mandates. They then examine the outcomes that emerge from the decisions of consumers and producers. Economists rarely pretend to superior sector specific knowledge about the technology of production that would allow them to argue for more of this or that particular input into the production process. Imagine how odd the equivalent of the "class size debate" would be in other contexts—economists recommending how much labor Dell computer should hire in computer assembly or the temperature at which steel producers should anneal or the mix of car colors GM should produce. Even with education related areas like private firms' on-the-job training, economists assume that firms—as demanders of education—choose the privately optimal amount of training and that

competition in the market for training (including the possibility of in-house production) leads suppliers to be productively efficient.

One possibility that needs to be seriously considered is that the desirable policy actions discussed above are not adopted when governments as direct producers of education face the wrong incentives. In this case, the "wrong" policy action is not a "mistake" but an endogenous *outcome* the existing rules of the game and incentives. If existing public sector actions are in fact the stable result of the existing institutional design (a word to describe the formal and informal "rules of the game") then "recommending" new policy actions might be pointless. If teachers have been hired by local politicians as a means of providing political patronage, then "recommending" these politicians hire fewer teachers is itself an odd behavior.

Alternatively, one could make government actions in the education sector partially or wholly endogenous and focus on "how reforms of institutional conditions for the provision of education would impact the level of education by changing the opportunities facing demanders and suppliers of education?" I want to stress up front that I am not saying that existing deficiencies mean public sector education is "unreformable" or that a private market with vouchers is a panacea. After all, as one examines the high performing countries in education in any given dimension—coverage, equity, learning achievement—one will find countries with public sector producing education. Teachers are just as much employees of the government in high performing Hungary, Korea, or Singapore as they are in low performing Morocco, Nigeria, Peru or the Philippines.

The question is how the system of responsibilities and accountabilities for production are structured. A "market" is one way of structuring responsibilities—individual firms control all of the production decisions (what mix of products to produce, what production techniques to use, what factors and inputs to hire) and are held accountable by the decisions of consumers to buy or not buy their products. A "centralized hierarchical organization" is another way to structuring responsibilities and accountabilities (internal, top-up reporting).

The question which part IV addresses is "What pattern of responsibilities and accountabilities and set of policy induced relative rewards will produce high levels of education as the *endogenous outcome* of the decisions of individual producers of educational services and parents and children?" This takes the role of analyst from direct recommendations about production decisions (e.g. class sizes should be higher/lower) to recommendations about the structure of responsibilities, accountabilities, and incentives from which the correct decisions about class size will emerge.

But once policy actions are endogenous "recommendations" about opportunities become problematic. To tackle the question of the opportunities available to address the lack of education, one must ask—the opportunity for *who*? For a national government? For a parent? For an individual teacher? For a headmaster? For a researcher/policy analyst? For an NGO? For a bilateral or multilateral donor?

Part III) Evaluating the four opportunities for policy action

This section evaluates the four classes of policy actions with the question: "are there policy actions with high returns in addressing the lack of education"

Opportunity number 1: Supply side, expanding the quantity of schools

The expansion of education has always had a high priority in development efforts[17]. Clemens (2004) reviews the long history of commitments to universal primary education from at least the Universal Declaration of Human Rights in 1948 to the Dakar Declaration of the World Education Forum and Millennium Declaration of the Millennium Summit in 2000. These rhetorical international commitments to education all implicitly adopt a similar conceptual framework— "needs are *the* problem, direct government supply is *the* solution, and available financing is *the* constraint" —that pervaded much of the approach to development assistance (Scott 1999, Pritchett and Woolcock 2004). Since schooling is a "need", the demand side is not considered a key issue and the "opportunity" most prominently on the international agenda to address the lack of education is an expansion in the supply of government produced school places. With the assumption that low-income governments are doubly constrained both by low income and low capacity for revenue mobilization, the key constraint on expansion is financing—so that external financing can play a key role.

This conceptual approach is reflected in some answers to the question "what is the cost of meeting the Millennium Development Goal for universal primary completion?" The most straightforward calculation is to estimate the sum, across all countries, of the number of children/years of schooling to reach the goal, times an estimate of the average cost.

$$TotalCost = \sum_j (Additional\ schoolyears)_j * (Average\ cost\ of\ a\ year\ of\ schooling)_j$$

Table 5 summarizes some of the existing estimates for total additional annual resources necessary to put every child who is out of school through primary education. There are different studies available regarding the number of children of primary school age who are currently "out of school" which provide totals between 100 (Devarajan et al., World Bank 2002) and 120 million (Delamonica et al., UNICEF 2001) children. The major discrepancies in these estimates are the different methodologies employed in calculating the average cost per student. UNICEF (2001) use actual average costs (public expenditures divided by enrollment) country by country. The World Bank (2002) estimate uses four methods: an average cost per student per year of $110.6 for all countries, assuming the average in each country is equal to the regional median, assuming existing country averages, or assuming the average cost is 13 percent of GDP per capita in each country. The range of estimates is from US$9.1 billion to US$27.6 put an additional 100-120 million children through primary school.[18]

[17] It is a pernicious myth that recent economic research on economic growth has "discovered" or even "emphasized" the role of human capital more than earlier approaches to development. While the Solow model had no formal role for education all of the major development theorists (e.g. Lewis, Myrdal) acknowledged the key role of education and knowledge acquisition at least as early as late 1950s/early 1960s.

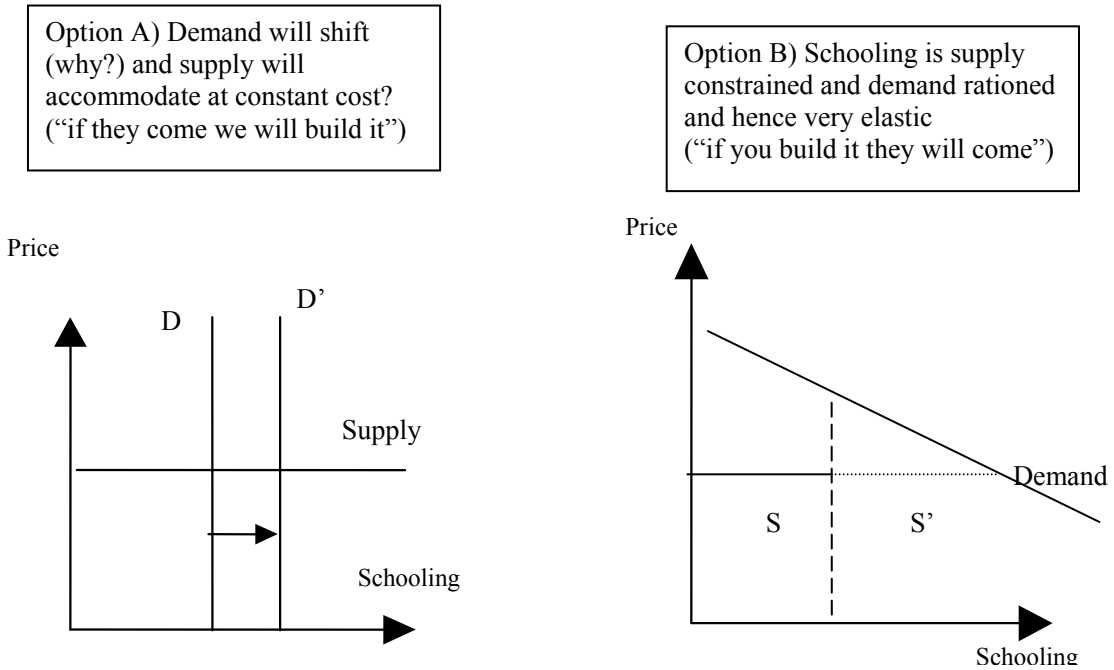
[18] The different assumptions have the obvious implications—if constant costs are assumed across countries it is expensive (US$4.94 bn) in Africa (where actual costs are lower and many children out of school) and cheap (US$730 mn) in Latin America (where costs are higher but few students are not in school). Alternatively, if costs are a fixed fraction of GDP per capita the estimated incremental cost is higher (US$ 8.1 bn) in Latin America (which has (relatively) high GDP per capita) and low (US$1.27 bn) in Sub Saharan Africa (where GDP per capita is low).

Table 6: Summary of additional annual spending at constant average costs to reach universal primary enrollment

		World Bank 2002				UNICEF 2001
Estimate Method	*Average cost per student* *No. students*	*$110.6 per out of school student* *100 million*	*13% of GDP PC spent on all school-age children* *100 million*	*Regional median level spending* *100 million*	*Country level spending* *100 million*	*Percentage of per capita income (country level)* *120 million*
SS Africa		4.94	1.27	2.63	2.15	2.90
South Asia		2.69	1.58	2.24	1.80	2.20
East Asia & Pacific		0.89	10.4	0.89	0.38	0.40
Middle East & North Africa		0.90	5.73	0.87	2.18	2.20
Latin America & Caribbean		0.73	8.10	1.45	3.23	0.90
E. Europe & Central Asia		0.30	0.46	0.47	0.63	0.60
Total		**11.4**	**27.6**	**14.9**	**10.4**	**9.1**

But it is not immediately obvious how to interpret these estimates, as they are only the answer to the question: "*If* children who are not currently attending school *were* to decide to attend school *and* each of those children could be accommodated with an expansion of the current system with marginal cost assumed equal to average cost *then* how much would additional spending be?" But this is not the answer to the question "how much will it cost to attain universal primary completion?" unless one believes that the causal mechanism whereby universal primary completion will be achieved and it is, more or less, "if you build it they will come." That is, another way to interpret this approach is that the supply of public education (e.g. school buildings and classrooms) is fixed and that the demand for education by the household is constrained—in the sense that the number of people who would like to be enrolled at the given price of schooling is far higher than the provided capacity—and hence are "rationed out" in some way or another. Therefore an expansion in the government-produced supply of education will translate one for one into increases in enrollment and attainment.

Figure 7: Interpreting the use of current average cost to "cost" the achievement of the MDGs of universal primary completion. Is the causal interpretation A or B?

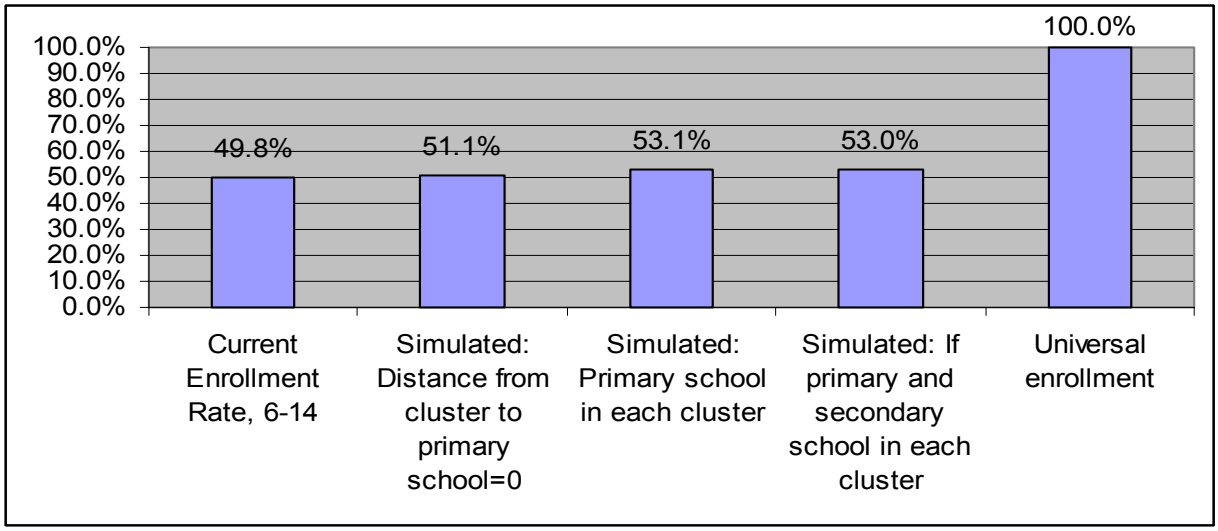


What is the evidence about the costs and benefits of an *expansion of the educational facilities (primary and secondary schools) as a means of addressing the lack of education?* Filmer (2003) examines the correlates of child enrollment using the (nearly) identical datasets from the DHS from 21 different countries (24 data sets as India, Bangladesh, and Niger has two) including child, household and cluster level covariates. These DHS datasets included a "community" level questionnaire that included information about the distance from the sampling cluster (a very small unit—usually a census tract) to a variety of facilities—primary or secondary schools, a post office, a market, a paved road, etc. His findings suggest very strongly two points. First, in 16 (14) of the 24 data sets the distance from the cluster to a primary school is a statistically significantly at the 10 percent (5 percent) level as a correlate of child enrollment, with similar findings for a binary variable for whether there was a primary school in the cluster.

Second, the estimated effects of school construction, while positive, are small, *very* small. The median enrollment of children aged 6 to 14 in this sample of country datasets (which are by no means an estimate of "global" totals) is 53.2 percent. Reducing the distance to a primary school to *zero* raises the median enrollment to 54.7 percent—1.5 percentage points. Alternatively, the use of estimates from a binary variable to simulate the enrollment impact of having a primary school in each cluster raises the median enrollment rate from 53.2 to 55.4 percent—2.2 percentage points.

Figure 8: Using the estimated relationship between enrollments and distances suggests ony modest progress in enrollments from building more schools

(Actual and predicted average enrollment rates of children aged 6-14 in the 21 countries using estimates in table 9)



Source: Filmer 2004.

There are methodological pluses and minuses of relying on studies that implement exactly the same regression across a number of datasets which represent a variety of circumstances. The plus is that it increases the generalizability of the reported results, in two senses. First, there is likely enormous publication bias—since a "failure to reject" is often treated as less interesting that a "rejection"—so a review of the *published* literature might overstate the typical correlation substantially. Second, there is clearly a good deal of variability to the results—the linear probability coefficient ranges from 3.1 percent to 0 (or even positive in the case of Senegal) and a selective review of the literature which emphasized the largest findings could overstate the "typical" response by a factor of 4, or more[19].

The minus is that datasets which are cross-nationally comparable on a large scale typically do not contain data sufficient to address the "identification" problem. Hence the step from a reported multivariate correlations--which is perfectly valid only as a descriptive statistic--to treating the coefficient on distance as representing a reliable estimate of an exploitable causal impact which any simulations of impact must assume cannot be taken. The bias of the OLS regressions as estimates of causal impact is impossible to predict *a priori*—it could be positive or negative. If schools are located in villages with characteristics that are not included in the

[19] These points are consistent with the existing literature that finds a mix of strong and small or no effects of distance to facilities on enrollments. Lavy (1993) found that the elasticity of the probability of enrolling in primary school with respect to the distance to middle and primary school is 0.30 and 0.07, respectively. In a study evaluating the relative importance of supply and demand factors in determining primary school enrollment in rural Ghana, Handa (2002) found that increasing school coverage by decreasing the travel time to school (i.e. more school buildings) would have a larger impact on enrollment (13 percent increase) than income changes (2 to 4 percent increase). Based on these studies in Ghana alone, one has no idea whether these results are large or small relative to the typical result. Other studies, such as Holmes (1999) in Pakistan and Burfey and Ifan (200) in a number of countries find little or not impact of distance.

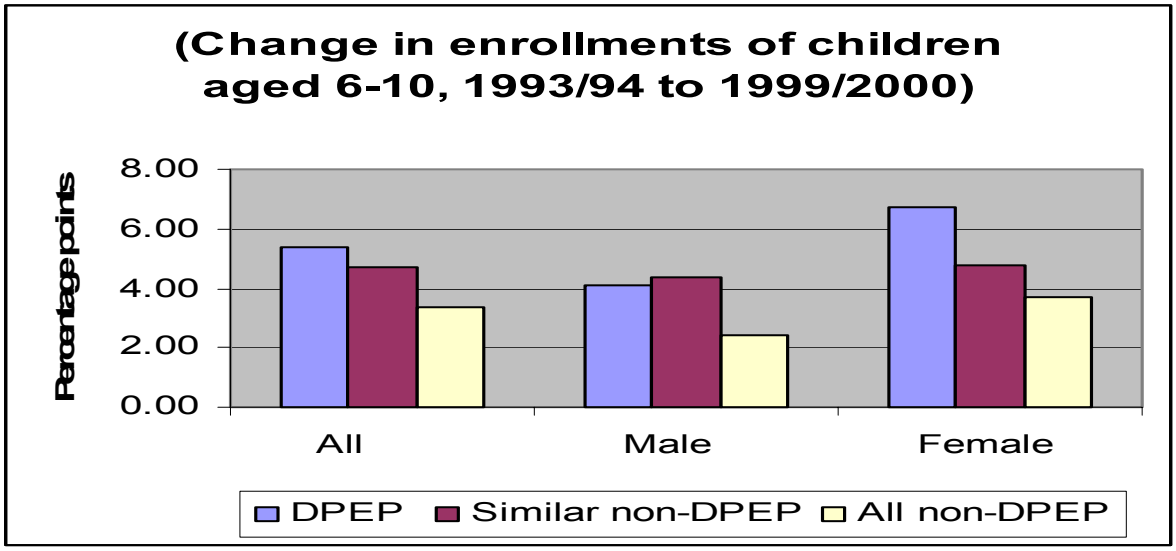
regression equation that are positively correlated with the (net) demand for education then OLS will *over* estimate the causal impact. Conversely, if the placement of schools were determined by some deliberately compensatory approach and schools were more likely to be present where enrollments would otherwise have been low, then the OLS estimates will *under* estimate the impact of additional construction[20].

Evaluations of specific interventions are another means of assessing the responsiveness of enrollments. The District Primary Education Project (DPEP) in India has spent 1.62 *billion* dollars since 1994 to expand schooling, and improve the quality of schooling, in India. Implemented in a phased manner in 242 districts in India—chosen because they had low female literacy. The project is much more than a "build schools" project--the project had many components including school construction provided financing for non-teaching inputs, teacher training, formation of village committees and mother-teacher associations. Therefore the impact should be much *larger* than a "build schools" approach. A recent evaluation finds that while enrollments in the DPEP districts did improve and the "before and after" if one either compares the performance of DPEP to either (a) all non-DPEP districts, (b) similar non-DPEP districts or (c) use propensity matching scores the impact of DPEP on enrollments is at most—about 1 to 2 percentage points. Figure 8 from her paper shows the evolution of enrollment in DPEP districts, non-DPEP districts, and in non-DPEP districts that were below average female literacy. The improvement in the DPEP districts is very difficult to distinguish from the progress in the non-DPEP districts[21].

[20] In his study Filmer (2004) uses fixed effects for the two countries with repeated surveys and common clusters and finds substantially smaller fixed effect than cross sectional OLS estimates—suggesting that, if anything, the cross sectional estimates in table 7 are too large. Pitt, Rosenzweig and Gibbons (1993) used fixed effects to eliminate the impact of selective placement based on geographically fixed characteristics and find that their fixed effect estimates are different. Duflo (2001) identifies the impact of additional school construction using data from Indonesia that combines the geographic distribution of a large-scale school construction program in Indonesia (1973-1978) with information about individual's location of birth. She found substantial impacts—children aged 2 to 6, when a school was constructed, received 0.12 to 0.19 more years of schooling for each school constructed per 1,000 children. The fact this is compelling, well identified, evidence about the causal impact of school construction, but that there was *some* impact at least in *some* places, was never in doubt and this does not address whether the OLS are over or understated.

[21] Jalan (2003) uses the econometric technique of propensity score matching to identify the counter-factual by asking essentially—"how much more did enrollments increase in DPEP districts than in districts that were otherwise comparable but did not participated in DPEP?" Across a variety of indicators of performance she finds that the naïve "before and after" approach dramatically overstates program impact and the "differences in differences" estimates of either DEPP and non-DPEP districts. Her best estimates of the program impact are mixed: (a) the impact on enrollment rates is either small—1 .3 percent increase in enrollment rates due to the program or, strangely, negative, (b) the impact of girls, who were a special focus, is uniformly small or negative, (c) the impacts were much more positive in one state, Madhya Pradesh which also implemented two other educational programs in tandem so, if anything, the estimates *over*state the pure DPEP impact.

Figure 9: Evaluation of a large scale intervention in India (District Primary Education Project) in 42 low female education districts from 1994 to 2000 shows positive, but modest, positive impacts on enrollment rates



Summary of opportunity 1: Expansion. On the opportunity of raising schooling through supply side expansion, the international consensus is in a strange position. Although much of the international rhetoric has a "if you build it they will come" flavor and the reporting "costing" exercises of international targets make assumptions that seem to be consistent with this view--almost no international expert on education believes this is literally true (except perhaps in a few of the world's poorest countries). Rather, it is now generally accepted that success depends on increasing the demand for schooling through either (a) increases in quality (opportunity 2), (b) raising returns (opportunity 3), or lowering costs (opportunity 4). What divides opinion is primarily whether these are possible through policy actions or require systemic reform (the fifth element).

Opportunity number 2: Improving quality, supply side policy actions

Given the enormous progress most countries in the world have already made in expanding the basic physical and human infrastructure of the schooling system, improving the quality of schooling is almost certainly a key to addressing the lack of education for two reasons. First, as shown above, the fact that a large fraction of the children do not complete basic schooling is due to children who begin schooling and fail to progress and/or abandon school. The low perceived quality of school is one factor behind low progression and high drop-out rates. Raising the quality of schooling therefore raises the demand for schooling. Second, many children are completing the primary and/or basic education cycle with astoundingly low levels of learning achievement. Raising the learning achievement per year of schooling will benefit not only those at the margin, but also all those who are currently enrolled.

A key debate in education today is between those who believe that increasing expenditures per student (perhaps generally, perhaps devoted to some key input) is key to improving learning achievement and those who believe that without serious systemic reform increases in expenditure are unlikely to be effective in raising learning achievement. As with any starkly stated sides in any debate the vast majority believe some combination of "both"—

that improving learning achievement will require both more money and more "performance accountability" (one of the key slogans of systemic reform). But "both" is not really an answer as the *proportions* and *sequencing* in the mixture of money and reform are both crucial—is the mix 90-10? 50-50? Or 10-90? Should money lead reform or vice versa?

A framework for reviewing the literature—and the example of the "class size" debate

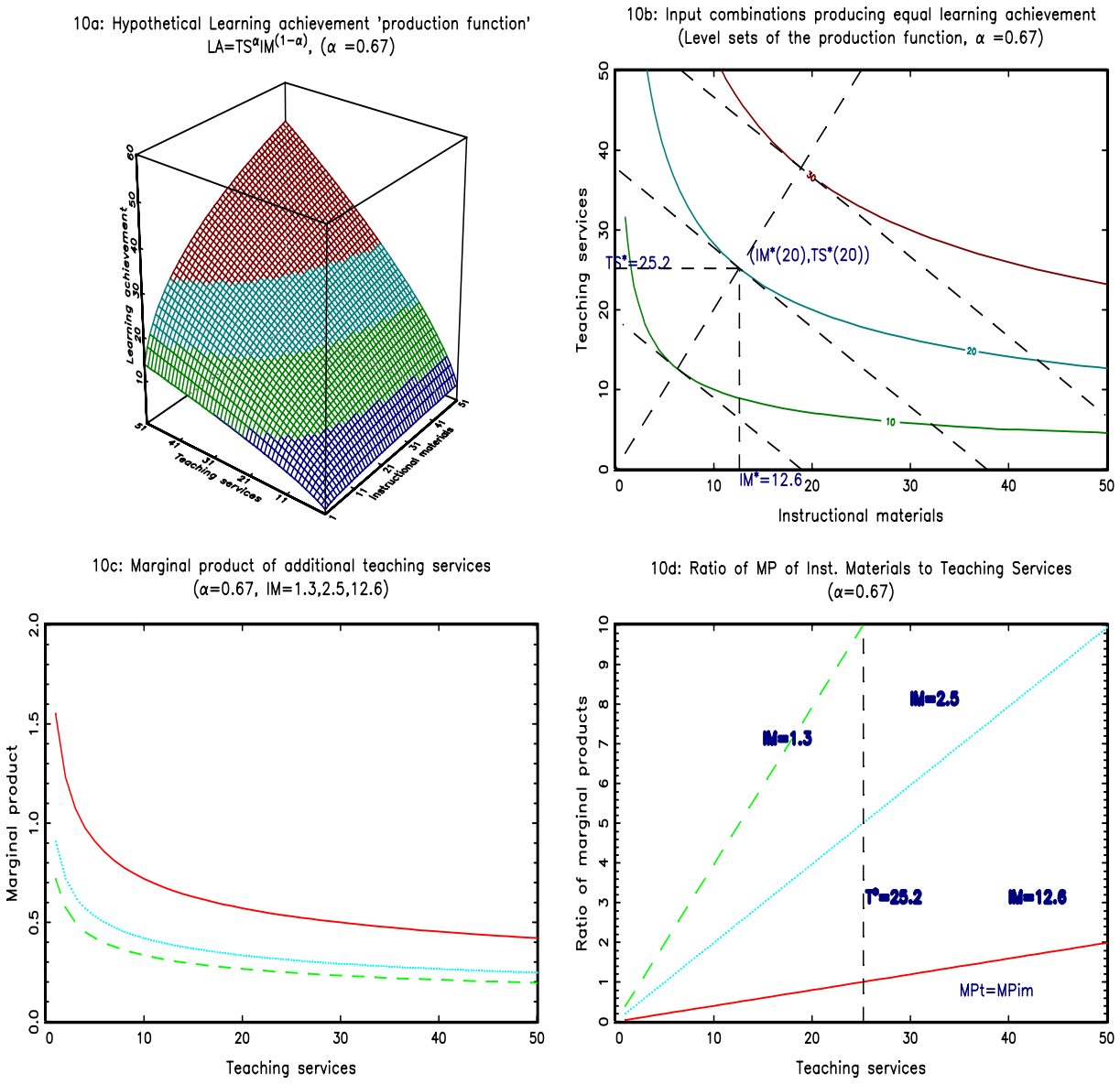
Reviewing the empirical literature relevant to this question requires some basics, which are presented in figure 10[22]. Suppose there is some connection between learning achievement and inputs, as pictured in 10a between two inputs 'teaching services' and 'instructional materials' where the surface represents the *maximum achievable* learning achievement from any combination of factors. To maximize learning achievement subject to a budget constraint and the relative price of teaching services to instructional materials producers should be on the surface (that is be productively efficient for any combination of inputs) and the particular point on the surface should equalize the marginal learning achievement gain per dollar of additional input. This is represented in figure 10b by the equalization of the slope of the budget constraint—which is the relative price of the two inputs—and the slope along the level sets of the learning achievement surface (combinations of teaching services and instructional materials that produce equivalent outputs). For instance, for this particular functional form of the learning achievement surface and the given parameters the optimal combination of inputs for produce learning achievement of 20 is 25.2 units of teaching services and 12.6 units of instructional materials (all on arbitrary scales).

Figure 10b also illustrates that as budget available expands, the optimal amounts of the two inputs increase and the maximum achievable learning achievement at the optimal allocation increases.

Figure 10c illustrates two common assumptions about the "education production function"--that the productivity of additional teaching services depends on the level of teaching services and illustrates *declining marginal productivity* and that the productivity of teaching services is higher the more instructional material inputs are available. Figure 10d illustrates that the relative marginal product of the two inputs varies widely. For instance, if inputs are at their optimum for producing 20 (12.6 of IM, 25.2 of TS) then the marginal products are equalized. If however, at that same level of TS one were only using 20 percent of the optimal instructional materials the marginal product of additional instructional materials would be 5 times higher than additional teaching services, and at only 10 percent of the optimal IM the marginal product would be ten times as high. This means that, away from the optimum the increment to learning achievement from additional budget can depend completely on how that additional budget is spent.

[22] Everything about this example is kept as simple as possible to illustrate the points—but everything could easily be generalized in nearly any way with the basic points still holding true.

Figure 10: An "education production function"--learning achievement the productivity of additional inputs



This simple analytic concept of an "educational production function" can combined with the enormous empirical literature estimating the impact of class size reductions to illustrate three fundamental points about the methodology and empirical findings:

- Uniformly "zero" marginal products is rarely an interesting hypothesis
- The "identification" problem—statistical associations do not estimate "treatment effects"
- That school "value added" accounts for only a small fraction of total differences
- The observed "treatment effect" will vary widely, hence a positive model of producer behavior is needed for interpretation.

Basic one: Uniformly zero marginal product is not an interesting hypothesis. There is an enormous literature estimating the relationship between class size (or more broadly, teaching services per child) and learning achievement. A central paper in this literature is a review by Hanushek (1986) which claimed that the bulk of the evidence suggested that the typical impact of reductions in class size on improving learning achievement (usually as measured by some type of standardized test) was statistically insignificantly different from zero. However, no one ever seriously suggested the effect of class size was everywhere and always zero. An analogy would be investigations into rice yields. If the effect of additional applications of fertilizer is non-linear and displays diminishing returns—or even negative returns past a certain point—then the impact of additional increments of fertilizer will depend on the rate from which the increases are made. One can imagine finding that if current rates of fertilizer application are high the impact of additional applications on yield would be quite near zero (or even negative). But this would not be interpreted as "the impact of fertilizer on rice yields is *everywhere and always* zero." In contrast, someone might suggest that playing Wagner music while the rice was growing would increase yields. In this case it is a meaningful hypothesis that the impact of Wagner on rice yields is zero at all rates of application (volumes, durations, pieces).

The first interesting question about class size and learning achievement is properly framed as the question of the *magnitude* of "treatment effects" of class size reductions starting from existing levels relative to the cost and relative to the marginal products of other inputs. The debate about "statistical significance" of a test of zero is therefore misguided from two points of view. First, a "failure to reject" can either be because the estimated magnitude of the coefficient is small or because of low statistical power. Second, the only reason zero is interesting is that zero is almost certainly a "small" effect from any point of view—but the converse is not true—one could easily reject that the class size was zero but also be able to reject that the class size was of sufficient magnitude to make class size reductions cost-effective[23].

Basic two: The identification problem--OLS regressions do not necessarily identify treatment effects. As a statistical technique multi-variate OLS uses the covariances in the data to estimate the partial correlations between achievement and class size. However, as economists have long recognized, if the data is generated by purposive choices then it is not clear how to interpret these partial correlations. For instance, if there are good teachers and bad teachers within a school and parents/students are allowed to choose teachers and good students choose good teachers then good teachers will have large class sizes with high performing students and bad teachers will have small class sizes and the data will suggest a negative link between class size and performance—even though for any given teacher a reduction in class size would improve student performance. Another possibility (Lazear 2001) is that if students are assigned to classes purposively by school administrators then they may choose to place disruptive students in small classes and these students may detract from learning of others so that smaller class sizes

[23] Therefore increasing refinements on what the existing studies show, either from "meta-analysis" (e.g. Hedges et al. 1994) or from differing weights on different studies, would only be interesting to the extent they are framed around whether the evidence suggests the marginal product per dollar is higher or lower than other feasible policy actions.

may again be associated with lower learning achievement—even though the "treatment effect" of lower class sizes is positive[24].

The empirical solution is to use variation in class sizes that is "exogenous"—that is, that component of the variation in class sizes that is *not* caused by choices of parents/students or administrators—to identify the "treatment effects." This has been done in three ways: using exogenous variation in class sizes caused by demographic fluctuations (and/or classroom assignment rules), using experimental variation, or most recently using within school variation in class size across grades.

The use of demographically induced variation in class sizes has tended to produce mixed results. Angrist and Lavy (1999) using the "Maimonides" rule in Israel, which limits class sizes, find statistically significant and modestly sized impacts in some grades and some subjects. Case and Deaton (1999) essential use the *apartheid* restrictions on residential mobility in South Africa to identify class size effects (as this eliminates people being able to move to good schools) that are statistically significant and substantial—in a situation in which class sizes were on average large and also quite variable. Urquiola (2001) finds substantial class size impacts in Bolivia using exogenous demographic variation. In contrast, Hoxby (2000) uses administrative data from Connecticut and class size variation induced by demographics and finds "high powered" rejections of zero—that is, she fails to reject zero impact and can reject even modest sized positive impacts. While this literature produces some positive and statistically significant results the reported results are "hit and miss"—there are class size effects in some grades and subjects and not others—with no particular pattern. Moreover, this literature is subject to enormous potential "publication bias" at every stage of the research.

The use of true experimental variation is much more rare, so that an experimental study that has gotten considerable attention is the Tennessee STAR (Student Teacher Achievement Ratio) Project in which class sizes were reduced to very small sizes (Krueger and Whitmore, 2001). This study found that there were large impacts at the very early grades, particularly for disadvantaged students. An issue we return to in the next section is that randomized experiments in education are not necessarily compelling identification of treatment effects because they are not "double blind"—the subjects know they are participating in a study. This leads to three types of bias. First, there are potential "Hawthorne" effects, in which performance might improve just because it is the subject of study. Second, there are potential "management" effects, in which the involvement of those administering a study assists in the solution of problems. Third, as pointed out by (Hoxby 2000) if teachers know that if the experiment "proves" the impact of class sizes is large this will lead to lower class sizes and if small class sizes are good for teachers, then teachers will be motivated to work extra hard temporarily during the experiment. All of these will lead the "experimental" treatment effects to be larger than the treatment effects of widespread implementation.

Recently, Woessmann (2002) has used a new approach which addresses several of the limitations. He uses the fact that the TIMSS examinations tested students from one class in two

[24] Of course merely because it is possible to construct a behavioral model in which choices cause the OLS to understate the LATE, does not mean it is plausible that any proposed behavioral model does affect the observed partial correlations and hence that OLS does, in fact, understate the LATE. Moreover, estimation bias is not an "all or nothing" phenomena—the behavior induced biases of OLS coefficients for LATE may be large or small relative to the application of interest.

different grades in the same school and recorded both the class size of the class tested and the average class size in the grade in the school. This allows him to use both school fixed effects (which should eliminate the bias of school sorting) and also use the grade average class size as an instrument for actual class size (which should eliminate the bias of within school sorting of students into classes). The TIMSS data allow him to produce school fixed effect-instrumental variables estimates for 18 different countries and implementing the same methodology for a number of countries and reporting all of the results reduces "publication" bias.

These results are clear that (a) there is a large bias in the weighted OLS (WLS) results but that (b) the statistical procedure to address the OLS bias (school fixed effects with instrumental variables (SFE-IV) produces estimates that are quite small (or the SFE_IV have huge standard errors). These two propositions can both be true because the WLS results give results that are mostly of the "wrong" sign as they suggest that smaller class sizes are associated with *worse* performance (only 11 percent in math and 22 percent in science are negative). By reducing the bias SFE-IV therefore makes a negative effect of class size into a small positive effect.

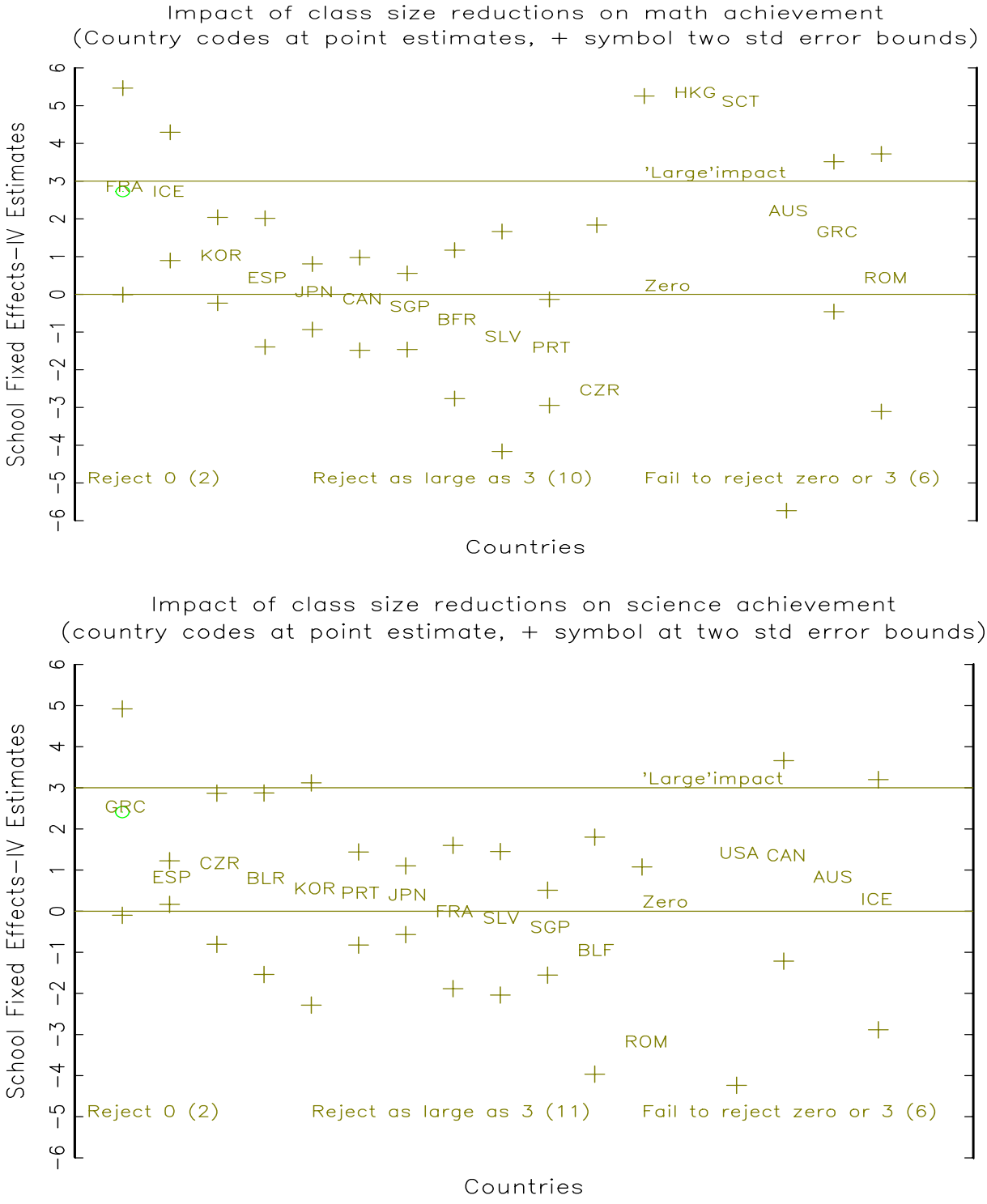
The final SFE-IV are consistent with class size impacts that are typically quite small. In the TIMSS results, the international standard deviation is 100. The estimated *median* of the coefficients suggest that a reduction in class size by 5 students (which is about a 20 percent reduction from a typical class size of 25) would increase Math scores by about one half of a point (5*.12) and increase Science scores by 1.5 points[25]. Woessman (2002) uses the calculations of Krueger (2000) to estimate that an improvement of 3 points for each one student reduction in class size would be large enough to justify the cost[26]. The results for the 36 combinations of countries and subjects break into three groups. First, there are 4 of 36 country/subject estimates that are statistically significant and of the "right" sign—but none of these has a magnitude as large as 3. Second, there are 21 of the 36 that fail to reject a zero impact but are sufficiently precise to *reject* an impact as large as 3. Third, there are 12 cases in which the estimation technique is essentially uninformative, in that the standard errors are too large to reject either a zero impact or a large impact[27].

[25] In elasticity terms, from the averages of 25 students and the mean test score of 500, a 10 percent reduction in class size would cause math scores to increase by .29 point (.06 of one percent) for an elasticity of .0059 and Science scores to increase by .90 point (.18 of one percent) for an elasticity of .018.

[26] This is likely to be extremely generous towards finding class size is effective as even with all of the assumptions the internal rate of return to class size reductions was 5.5 percent—well below most investment thresholds. Moreover, in their commentary Carneiro and Heckman (2003) characterize Krueger's cost-benefit analysis as "whimsical." (p. 345).

[27] The large standard errors are not particularly surprising as both fixed effects and IV reduce the total variation used and hence tend to increase standard errors.

Figure 11: The estimates of the size effect of class size reduction on test performance are usually positive, but quite small...
(School fixed effect, instrumental variables (SFE-IV) estimates)



Source: Based on Woessmann (2002).

Basic three: Schools only account for a fraction of variation in learning achievement. A third basic that is learned from the class size debate is that part of the reason it is difficult to establish a clear association between factors at the school level and student performance is that (a) measured learning achievement varies so much across students within schools and (b) so much of school specific variation depends on the fact that student household background characteristics vary across schools (since children with similar income and parental education tend to attend similar schools. So for instance, in the PISA examinations in Brazil 55 percent of the total variation in child performance was within schools. Of the remaining 45 percent that was due to differences in averages across schools, 25 percentage points were due to the fact that student background varied across schools. This means that the *maximum* amount of student performance that could be explained by *everything* about the schools is 20 percent of the observed variation in student scores[28]. This poses large problems for the researcher in trying to disentangle the causes of higher student performance as modest amounts of sorting by parents and students into schools can lead to large bias in estimates of the relationship between school factors like class size and learning outcomes.

Table 7: Schools account for only a small part of variance in student learning outcomes

(Results from analysis of the PISA examinations)				
	Share of total variance across students due to:			
	Col. I	Col II.	Col. III	Col IV.
	Students within schools	Total across schools	Across schools due to school student background	Maximum variation *possibly* attributable to school specific factors (col. II minus col. III)
Brazil	55	45	25	20
Russian Federation	63	37	17	20
Czech Republic	48	52	4	18
Korea, Rep. Of	62	38	14	24
Mexico	46	54	32	22
OECD average	66	34	20	14
Source: OECD (2001), Annex B1, table 2.4. Share of total variation in student test performance that is (a) school specific and (b) not attributable to student background differences across schools.				

[28] It is perhaps worth noting that at least some the motivation of the research into school impacts in the USA in the 1960s was the question of how large a role compensatory school spending in schools attended by African-Americans could play in reducing existing racial inequalities. What made the Coleman report (1966) so shocking was that he found not just that there was no association between school spending and performance—which implies that no amount of spending will equalize outcomes—but that there were no school specific effects at all. In many ways these findings have structured all future research: (a) are there school specific effects, (b) are those effects associated with observable (and controllable) levels of inputs (e.g. Hanushek 1986), and (c) are those effects associated with other characteristics of the school (e.g. vision/mission, school management practices, pedagogical practices).

Basic four: Interpreting the observed "treatment effects" for policy implications requires both a model of instruction and a positive model of producer behavior—and the usual one won't do. There are two reasons why moving from observed "treatment effects" to policy implications requires a positive model. First, almost certainly the magnitude of effects varies as the level of the treatment varies. Second, the treatment effect almost certainly varies widely from situation to situation in ways that depend on other choices of producers.

First, *if* one applied the standard economic model to schooling and *assumes* that producers are choosing inputs and production techniques to maximize learning achievement then the observed levels of class size should be consistent with equalization of marginal products per dollar. *But there is absolutely no reason to adopt this positive model of the behavior of the producers of schooling when they are in the public sector.* That is, suppose that decisions about inputs are affected by lobbying, negotiations, political pressure on the public sector and suppose that teachers are able to mobilize more effectively than the representatives of other inputs (or, more crudely put—teachers strike and books don't). Then one might expect that teaching services are over-used relative to the cost minimizing optimum allocation so that, evaluated around existing levels the marginal product per dollar will be low. The effects of class size on achievement could be non-linear so that class size reductions could be enormously important for learning starting from very high levels—but the effect tapers off and political factors may end up with class sizes that are either much too large or much to small relative to the efficient size.

Second, the effects of class size reductions are almost certainly heterogeneous in many ways: the "class size effect" is not a physical constant. For instance, the empirically observed impact of class size reductions almost certainly varies by the motivation of teachers for students to perform better. Suppose a teacher usually has 30 students but by demographic chance one year happens to have 20. If the teacher applies constant effort then one would expect higher performance. But what if the teacher reduces total effort and does not take advantage of smaller class sizes to change learning technology or provide more group work? Then the expected empirical impact of class size reductions is small. Without a behavioral model for teachers it is impossible to know what to expect.

Another example is that almost certainly the impact of class sizes depends on teacher competency and on the capability of teachers to take advantage of smaller class sizes to apply superior teaching methods. If learning achievement is low because the teacher does not know the material (as has been shown in Pakistan [MS—Sabot has claimed they tested both students and teachers]) then reducing class size is unlikely to matter much. Also, it may well be that reductions of class sizes to very small class sizes facilitates the application of "student centered" learning in which teachers tailor the teaching to the individual student's learning style. This could lead to large improvements in performance at small class sizes with very highly capable teachers but no impact of small class sizes with poorly trained teachers. So, for instance, it would not be contradictory to observe the highest quality education choosing small class sizes (e.g. expensive private schools in the USA) and still find no impact of small class sizes resulting from low attendance in rural areas of Pakistan[29]. There is no reason to believe that the existing

[29] As a final note, there is a slogan which is popular among some economists that: "one rigorous study trumps a thousand bad regressions" but which is just so deeply unscientific as to be ridiculous. A fundamental of the *physical* sciences are "invariance" principles—that physical laws are invariant with respect to "translations" (Feynman 1989). If the slogan were applied to estimates of some physical constant like the charge of the proton or the gravitational

observations on learning achievement and class size are on the "education production function" so the *empirically observed* treatment effect may have nothing to do with the *possible* treatment effect.

Opportunity 2A) Radial budget expansion

As with supply side expansion, international policy discussions of options for improving quality tend to be dominated by assumptions that expanding budgets is necessary and sufficient for quality expansion—even though no one actually believes that to be true. There are countless examples of countries' "commitment" to education measured by the share of GDP devoted to spending and the "quality" of schooling proxied by crude indicators such as class size. Economists are pre-disposed to believe that expenditures are positively related to outputs because of their positive theory of producers. If producers are maximizing cost, expenditures must expand to increase output. However, this simple theoretical intuition does not apply for the relationship between public sector schooling costs and learning achievement for just so many reasons.

This section asks the question "when and where does expanding the budget *alone*—in the absence of changes in the composition of expenditures, system reform, or demand shifts--constitute an opportunity for remedying the lack of education?" The combination of (a) the facts measured learning achievement and schooling expenditures from time series, cross-national, cross-regional, across schools and (b) a review of the educational production function literature suggest this is only rarely a promising activity.

Expenditures and learning achievement. There are four literatures that examine different aspects of the co-variation between learning achievement (almost always as measured by student learning achievement) and expenditures per pupil: over time, cross-national, cross-regional, and across schools.

Time Series. Hanushek (1995) details two major trends in schooling from the 1970s to the 1990s were rising per pupil expenditures and stagnant test scores as measured by the inter-temporally comparable, national representative National Assessment of Economic Progress (NAEP). This implies that the ratio of "test question correct per (real) schooling expenditure per

---

constant is the best estimate. But *none of this applies to schooling* and especially not schooling *undertaken by public sector, about which there is no widely accepted positive behavioral model.* Suppose one accepted the STAR experiment results that reductions in class sizes (a) to very small levels, (b) in early grades, (c) in Tennessee (with all that implies about availability of school infrastructure, class materials, quality of teachers) and about the overall quality of public sector governance had substantial effects primarily on (d) disadvantaged students. How does this empirical knowledge inform decisions to reduce class size from 50 to 40 at for median income students in grade 4 in a province of Pakistan? If one adopts equal methodological skepticism about generalizability as economists have adopted about identification then a rigorous experiment tells *absolutely nothing* outside the conditions and ranges in which the experiment is done. There is simply no scientific reason to believe the experimental STAR estimates in Tennessee or the well identified "Maimonides" rule estimates in Israel are better predictors of the impact of a reduction in class size in Pakistan than the simplest possible OLS regression done on Pakistani data. Since there are two fundamentally different reasons why the two methods might mis-estimate the "treatment effect." While a good study trumps a better study *of the same thing* rarely do we have such luxury to pick and choose and a review of the literature has to do its best to incorporate all existing findings. That said, given the bias in simple OLS regressions it does make sense to pay careful attention to identification in assessing the implications of any study.

pupil" (TQC/SEP) was falling sharply[30]. While this kicked off a sharp debate in the United States about the explanations for these trends, most of this debate focused on USA specific explanations.

Woessmann and Gundlach (1999) and Woessmann, Gundlach and Gmelin (1999) combine a clever insight with one new set of data to produce a striking finding that the TQC/SEP decline is general in the OECD and East Asian countries examined. Their clever insight was that if the USA has an inter-temporally consistent evaluation of learning achievement and if the USA has participated in multiple internationally comparable assessments of learning achievement then estimates for the evolution of the learning achievement of a large number of countries can be recovered by linking the two data sets. The new set of data are calculations of the "real" expenditures per pupil—where the deflator for schooling expenditures is based on the price of "productivity resistant services" so that these real expenditures already take into account that the costs of services rise relative to manufactures because of slow productivity growth (the Baumol effect). The striking finding produced is that TQC/SEP decline is not a US specific phenomena, rather the opposite, the fall in the USA is second smallest of all OECD countries.

As shown in table 7 the rise in real per student expenditures (where "real" uses a services deflator that already accounts for a general service sector specific Baumol effect) is enormous. At the same time, no country shows a large increase in measured learning achievement. By these estimates math and science learning declined modestly in France—while expenditures *tripled.* Although, as discussed above, Japan leads the USA in the *level* of learning achievement the *gap* has not widened, which implies the *growth* of achievement is about the same—even though expenditures per pupil have doubled in Japan.

[30] While some would refer to this as a "productivity shock" this presupposes that the test measure the desired "product" of schooling—which is far from obvious—so I will use the deliberately awkward TSC/SEP to be clear this is a descriptive statistic which does not pre-suppose any given explanation nor normative conclusion.

Table 8: Real expenditures per pupil have risen enormously while measured learning achievement has stagnated…

Country	Estimated change in the assessment of math and science learning achievement, 1970-1994	Estimated change in real expenditures per pupil, 1970-1994.	Estimated change in "test questions correct per school expenditures per pupil"
Sweden	4.30%	28.50%	-23.20%
United States	0.00%	33.10%	-33.10%
Netherlands	1.70%	36.30%	-34.10%
Belgium	-4.70%	64.70%	-72.80%
United Kingdom	-8.20%	76.70%	-92.50%
Japan	-1.90%	103.30%	-107.20%
Germany	-4.80%	108.10%	-118.60%
Italy	1.30%	125.70%	-122.80%
France	-6.60%	211.60%	-233.70%
New Zealand	-9.70%	222.50%	-257.20%
Australia	-2.30%	269.80%	-278.50%

Source:: adapted from Woessman (2002), tables 3.3, 3.4

While the time series evidence is specific to the OECD—and a few high performing East Asian countries—the general point that expenditures per pupil *can* expand enormously with very little impact on measured learning achievement is likely true elsewhere as well.

Cross-national/cross regional. There are large differences in the performance of countries on the international examinations but very little of that variation is associated with differences in expenditures per pupil--or any other physical indicator of educational systems such as pupil-teacher ratios or teacher wages—either unconditionally or after conditions for basic background factors. Hanushek and Kim (1995) and Hanushek and Kimko (2000) find no statistically significant correlation of test scores and either pupil-teacher ratios, current expenditures per pupil or expenditures on education as a fraction of GDP. Barro and Lee (1997) also conduct cross-national estimates of the correlates of cross-national examination performance and find that the pupil-teacher ratio, expenditures per pupil or average teacher salary are not correlates of math or science score performance, while for reading scores there were positive associations with pupil-teacher ratios and teacher salary but not with education expenditures per pupil.

There is also very little connection between spending and enrollment and attainment outcomes. Just to mention two of many, Jayasuriya and Wodon (2002) find no connection of net primary enrollments and education expenditures per capita (controlling for adult literacy), Roberts (2003) finds no connection between gross enrollment rates or primary completion rates and either primary spending as percent of income or primary spending per pupil as a fraction of per capita income.

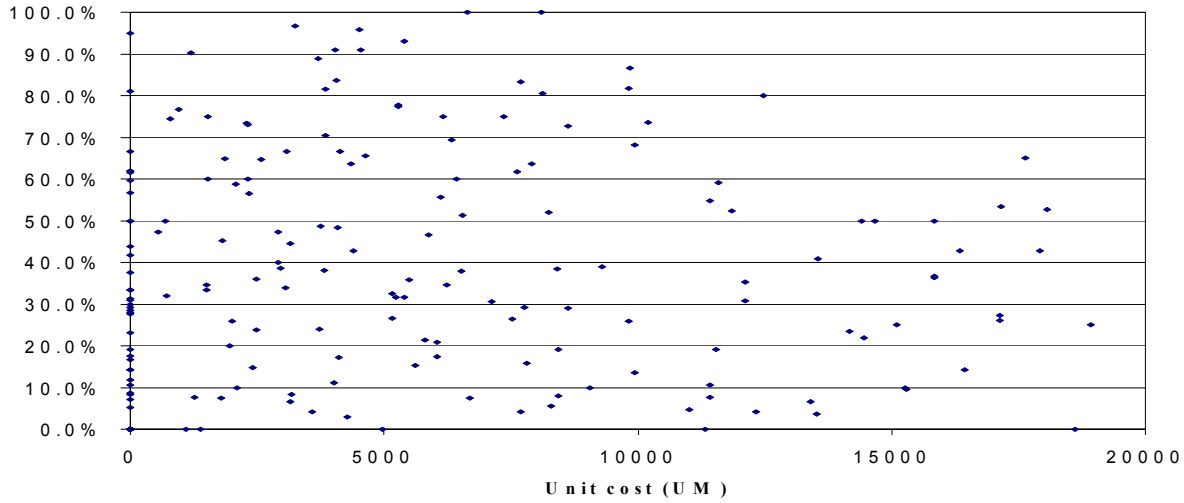
Of course the methodological problems are so overwhelming that it would be hardly worth mentioning these type of results if these same crude indicators that show no correlation with learning outcomes (e.g. 'class size' or 'spending per pupil') were not constantly used as if they were reliable proxies for school quality.

Cross schools. Woessmann (2002) uses the massive (over a quarter of a million observations) dataset from the TIMSS to examine child, school, and country specific factors associated with performance. He finds that *none* of the standard measures of school inputs—

expenditures per child, class size—are associated with improved performance. There are an enormous number of studies that relate performance to school level factors—almost none of which find that expenditures are an important independent determinant of school performance (particularly once student background characteristics, which tend to be strongly correlated with school expenditures are controlled). On one level it is hardly worth saying that *total* school expenditures have little correlation with performance in public sector systems, for three reasons. First, schools are not managed at any level to think of cost-effectiveness as a performance measure—it is actually rare for schools to even *know* unit costs, much less manage to minimize the cost of a given performance (or maximize performance for a given cost). Second, in standard public sector systems individual schools do not typically have control over their total expenditures, which are mainly driven by teacher wages, that are typically completely beyond the managerial control of the schools themselves. Third, since student background accounts for so much of the variability in performance the school effect itself is typically quite small, the amount that can be empirically attributed to any school specific features at all necessarily even smaller, and the fraction of the variance attributable to one factor like total expenditures smaller still.

Figure 12: Primary school success varies hugely despite similar unit costs

(Pass rate (percent) at the end of primary education exam (vertical axis) versus unit cost (horizontal))



Source: Mingat (2003).

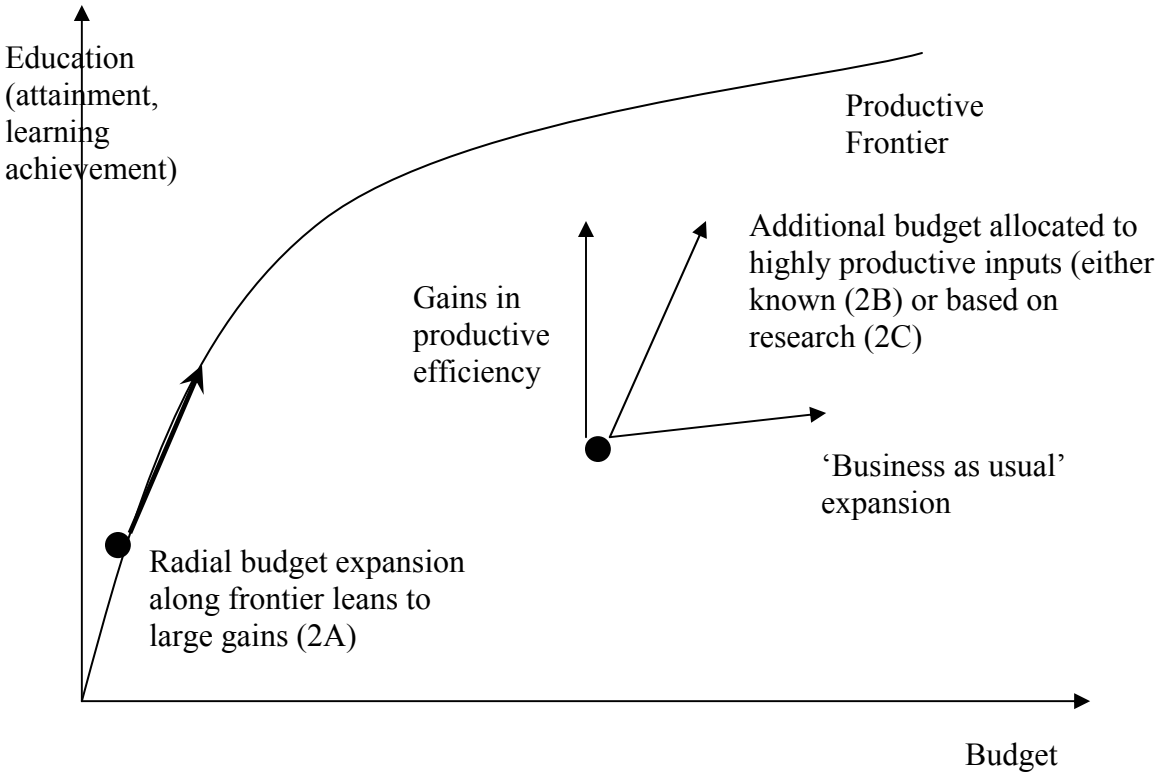
Equalization of marginal learning achievement gain per dollar spent across inputs. The strongest case for a budget expansion is if schools are already both productively efficient and have allocate their expenditures efficiently. Otherwise, it makes sense to allocate any available incremental expenditures to inputs with higher marginal product per dollar.

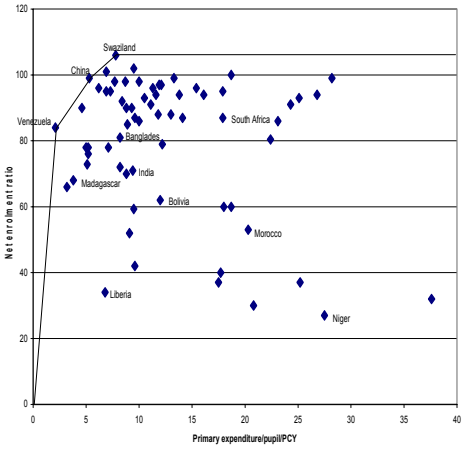
Table 9: Implications of productive efficiency in the allocation of inputs		
Implication of allocation of inputs for productive efficiency	Source of test	Findings [Caveats]
Marginal product per dollar should be equalized across all inputs	Rigorous evaluation of interventions	No rigorous evaluation in a developing country has failed to reject equal marginal product per dollar. (e.g. Hanushek and Harbison 1992 in Brazil) [So few have been done]
	Cost-effectiveness calculations from econometric studies	Typical finding is that the marginal product of teacher related inputs is small and non-teacher related inputs have cost-effectiveness between one and three *orders of magnitude* higher. (e.g. [Identification across different inputs difficult to impossible]
	Non-experimental variation, "confirmation frequencies	Reviews of the literature that do counts find that measures of infrastructure adequacy and availability of learning materials are much more likely to have "right sign/statistically significant" associations with learning achievement than expenditures, class size, teacher wages, or teacher formal qualifications. (Hanushek 1986, Fuller and Clark, 1994) [Most of the studies reviewed had no plausible identification strategy]
If resources are used efficiently then only total expenditures should matter—not sources	Variations in degree of finance from local sources	Financing from local sources has more impact on performance than centrally provided financing. (e.g. James et al. 1996 in Indonesia, Birdsall and Orivel 1996 in Mali). The composition of inputs financed by central and local sources differs systematically—with central sources providing teachers and local sources providing other inputs. [Inputs from communities could easily be correlated with other community characteristics that promote high quality learning]
If resources are used efficiently in both public and private sector then input allocation should be equal and performance per dollar equal.	Comparisons between private and public schools	Even after controlling for selection effects private schools tend to produce higher learning achievement with less resources. (e.g. Jimenez and Lockheed 1995 On five countries) Public and private schools tend to have very different input mixes. (Alderman et al. 1996 in Pakistan). [Very difficult to control for selection of students into private or public schools]
Based on Filmer and Pritchett (1997), tables 1-6		

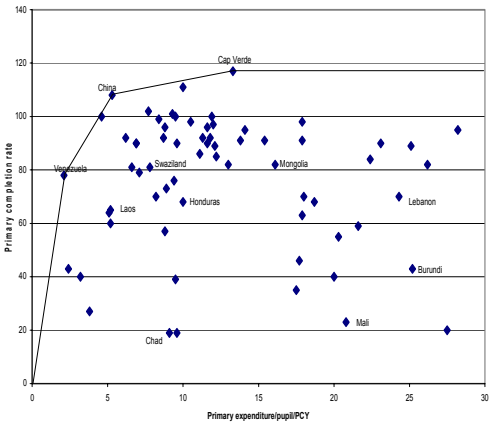
Summary on budget expansion. There is an obvious contradiction between the surface interpretation of the rhetoric in the development community and the existing evidence. The conventional wisdom seems to be inputs, inputs, inputs: a country's "commitment" to education can be judged by *spending* (as a fraction of GDP or total government expenditures); educational

spending should be "protected" during budget adjustments; debt relief is important because it will allow more *spending* on education; international goals for education are crucial because they will allow more donor finance for *spending* on education. Yet I say this "seems to be" the conventional wisdom because my perception is that *no one* actually engaged with education actually believes that spending *per se* is sufficient to raise the level of education. Everyone is aware that *none* of the available evidence supports the view that the major, or even a sizeable portion of the existing variation in schooling performance (either in attainment or learning achievement)—across time, countries, regions, schools—is accounted for by differences in spending. *No study* has ever tested any implication of the hypothesis that public sector production is productively efficient and failed to reject it. There is agreement that in many countries they are far from the "frontier" in either productive or allocative efficiency.

Figure 13a,b,c: If countries are on (or near) the efficiency frontier and spending is low then radial budget expansions can have substantial impact…but if countries are not efficient then "business as usual" expansion are of indeterminate impact and improvements in productive or allocative efficiency are more promising…







Souce: Robert 1993.

The current conventional wisdom among practitioners is that additional spending will be necessary, but that it will only service to raise attainment and learning outcomes if:

a) it is devoted to categories of expenditure that are highly productive—including either better productive efficiency or better allocation across inputs (opportunity 2B),

b) if it is devoted to, or combined with, *new* techniques and pedagogical approaches (opportunity 2C),

c) it is accompanied by systemic reform.

Opportunity 2.B) Quality improvements, supply side: Expanding chosen known specific elements

I will devote only relative short space to what most might regard as the single most important element of improving education: identifying and expanding the particular actions that lead to increased learning achievement. That is, while most will admit that budget expansion will not necessarily improve performance, nearly everyone has a "favorite" input, action, or activity that *if the incremental resources were devoted to that activity* then budget increases would lead to greater attainment and learning achievement. There are three classes of inputs that are frequently identified as promising in the truly voluminous empirical literature relating measures of student learning to inputs (see reviews of the 'education production function' literature by Fuller and Clark, Hanushek, Villegas and a slightly different tack on the literature in Pritchett and Filmer 1997).

- *Instructional materials*. Many studies find that some measure of instructional materials at either the class room (presence of absence of textbooks, for instance) or school (having a library) is associated with higher test scores. Many education projects and programs build on those findings and focus on increasing those identified inputs.
- *Key Infrastructure.* Many studies find that some measure of infrastructure (e.g. having desks, having latrines, having running water, having windows) is an important determinant of progress. Again, many education projects and programs are structured around not just school construction but school upgrading of one type or another.
- *Teacher training.* Many studies find that some measure of teacher knowledge of the content and/or pedagogical practices is associated with higher test scores. Nearly every educational project or program has some component for pre-service or in-service "teacher training" that attempts to improve either teacher content knowledge.

Conclusion on expanding known quality improvements. I will not even attempt to review this literature, rather I will go straight to four points.

First, almost certainly there are interventions in each of these dimensions, which, *if successful in their intermediate objectives (e.g. raising textbook availability, improving teaching) will have substantial, cost-effective, impacts on attainment and learning achievement.*

Second, what the particular interventions are will vary across countries, regions, schools, etc. Since the "learning achievement" function is so complex and depends on the presence of a variety of inputs the impacts will vary.

Third, while "associations" are easy to establish using the data at hand the question of *causal* impact has been underplayed and the findings from studies using randomization find that the usual methods (OLS or HLM) would produce estimates that were quite wrong as predictors of program impact (see below on Kenya). This goes back to the issue that, since most of the observed variation in the data is the result of choices made by parents or producers, without a positive model and "identification" the associations are consistent with a very wide range of underlying causal models. For instance, lack of key inputs may just be a proxy for a dysfunctional school situation. Take the example of textbooks. Suppose in the data there is a strong correlation of textbook availability and scores in a cross section of schools. This could be because the schools that did not have textbooks were dysfunctional in a variety of other ways. If this is the case then a program that gave textbooks to all schools might have no impact on performance—even if it saturated the school with books--in spite of the strong positive association of textbooks and performance--because the incremental textbooks would all be used in a still dysfunctional school and hence might have little or no impact.

Fourth, the key pragmatic objection is whether programs designed at eliminating "bottlenecks" to learning by devoting incremental resources to high productivity activities can be successfully implemented. For instance, the world is littered with projects and programs for teacher training that had no impact on subsequent teacher classroom practices (see Box 1 for a particularly depressing example from Pakistan).

Box 1: The dismal state of teacher training in Pakistan, early 1990s.

Teacher training in this province is a mockery. We should close down the teacher training institutes and stop this nonsense. I have been teaching in a B.A./B.Ed. program for many years and see no signs that I have any impact on the students I teach.

—A university education instructor quoted in Warwick and Reimers (1995).

Most inmates of this system [two teacher training institutes] have no respect for themselves, hence they have no respect for others. The teachers think the students are cheats, the students think the teachers have shattered their ideals. Most of them are disillusioned. They have no hopes, no aims, no ambitions. They are living from day to day, watching impersonally as the system crumbles around them.

—Nauman (1990).

A national survey of Pakistan's primary schools suggests that these anecdotal accounts are only too true. Survey data on teaching practices "provide no basis for statements that . . . teacher training makes a substantial difference to how teachers teach." A 1998 study of teacher training suggests that "staff and faculty are professionally untrained, political interference is common, resources and facilities are poor and badly utilized, motivation and expectations are low and there is no system of accreditation to enforce standards." Embedded in an education system that was fundamentally unaccountable and lacked any outcome orientation, teacher training reflected worst practice.

Sources: Adapted from World Bank 2004, Warwick and Reimers (1995); Kizilbash (1998), p. 45.

In examining the claims that incremental spending on this or that activity would increase scores enormously, one needs to ask the question "then why is this not happening already?" Convincing evidence that some inputs have incredibly high marginal products relatively to other inputs is at the same time convincing evidence that existing producers are not efficient (since the implication of efficiency is equalization of marginal products per dollar). But if the system is such that there are no built-in tendencies towards

productive efficiency then why will a particular intervention to raise efficiency *with no changes in systemic incentives* succeed? I am not arguing that such a case can *never* be made—there are numerous successful piecemeal initiatives to act as counter-examples--but it cannot simply be assumed that piecemeal policy action can be effective as the many failures attest.

Opportunity 2.C) Quality improvements, supply side: Experimentation with rigorous evaluation

A third opportunity is to improve the quality of schooling through empirical research which investigates the impact of interventions, better establishes the determinants of learning, and evaluates new techniques for raising learning. If successful in raising quality, this would increase the learning achievement of students currently enrolled and likely increase retention of students who now enroll but abandon early.

Rigorous evaluations of the impact on learning achievement of a variety of actions—both the proximate determinants of learning and the institutional/incentives determinants of the proximate determinants—are notable by their almost complete absence in nearly all countries have been. Work has begun to apply randomized evaluations to educational interventions, including a series of experiments in rural Kenya led by Michael Kremer involving the effect of subsidies (free uniforms), text books, teacher incentives, class size, health interventions (deworming). These evaluations confirm almost none of the "conventional wisdom" about education. In fact, in a paper that summarizes almost a decade of experience Kremer concludes that (a) fees are too high, (b) class sizes are *too small*, (c) school 'choice' while practiced by parents has perverse incentives for headmasters, and (d) most programs that provide additional inputs for "poorly performing" schools are wrong-headed—and that all of these outcomes are *endogenous* to the institutional conditions for the production of education in rural Kenya (Kremer, Moulin, and Namanyu, 2003).

Evaluations often find that what "everyone knew" would work—doesn't. They also sometimes identify actions that have enormous impacts. A recent randomized evaluation of a tutoring program created by a local NGO and targeted to students who are lagging behind in early grades is enormously effective and cost-effective (Banerjee, Duflo and Linden 2003). In addition to evaluating specific techniques, there have been increasing attempts to investigate the impact of other innovations such as greater school autonomy in Nicaragua (King and Ozler 1998), 'vouchers' to secondary school students in Colombia (Angrist, Bettinger, Bloom, King and Kremer 2002).

Not all evaluations need be randomized to be rigorous. The key is to have a research method that is capable of identifying the true causal impact of a particular intervention or policy action. The comparison of "before and after" is simply inadequate to estimate the "with and without" impacts that are needed[31].

[31] There is an ongoing debate among US academic economists. One group seems to believe the identification problem is so severe that essentially only randomization can produce clean estimates as otherwise the "instruments" will be weak or strong "structural" assumptions will need to be made. Another group believes the estimate of structural models is both possible and essential to understanding the economics of the intervention in any case.

Box 2: Randomized experiments in schooling: Busia district in Kenya

Since 1996 a group of researchers has been working with a Dutch nonprofit (International Christelijk Steunfonds) supporting schools in rural Kenya to estimate the impact of various interventions. Through random selection some schools were chosen to implement the interventions first, with the other schools to follow. This allowed the researchers to test a number of ideas.

Textbooks. Everybody knows that textbooks are important and that their lack is a major constraint on effective instruction. Yet the first study found "no evidence that the provision of textbooks in Kenyan primary schools led to a large positive impact on test scores, nor is there any evidence that it affected daily attendance, grade repetition, or dropout rates." Does this mean that textbooks don't matter? No. Although textbooks did not increase the performance of the *typical* (median) student, they did improve performance for students who did the best on the pre-test. This suggests that because the textbooks in this instance were too difficult for the typical student, the books did not matter.

Teacher incentives. Everybody knows that teacher incentives are crucial since teachers are under motivated. Yet a study on incentives for teachers based on student test scores found that "teachers responded to the program primarily by seeking to manipulate short-run test scores. . . . [T]eachers' absence rates did not decline, homework assignments did not increase, teaching methods did not change." Does this mean that teacher incentives don't matter? No. Teachers did change their behavior—they "conducted special coaching sessions and encouraged students to take the test." This suggests that you get what you pay for—whether you like it or not.

Deworming. Deworming does not feature widely in the education effectiveness literature. Yet a randomized trial of an inexpensive medical treatment for hookworm, roundworm, whipworm, and schistosomiasis found that it reduced absenteeism by a quarter. Does this mean that health is all that matters? No. While attendance improved, test scores did not.

Three observations. First, things that everybody "knows" to be important did not work as planned, whereas the intervention with lower expectations had large impacts. Second, these results from a hundred schools in an isolated area of Kenya have been getting enormous academic attention because there are so few rigorous, randomized evaluations of schooling interventions. Third, the findings from each intervention do not reveal universal, immediately generalizable results, but they reveal that specifics matter and that learning about what works needs to be local to be useful.

Sources: Miguel and Kremer (2000); Glewwe, Ilias, and Kremer (2000); Glewwe, Kremer, and Moulin (1997)

There are, however, four questions about the use of rigorous evaluation as a key to improving education: two methodological and two about impact. First, there is a methodological problem even with "randomized" experiments, which is the lack of a true "double blind" control that allows for a control of "placebo" or "Hawthorne" effects, which tend to find false positives of *any* intervention[32].

[32] There are many ways in which 'false positives' could be produced, particularly in complex experiments that alter incentives (e.g. school autonomy). For instance, if talented managers are attracted into managing the experiments they might also, in the course of implementing the experiment, have impact on improving the "treatment" schools through their contacts. Or, as Hoxby argues may be true for the Tennessee STAR class size experiments, teachers/administrators might make extra temporary effort to make an experiment succeed if its success serves their long-run interests.

Second, the results of randomized evaluations cannot be applied with any confidence outside the range of the experiment. If the effect under study is plausibly homogenous across time and space, and one rigorous study can be extrapolated to a variety of contexts, then the benefit/cost ratio for evaluation is enormous. If however the impact of specific interventions varies widely from place to place, then the benefits of any given study are smaller, potentially dramatically smaller. The mixed evidence reviewed above, about one seemingly simple determinant of learning—class size—suggests large heterogeneity in the effects of class size. If reducing class size is essentially a different "treatment" depending on a host of other factors-teacher quality and motivation, availability of complementary inputs, etc. then the cost-effectiveness of evaluation is more difficult to defend.

Third, the principal costs of randomization are political organizational risk. Few governments have proved willing to participate in randomized experiments. Nearly all of the experiments to date have been carried out either with NGOs (e.g. Kenya, India) or implemented outside of mainstream agencies (e.g. Colombia). Moreover, reform agendas tend to be promoted by those who are truly convinced of the desirability of their idea and hence hesitate to run the risk of being disproved[33].

Fourth, the question is the diffusion of information and adoption and the fact that knowledge itself is potentially endogenous. If what limits adoption is not knowledge or even the perceived reliability of knowledge, but rather deep incentive problems then the benefit/cost ratio of new knowledge conditional on an endogenous lack of adoption, is small[34]. Many observers have concluded from the observation of many waves of reform initiatives that knowledge is not the problem. Allow me to quote at length a passage from the introduction to a excellent introduction to education reform in Africa:

> *There is a particular irony to education reform. Pockets of good education practice (such as enlightened and effective classroom management, novel curricula, and innovative instructional technologies, many of them cost-effective) can be found almost anywhere, signifying that good education is not the result of arcane knowledge….Yet the rate of uptake of effective practices is depressingly low. As a result, these innovations exist on a very small scale and many soon sputter out….Most projects that introduce innovations are in one sense or another meant to be demonstration projects. They are supposed to yield and disseminate practical information about good pedagogical practice. Yet reality contradicts the "information" assumption on which such projects are based. Effective practices are often found in areas far from the world's information centers. And effective schools are often found just a few blocks from dysfunctional ones.* Crouch and Healey 1997.

[33] See Pritchett (2003) "It Pays to be Ignorant" for a simple political economy argument explaining the lack of rigorous evaluation.

[34] The constraints to adoption can result in ideological arguments deeply tied up with interests as well. There are some who argue it is no coincidence that "progressive" notions of how to teach get interpreted by many teachers and teacher organizations in ways that imply less individual accountability for effort and results by teachers.

This leads to the next section, "systemic reform" as it suggests the real problem is the *motivation* to know.

Box 3: Randomized experimentation on radio instruction

Randomized evaluations are not a new technique. Jamison et al. (1981) report on an experimental study of the impact of textbook availability on mathematics achievement of students in Nicaraguan first-grade classes. The evaluation compared control classes in which textbooks are relatively rare with two treatments: textbooks alone and with textbooks plus a radio-based instructional program. The evaluation revealed that while both the textbooks and the radio treatments had significant effects on achievement, the radio group improved at a substantially higher rate than the textbook group. Since the incremental cost of the radio instruction was low this was an enormously cost-effective innovation.

What was the impact of this rigorous demonstration of a cost-effective technique? While there was a wave of adoption of radio based instruction in some countries in the 1980s it has not become widespread and Glewwe (2002) doubts whether the results of randomized evaluation played any role.

Interestingly, Jamison originally predicted that the results on textbooks proved more popular as they were easier to implement logistically. In addition, the political economy is more favorable to textbooks, which are clear complements to teacher inputs than to radio instruction, which is a clear possible substitute.

Summary on evaluation. It is almost certainly the case that the amount currently invested in rigorous investigation into the impact of educational projects and into the conditions for promoting learning is far less than the "optimal." Since the amounts of money involved in rigorous research are typically small (relative to budgets) and since the potential returns in improvements are large expanding rigorous impact evaluation is a promising opportunity. In particular, there is no reason to not demand much greater attention to impact evaluation in donor financed projects. In particular the problem is to insist that attention be paid to the counterfactual—while nearly every donor financed project has "monitoring" and nearly every project now does some impact assessment—they are almost always simply estimates of outcomes "before and after" rather than a serious attempt to measure "with and without."

The opportunity is mainly limited by two factors. First, whether the results will actually change actions—obviously the returns to rigorous evaluations which are ignored in future planning is near zero. In many countries of the world the slow diffusion and adoption of innovations is due to existing institutional incentives and hence is a reform issue and without the fifth element of system reform the pay-off to increased knowledge will be small. Secondly, careful attention to the viability and 'scalability' of the proposed intervention—again, obviously devoting expensive and scarce high quality researcher attention to investigating 'pilot' programs that cannot be replicated to scale is a waste of resources.

Conclusion on supply side policy actions to improve school quality. This long and complex section is perhaps best summarized as a decision tree, as illustrated in figure 14.

- Are producers already optimizing and is the marginal product high?

If this is so, then key to improving education will be to mobilize financing to *increase expenditures per child* in roughly the way that expenditures are currently allocated. Personally, I

judge this state of affairs to be extremely rare precisely in those situations in which education is the most lacking. In fact the on-going empirical debate is not so much whether countries are at, and would move along the production frontier (for which there is no evidence), but whether a "business as usual" expansion in expenditures (e.g. class size reductions, higher teacher wages with same structure of incentives, modestly more inputs) would have roughly zero impact (my view) or a positive impact that is high enough to be justified.

- Are there inputs with higher learning achievement gain per dollar in existing conditions?

And

- Can these be implemented in the existing institutional structures?

If this is so, the key to improving quality is to adopt better *expand expenditures on cost-effective inputs*—those that improve productive efficiency (e.g. teacher training) or allocative efficiency (e.g. more spending on instructional materials, key infrastructure deficits).

- Are there promising alternatives to be evaluated?

And

- Would demonstration of the superiority of these techniques lead to their adoption?

If this is so, then *rigorous research into alternatives* is key.

- Is the lack of productive and allocative efficiency the *endogenous result* of the incentives created by the existing institutional organization of the production of schooling?
- Is the lack of widespread adoption of existing innovations and the lack of rigorous research into alternatives and the lack of rigorous impact evaluation the *endogenous result* of existing incentives for the producers of schooling?

If the latter is the case then the only option is some type of *systemic reform*—which could be either entirely within the public sector or could involve non-public sector providers either on their own or with government support via contracts or demand side financing[35].

---

[35] I would like to quote at length from a comment from one reader of an earlier draft, as it is a succinct expression of a valid objective to a bandwagon of randomized evaluation:

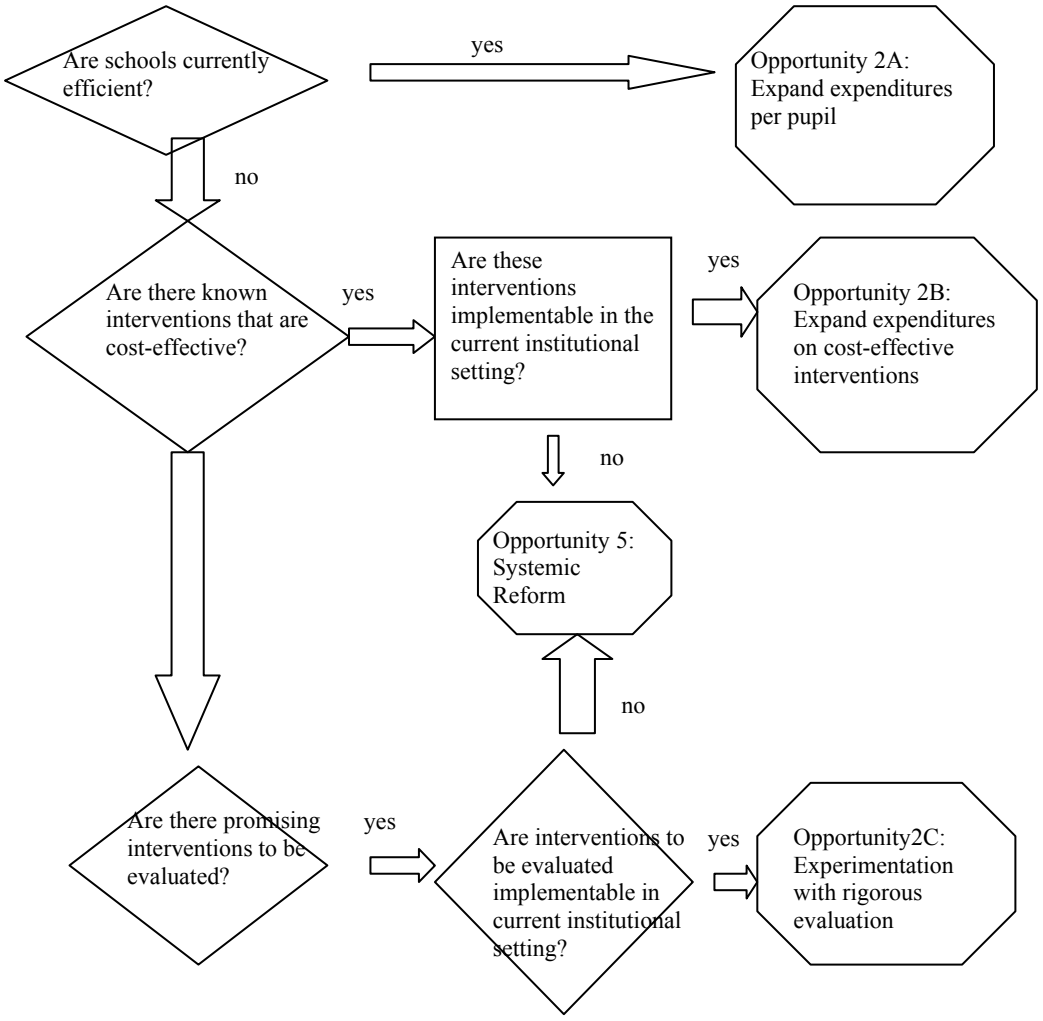
> I just don't think economists should be production engineers for state monopolies. The idea behind all this stuff is that if we can discover the "magic bullets" that work, then we can make *everyone* take the magic bullet. But that's a really, really naïve view of how anything, at any point in history, has ever improved. In fact, things improve because of Schumpeterian innovation and destruction. Even good bureaucracies improve that way. Good bureaucracies, including effective private ones like McDonald's observe what effective real administrators do, and then force others to do it. The market does it via the pressure of competition, and because in good markets entrepreneurs can imitate the technology other entrepreneurs discover (including the organizational technology). We tend to forget the Econ 101 lesson ignore that the information that has to exist is not just about the quality of what is produced, so that consumers can make wise decisions, but about production technology so that entrepreneurial rents are bid away. Note that you have to have a real feel for the actual technology and human interaction for how this happens. Entreprenerial information is often diffused through employees. They often do this by poaching employees from each other. Observe how technique (say, a better filing system, or a better way to make proposals) spreads from firm to firm, and it is by the interchange of employees. Production technologies in one firm are always set up by someone who either discovered it, or, more often, brought it from another firm. Anyway, a good (public OR private) bureaucracy forces principals to adopt practices by laterally

Again, this is the essence of the current debate. The debate is not principally about the *proximate determinants* of learning—nearly everyone agrees that exposure to school is essential and there is broad agreement on the fundamentals of learning while in school. The debate is not even principally about what policy actions by producers of education that *could* lead to improvements—nearly everyone agrees that better teaching practices and better instructional material inputs *could* lead to better outcomes. The debate is also not about whether it is possible for public sectors to produce high quality schooling—*of course* they can, and do. The key question is whether in *any given country* the current institutional conditions lead to a public sector that is *motivated to* and *capable of* implementing the required actions.

My personal view of the decision tree in figure 14 is that in nearly all developing countries the path through the decision will lead to systemic reform. The existing failures are almost entirely the endogenous outcome of the existing incentives of the public sector (which is unconcerned about quality and unresponsive to citizen demands for higher quality schooling) and of public sector producers who are given no support in increasing quality.

Figure 14: Decision tree for evaluating the supply side policy actions for improving education

disseminating what it observes works. Not by making grandiose experiment to discover what works. And a good bureaucracy ratchets things up by allowing enough mavericks to do things a little differently, so that they can learn from them and spread it to the non-mavericks. But some bureaucracies get really good at what they do, and the norms are really effective, so they ossify. Happens all the time, which is why firms go broke. Schumpeterian destruction. In my view, the public sector, and education, need to work more or less this way, and that is how in fact they do work, when they work well. This randomized experimentation stuff seems, to me, to ignore how social change and social improvement works, and hence is sort of barking up the wrong tree. In short, if experimentation and then Stalinist diffusion of "what works" could work in education, it would have worked for Stalin in tractor production. It didn't."



However, I will be the first to admit there are many informed and well-meaning analysts who disagree and believe that policy actions—to improve teaching practices (content knowledge and pedagogy), to raise the quality and availability of instructional materials, to improve infrastructure are possible within the existing institutional structures for the production of schooling. Moreover, the problems with schooling are so pressing and the damage to children of not having access to quality schooling are so irreversible it is very difficult to "give up" and simply do nothing even in environments which all acknowledge are not promising—e.g. Nigeria.

This debate comes back to the key question of a *positive* model of public sector producers. Everyone who advocates the adoption of some policy action that promises to dramatically improve educational outcomes should be able to answer the question: If what is being proposed has the benefits claimed then *why* is the public sector currently not doing it *already*?

Opportunity number 3: Demand side: Raising the benefits of schooling.

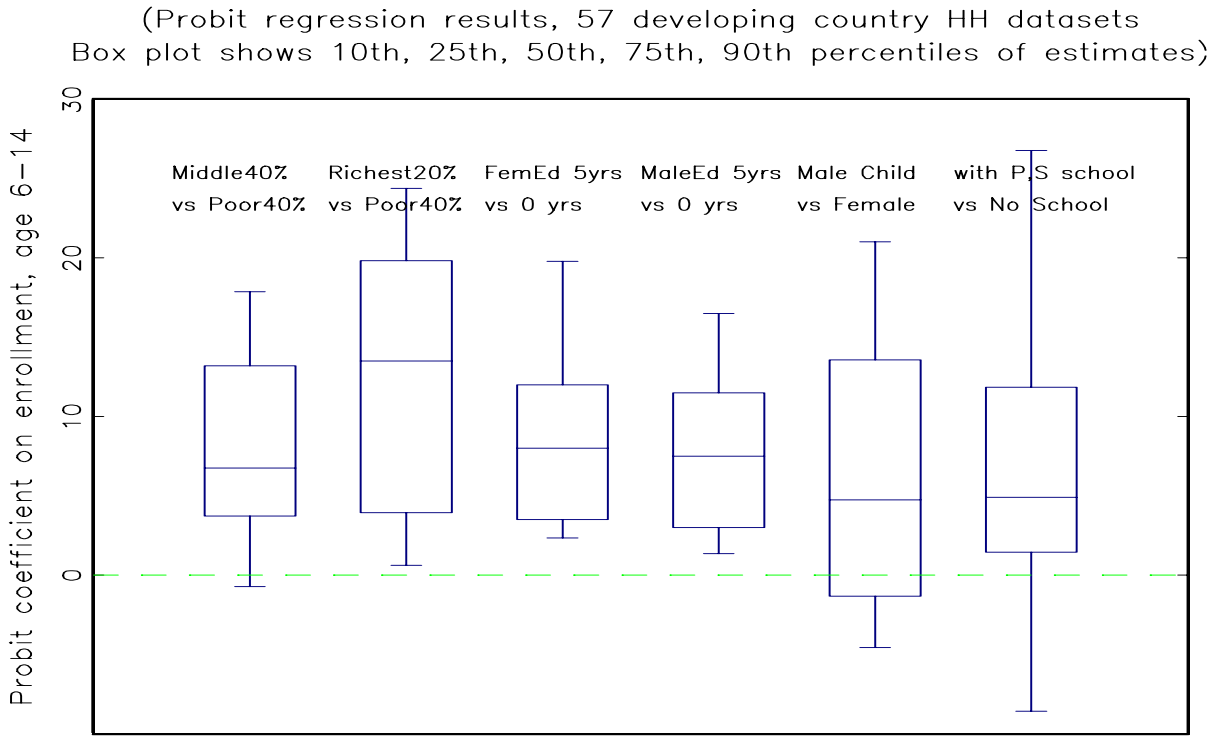
There are two key opportunities to address the lack of education that are indirect, but which may be far and away the most powerful empirical correlates (and perhaps determinants) of increases in education. First, since the most important determinants of a child's educational attainment appear to be household income and parental education, generalized increases in productivity and/or income—from whatever sources—that raise the incomes of parents will increase schooling, and increases in education will beget future increases in education. Second, are policy or institutional reforms that raise the returns to education. However, while there are, in all likelihood the most powerful determinants of education I will not explore them in much depth because they are not education specific.

3.A) The effect of income on education

Filmer (2000) uses the DHS datasets to investigate the factors associated with child (age 6-14) enrollment. Using a probit regression that controlled for child age and age squared, urban versus rural residence, wealth of the household, education of the adult females and males in the household, and child biological sex, he finds that the effects of parental income and parental education are consistently empirically large and nearly always statistically significant. The median estimated marginal effect of living in a middle wealth versus poor household is 8.8 percentage points, richest 20 percent versus poorest 40 percent had a median 13 percentage point effect, having a mother with 5 years of schooling versus no years had a median impact of 8.0 percentage points, while having a father with 5 versus zero years of schooling had a 7.5 percentage point effect. One would expect a child from a rich household with father and mother with 5 years of schooling to have an enrollment rate 28 percentage points higher than a poor child with parents with no education. The median association with having both a primary and secondary school in the cluster is 6.9 percent.

These strong correlations at the household level are also apparent in cross country comparisons and in comparisons over time. The same studies that find only weak impact of public spending on enrollment and achievement outcomes tend to find quite strong correlations with adult education/literacy and per capita income. Similarly studies that trace out the impact of income changes over time find a large role for improvements in income. One example is Edmonds (2003) examination of the of rapid growth in Vietnam in which he finds that 80 percent of the reduction in child labor observed in households that emerged from poverty can be accounted for by increases in per capita expenditures (as a proxy for income increases).

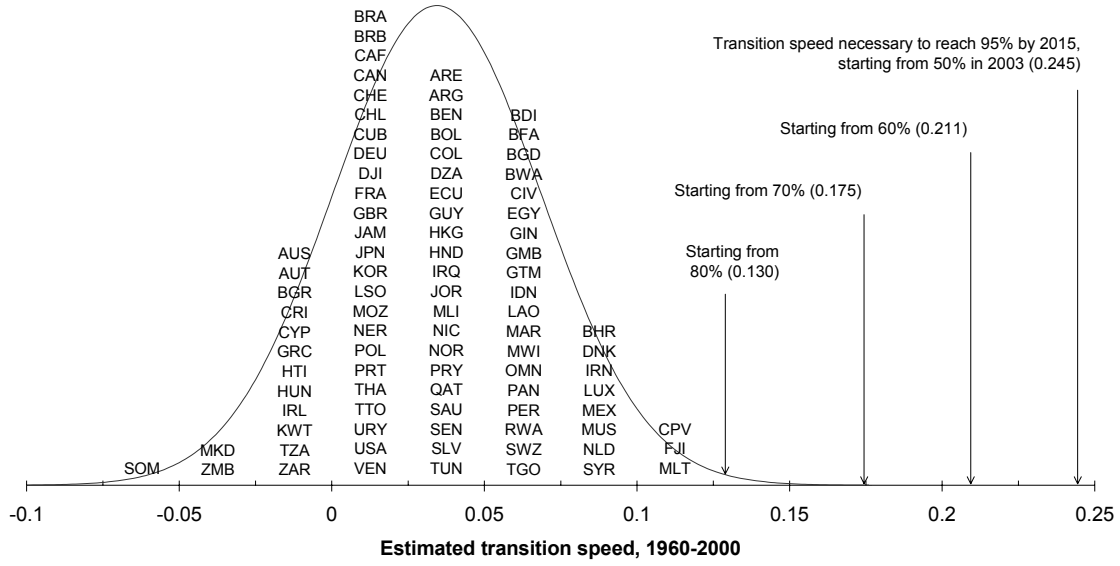
Figure 15: The estimated association in household datasets between child enrollment and household wealth and parental education is consistently large and significant… while the association with school availability is much less consistent



Clemens (2004) in a recent paper emphasizes that in spite of decades of focus on the "supply side" of public action the primary determinant of progress towards universal education has appeared to be principally demand determined and that very little that countries have done on the supply side has significantly altered progress. One implication of this is that since the demand side factors—particularly parental education—tend to evolve slowly in most countries the transition from modest levels of education (50 percent primary enrollment) to near universal has been steady but slow[36]. In fact, he points out that the rate of progress necessary for countries to meet the currently Millennium Development Goals by 2015 are wildly beyond the bounds of any observed rates of progress.

[36] One factor that I don't discuss, but which almost certainly is a large factor is demography. Some simple analysis suggests a strong correlation between total educational achievement and age dependency (Crouch, personal communication) which is plausible given the simple reality is that if you have lots of kids per adult, they are harder to educate. As an independent "opportunity" controlling demography does not come onto the list.

Figure 16: The distribution of the observed transition speeds to universal primary enrollment shows that even the most rapid are far slower than the rates needed the meet the MDGs in 2015 in countries starting from anything less than 80 percent enrollment today



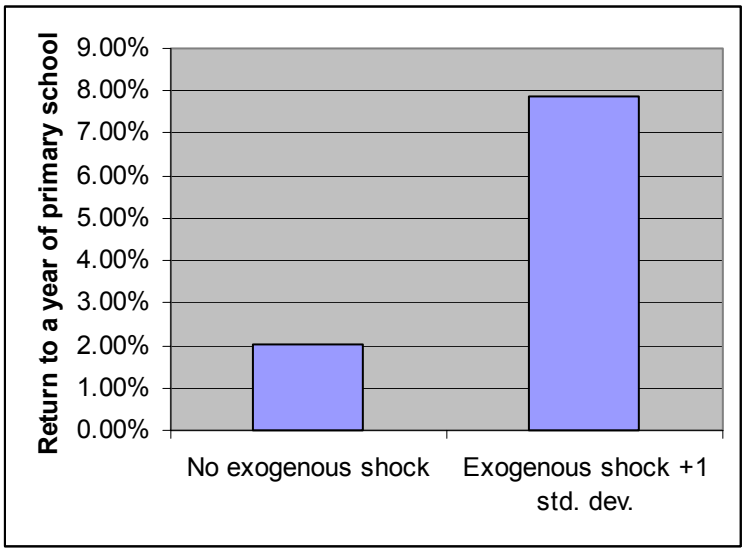
3.B) The effect of raising the returns to education.

There is no question that returns to education vary widely and that in stagnant technological and economic conditions the returns to education can be very low—sufficiently low that *if* current labor market conditions are expected to persist then *not* attending school may well be the *optimal* decision from a narrowly drawn economic calculation. I would ever recommend adopting this narrow view as the view that basic education is a human right and an essential element of well-being in the modern world is pretty compelling. But it is worth pointing out that when returns are low the economic calculation and the human rights desire may be at odds.

There is powerful evidence that certain types of technological shocks raise the returns to schooling[37]. There is also pretty strong evidence that parents/students respond to perceived returns to schooling. In an important paper Foster and Rosenzweig (1996) show three important things. First, the technological shock of the introduction of High Yielding Varieties caused returns to education to be higher in those regions conducive to these changes than in other regions—so the return to primary schooling in regions with a technological shock to farm profits a standard deviation above average was estimated at 7.8 percent. Second, the returns in regions that did not have a positive technological shock were very, very, low—only around 2 percent. Third, in regions in which returns were higher the growth in schooling was much higher.

[37] There is of course the enormous literature on the rise in the wage premia to education in the USA (and to a lesser extent, Europe) in spite of rising supplies of educated workers that attempts to determine the sources of the rise.

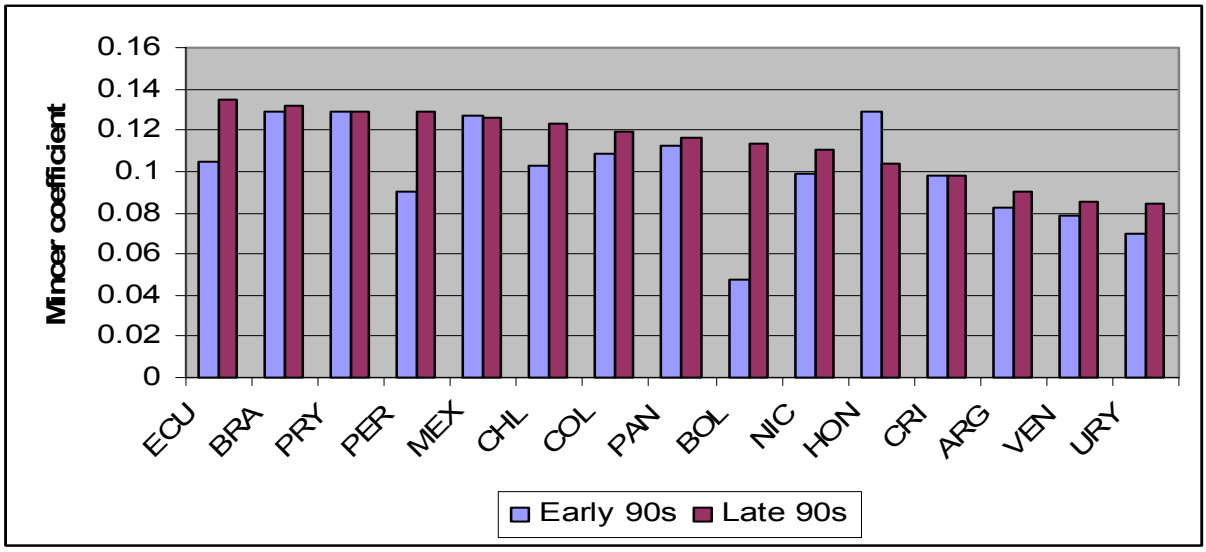
Figure 17: The returns to primary schooling (from increased farm profits) were much higher in regions conducive to the adoption of Green Revolution technologies in India in the 1970s.



In Latin America there has been an enormous expansion in the returns to schooling. Behrman, Birdsall and Szekely (2001) show policy changes affect technological progress which then creates higher returns to education in the "adaptation to disequilibria." This is the macro-economic counter-part of the effects seen in the Green Revolution. In contrast, in economies that are stagnant the demand for educated labor will be stagnant and the dramatic expansions in the supply of educated workers will drive down the wage gains from education.

Figure 18: Returns to schooling typically increased during the 1990s in Latin America

(Coefficient on schooling in Mincer regression, with data sets from the early versus the later 1990s)



It is not impossible to imagine that one reason that the expansion of schooling has stagnated in Sub-Saharan Africa is that there have been neither large positive shocks to agricultural technologies (*a la* some regions of India) nor has there been a large expansion in employment in industry or general economy wide increases in productivity. Almost certainly the returns to education in most of Africa today are much lower than they could be.

Conclusion on raising incomes and benefits. Even if one accepts that reforms that raise the general level of income raise demand for education and that raising economy wide returns to skills acquired through schooling is key to expanding educational attainments, the question remains: "what is the 'opportunity'?" It is "policy reform"? And if so, what are the costs and benefits? Is it investing in technological innovations that raise returns to schooling? If so, what are the costs and benefits of investing in innovations?[38] Generally, reforms that raise output are desirable on their own terms (without taking into account the feedback to improved education) and so 'should' be adopted in any case.

The main point is that there is a tendency by education analysts to overstate the importance of sector specific measures (a natural tendency since this is what they might control or influence) and understate the importance of economy wide factors. Generalized increases in per capita income are an enormously important route for improvements in "human development" indicators of all types and should not be ignored.

Opportunity number 4: Demand side: Direct (targeted?) support to households that lower the cost of schooling.

The second demand side opportunity to be considered are mechanisms to increase enrollment, attainment, or learning by lowering the (relative) costs of schooling. The reduction in costs can either be untargeted (as in fee waivers) or targeted by some criteria (gender, location, caste) and can either be reducing costs or a positive reward, such as a scholarship or conditional cash transfer.

Table 9: The variety of demand side transfers

			Cash (or equivalent)	Non-Cash
Untargeted			Fee waivers (Malawi, Uganda, Kenya)	
Targeted	Pure education	"Need"	Scholarships/vouchers (Colombia)	School feeding (e.g. Tamil Nadu)
		"Ability"		
	Social transfers conditioned on education		PROGRESA (Mexico)	FFE in Bangladesh) [Vermeersh 2002]

[38] This is particularly striking as many of the "opportunities" that would raise the returns to schooling would also be beneficial in their own right and, in fact, the impact on raising schooling would be only a tiny part of the overall benefits. Which implies that if these could be justified on schooling grounds alone the total benefit/cost ratio would be spectacularly large, which raises the question of why they have not already been done—that is, what is the "market" for income/human capital return raising economic reforms that it is so inefficient.

Costs of raising enrollments by lowering relative prices. To an economist there is no question that lowering relative prices will raise the quantity of schooling demanded—the only question is whether this is "cost effective" as a means of addressing the lack of education. A helpful way of thinking about that question is table 10 which distinguishes the *incremental* enrollment from the transfer and the total cost of the transfer plus the additional cost of schooling.

Table 10: Illustration of the implications for cost effectiveness of targeting transfers

	Do receive the transfer	Do not receive the transfer
Enroll with or without the transfer of magnitude T	**A** **('type II' error-transfer where it has no impact)**	**B**
Would enroll with the transfer T but not without transfer	**C**	**D** **('Type I error'—transfer where it would have an impact)**
Would not enroll even with the transfer	***E*** **('type II' error-transfer where it has no impact)**	**F**

Examples of alternative policies

	Incremental cost	**Incremental enrollment**	**Cost per incremental enrollment**
Reducing fees from *f* to zero with cost of schooling *c*	**A*f+C*c** **(A=all previously enrolled)**	**C**	**(A/C)*f+c**
Targeted transfer of *t* to 'poor' households with cost of school of *c*.	**(A+C)*t+C*c** **(A=just 'poor' who would have enrolled)**	**C**	**(A/C)*t+t+c**

Obviously the two key aspects of the cost-effectiveness is the ration A/C, which depends on two elements: the price elasticity of demand for schooling and the ability to target so that A/C is low.

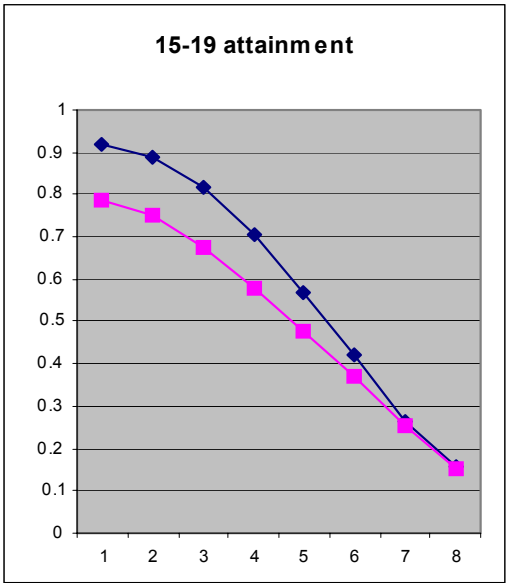
Price elasticity of the demand for schooling and universal reductions in cost. If enrollments are to be increased by lowering the price of schooling universally—say, reducing or eliminating school fees across the board—then the cost per additional student enrolled is a function of the elasticity of the aggregate demand for schooling. If enrollments are already high and the elasticity is low then the costs per additional student enrolled are astronomical.

Most of the available evidence suggests the demand for schooling is quite price inelastic. Schultz (1991) found that the school subsidy offered by the Progresa program – which provides education grants for poor families in rural Mexico – reduced the private costs of school by more than half, but increased the educational attainment by only ten percent hence the implied demand elasticity was around -0.2. The evidence from the distance regressions reported (plus many assumptions) allows a calculation of the price elasticity implied by the enrollment response to reduced travel time. These calculations suggest an elasticity of around -.3 and that above -.5 is substantially too high.

In contrast, the experience with the elimination of fees followed by huge increases in enrollments in Africa suggests that aggregate demand for *enrollment* is quite elastic (e.g. Deininger 2001). There are however, three aspects to the experience that makes one slightly cautious about a policy of universal fee waivers. First, by making the fee waiver universal the revenue foregone per additional student enrolled is quite high.

Second, the evidence is mixed of the impact on *attainment* as, at least in Malawi, there is some evidence that the reductions have increased ever enrollment but that persistence is lower (perhaps because quality is lower) and attainment has not been influenced nearly as much. The attainment profile of those aged 15 in 1992 (unaffected by the elimination of fees in 1994) and in 2000 suggests that while completion of the early grades has increased the fraction of 15 years olds completing 6th, 7th or 8th grade is nearly unchanged.

Figure 19: Attainment profiles after the elimination of fees in Malawi show unchanged grade attainment.



Source: Martyris (2004)

Third, unless there is the political willingness to replace the revenue foregone in a sustained way, reductions in budgets may lead to reductions in learning achievement for all students so that the net impact on learning achievement could actually be negative. One hint of these problems is that repetition rates in first grade have increased from 20 percent in 1992 to 45 percent in 2000.

Conditional cash transfers. Morely and Coady (2003) review the experience with "cash for education" or conditional cash transfer schemes like Mexico's PROGRESA, Food For Education (FFE) in Bangladesh, Bolsa Escola in Brazil (taken nation wide in 2001) and others. It is clear that these programs have substantial impact on enrollments. The PROGRESA scheme had a built in randomized evaluation and hence has about as reliable estimates as one could hope—which suggest that grade 8 completion was roughly 12 percentage points higher for PROGRESA recipients than non-recipients.

Whether CFE schemes are "cost-effective" all depends on whether one views the program as a social transfer program with the additional benefit of increasing enrollment—that is, the

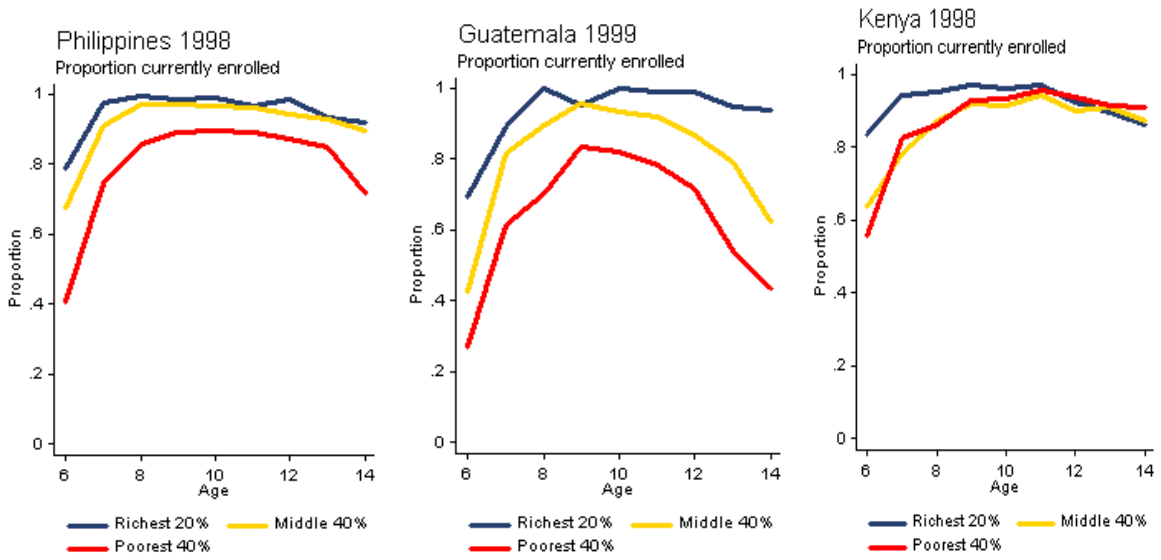
transfer would have been made anyway and the education benefits are attributed only to the incremental costs—or whether all of the transfer is attributed to the education objective. The basic point is illustrated with the cost-effectiveness calculations in Morley and Coady (2003). The calculate that the transfers produced an additional 393 years of schooling for a cohort of a 1000 with a total transfer of 3.43 million pesos in grants for that same cohort for an incremental cost per extra year of education generated of 8,727 pesos—this is 15 percent of GDP per capita per year of additional schooling. A rough estimate of the cost of per year of schooling in Mexico in 2002 is about 8,800 pesos. This means that the total public sector cost of an additional year of schooling via PROGRESA was about double the current average cost per student. It is hard to believe that is the most cost-effective way to increase schooling[39].

Targeted transfers. The key questions for these types of interventions as a means of expanding education are design questions.At what age/stage of education? There is evidence from the USA that "the earlier the better", although many scholarship programs focus only on retention or progression. Is the issue of late enrollment underemphasized? The key issue is the trade-off between cost effectiveness of targeting. This is function of the cost and feasibility of distinguishing the enrollment of various groups (which the attainment profiles in part I give some indication of how much the enrollment differences vary across easily identifiable characteristics in a given country). On the down side sharp targeting to maximize cost-effectiveness may miss many who are not enrolled. So if the groups with low achievement are small (e.g. poor girls in remote villages) then cost-effective targeting will still leave many out of school.

One key issue that has not been sufficiently addressed is targeting to late enrollment. There is increasing evidence that late enrollment also an important determinant of total attainment. This would suggest that targeting to the age group where the enrollment rates are the most different would have the least "leakage" to infra-marginal (these would attend in any case)—but targeting can have high administrative or political costs.

[39] Morley and Coady (2003) compare that to the cost-effectiveness of raising enrollments via school construction and, needless to say given the weak connection between school construction and enrollments noted above find the cost 11 times higher via construction.

Figure 20: Late enrollment implies much larger differences in enrollment rates by household wealth at early ages than later…



Source: Filmer and Pritchett and website.

Finally, a large issue is targeting ability versus some combination of ability and need. This is typically not even addressed.

Conclusions on demand side transfers. This brief review suggests the following conclusions.

- *Conditional cash transfers:* If the government is going to make cash transfers to individuals/households are part of a *poverty* program, there is little reason to not make these transfers conditional on the children of recipients enrolling in school. In this sense, conditional cash transfers are "cost effective" in raising enrollment because most of the cost of the program—the financial transfer—is a cost that would be incurred in any case and hence the education gains need to be judged only against the incremental costs of making the transfer conditional.

  By this same logic it is almost inconceivable that most conditional cash transfer programs could be justified exclusively (or even primarily) by their education benefits as, in nearly any setting capable of implementing such a program, nearly all of the transfer will be infra-marginal (that is, go to families whose children would have enrolled anyway. There is some hope for transfers sharply targeted to stages of the cycle—e.g. to reduce late enrollment, repletion in early grades, and transitions across levels of schooling

- *Untargeted reductions in fees.* While most are favorable about the experience with fee waivers in Africa I would be much more cautious.

- *Targeted programs.* There is no question that targeted programs such as in-kind transfers like meals in school can have substantial enrollment impacts. The cost-effectiveness question hinges on the ability to target to make the enrollment increase per dollar spent substantial.

- *Why not portable demand side transfers?* One large question is whether "demand side" transfers should be limited exclusively to public sector producers. If there is the capability of providing a cash transfer, an in-kind benefit to targeted groups, or a scholarship the logic of limiting the use of that to only public producers escapes me.

Part IV: The fifth element: Systemic reform to create performance management and producer accountability

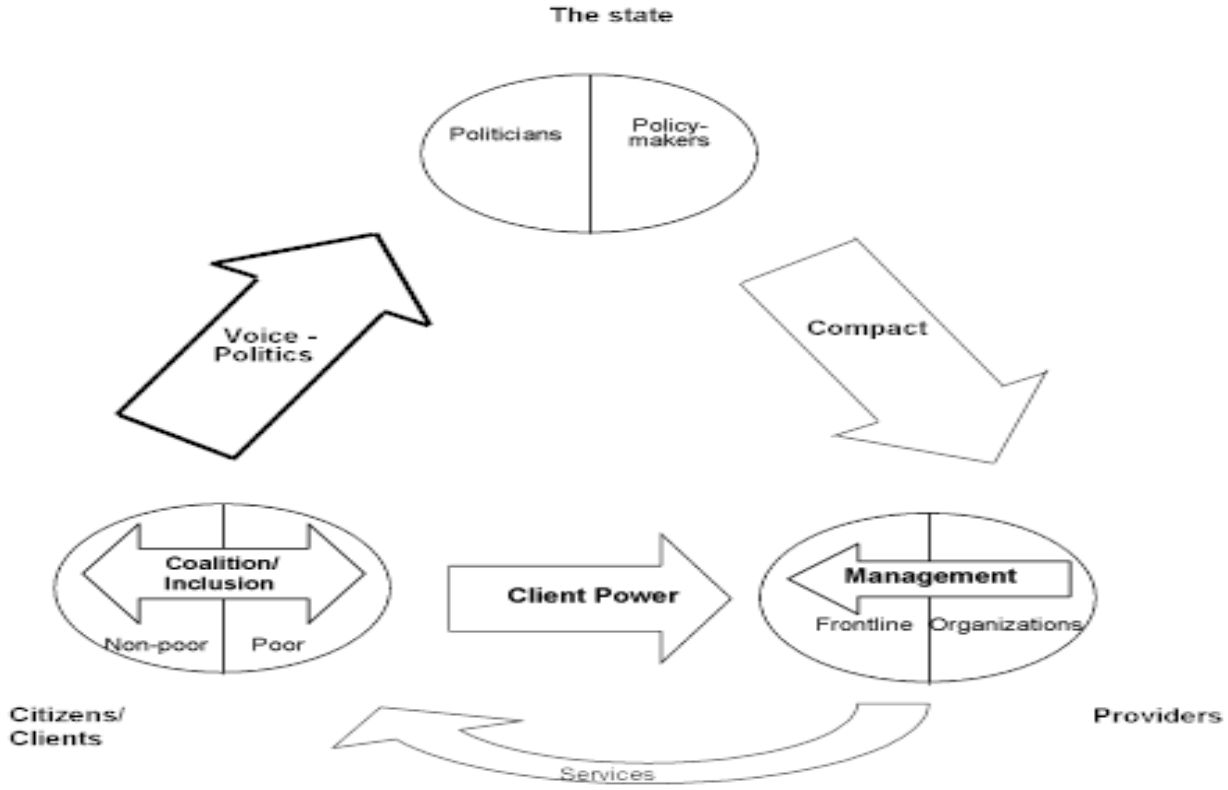
The World Bank's *World Development Report 2004: Making Services Work for the Poor* documents the problems with public production of schooling—low levels of budget commitment, high levels of teacher absenteeism, teachers hired for political patronage, low technical quality, misallocation of resources by level of education and across inputs, ineffectiveness of teacher training, stagnant productivity[40]. More importantly, it argues that these failures are *endogenous* and *systemic.* It proposes what is essentially an extended principal-agent analysis in which the relationship between the principal and agent is called a relationship of accountability. The principal thesis is that to a large extent the currently observed poor educational outcomes are the result of failures of the relationships of accountability.

This implies that the key problem is not applying the discrete policy actions (opportunities 1 to 4) that attempt to directly affect the "proximate determinants" of learning (illustrate in figure 5)—even when it is clear there are "off the shelf" innovations that would substantially improve the quality of schooling. The key problem is that within the current institutional conditions for government production of schooling, there are too few incentives to create *performance oriented* management—in either the government or market sectors. That is problem is that the choices are the outcome of the existing incentives (or lack thereof) for producers to raise the quality of learning. Where the key constraint is that the system does not produce incentives for producers to be effective, then neither budget increases nor new knowledge are likely to be effective[41].

[40] I was a member of the WDR 2004 teach and had responsibility for chapter 7 on education and hence, naturally, believe the analysis presented there is largely correct.

[41] One would think that analyzing education policy in terms of the incentives it created would be second nature to economists, but not. One conjecture of mine is that the "externality" version of the "normative as positive" model had such a powerful attraction as a formal apparatus that economists could easily assume that if welfare maximizing governments had a rationale for intervention in the market for education and governments did in fact produce schooling that the *reason* governments produced schooling was to maximize welfare. If that is the case then all that needs to be worried about are issues like the allocation of budget as, after all, if governments are producing schooling to maximize welfare then they are doing so efficiently. But, since not all (any?) governments produce schooling to maximize citizen welfare then it also goes without saying there is not particular reason to assume productive or allocative efficiency.

Figure 21: An illustration of the four (potential) relationships of accountability involved in the provision of services via the public sector: voice-politics (between citizens and the state), compact (between the state and producers), management (between organizations and frontline producers), and client power (between citizens/consumers and producers).



Source: WDR 2004

In order for (near exclusively) publicly produced education to be effective *either* an authoritarian nation-state has to create powerful incentives for performance without either the 'voice' of citizens or direct 'client power' on providers (which the high levels of learning achievement in Cuba, for instance, prove is possible) *or* the "long-route" accountability has operate. The 'long-route' is that citizens have to make demands felt via the political system, those demands have to translate into clear signals to public sector producers (via 'compact' and 'management') with or without a great deal of direct 'client power.' Of course, a country can also provide education without direct production by transferring resources directly to citizens and allowing direct client power to operate via their choice of schooling (with of course oversight of content and quality) to discipline producers. But systems in which the only form of government support for basic education is the direct production of schooling can become dysfunctional if any of these key links break down[42].

[42] A key limitation of many international efforts to improve education is precisely that they are inter*national*—that is, the key agents are representatives of nation-states (or agglomerations of nation-states, such as the UN or World Bank) and hence tend to assume the perspective of the nation-state, which may have very different interests in schooling than the typical citizen.

This is *not* an ideological case for 'vouchers' or the wholesale 'privatization'--some of the world's best schooling systems rely on direct public production and many of the world schools in the world are 'private." Nor does "performance management" mean slavish reliance on standardized tests or simple minded proposals for teacher pay. But much of the world's population is not well-served by existing systems.

A well ordered system for schooling has to solve several problems simultaneously:

- *Create clear objectives for the schooling system,*
- *Provide sustained, adequate, financing to line producers to achieve those objectives,*
- *Give "line producers"—principals and teachers--sufficient autonomy to manage for results,*
- *Maintain producer accountability for results--which implies transparent, consistent, measures of progress towards the objectives.*

There is no unique, universally applicable, way to accomplish these desiderata. The industrial economies have all found different ways to create systems that are reasonably effective (if not efficient or dynamic)—the Dutch have almost complete choice of publicly subsidized, privately produced schooling, the USA has historically relied on extremely decentralized control of school districts[43], France has maintained quite tight centralization of schools at every level. By the same token, the basic structures of the educational system in Singapore and Nigeria are similar--as descendents of the British system—and yet have almost polar opposite outcomes.

At the same time, while discussions, especially in international forums, tend to focus almost exclusively on government production of schooling the provision of basic schooling outside the traditional public sector model has been demonstrated to be feasible at scale in a wide variety of formats and in a wide variety of settings. A variety of alternatives to traditional public production has emerged in a wide range of settings in response to conflict (community controlled schools in El Salvador), lack of government alternatives (BRAC non-formal education in Bangladesh), changing demands in the labor market (English as medium of instruction schools in Mumbai), parental demands for religious education (*Fe y Alegria* in Venezuela), or simply a collapse in government schools (Cameroon). This is of course in addition to the usual alternative of 'vouchers' that allow students to attend private schools that has been demonstrated as workable system for the provision of education in Chile and the Czech Republic.

[43] But with a high degree of centralization within school districts so that principals are perhaps even more constrained in this "decentralized" system than in a "centralized" system like the UK.

Table 11: Alternatives to the traditional organization of government production of schooling—examples of different types of arrangements—at scale	
	Description of organization, performance
NGO	The Non Formal Primary Education (NFPE) provides basic education in small one-room mud structures, located within the village.for 8 to 10 year-olds in rural Bangladesh with an emphasis on girl's schooling--of the 1.1 million students currently enrolled, 70% are girls. **Teachers** are mostly local married women, part time workers. They must have at least 9 years of schooling, and undergo intensive training. **Decentralized management, Active parent and community involvement** including in determining the location of the school, assisting in construction, setting schedule, hiring and supervision of teachers. Parent-teacher meetings are held regularly. **Performance.** BRAC students have equal or better performance on national exams to test mastery of basic education. Over 50% of BRAC graduates passed, compared with 20% of Government students.[1]
Community control (EDUCO)	EDUCO schools in El Salvador are community managed schools which receive a per student subsidy from the government. Community committees have complete control over finances, including the hiring and firing of ***teachers*** who are on one year, renewable, contracts. Evaluation of the **performance** finds that controlling for student background EDUCO schools do no worse (and arguably better) than government organized schools.
Religious	Schools run by the Catholic church aimed at disadvantaged students in Venezuela (among other countries). Evaluations suggest that performance is higher, costs are lower, and in spite of much lower salaries, teachers are more satisfied.
Private for profit—without support	As a response to changing economic conditions *English as medium of instruction* schools have emerged spontaneously in Mumbai, India. These schools now account for a large percentage of all enrollment.

Centralized, monoposonistic production of basic schooling has a tendency to blur everything. Since everything is within "the government" the roles of government as financer of schooling, as regulator, as monitor, and as producer are all conflated. This can lead to a lack of clear objectives. But if there are no clear objectives (which of course includes the case of having many objectives which cannot be measured) then, as the literature on principal agents and compensation (e.g. Holmstrom and Milgrom 1997, Milgrom and Roberts 1997) shows there is not clear basis to manage or to create incentives within organizations. Nearly inexorably leads to "input" based management and "rules based" behavior.

That said, the tasks are difficult and the challenges of making any system of education work formidable.

Clear objectives. Schools are not machines for teaching facts—schools are the mechanism whereby societies replicate themselves. This means that while learning achievement is obviously a key element of schooling the objectives for schools are much broader. Since 'what gets measured gets done' one cannot produce a set of educational goals that include lists of desired outcomes but then set about only measuring enrollments or physical targets that measure "quality" like pupil teacher ratios. By the same token one cannot set education goals for creativity and measure only standardized tests.

Sustained adequate financing. The financing of schools is an issue that is, in many ways, independent of the structure of production—as one can always finance schools from general

revenue irrespective of who produces. But any system has to face a series of challenges—particularly the issue of financing teachers.

Autonomy to manage for results. One there is a clear statement of the objectives of using tax revenues for basic schooling and a source of financing then one can create a mechanism that allows producers to make decisions as how those goals can be most effectively met. Accountability for performance requires *autonomy* to act. In general this means pushing responsibility and power for decision making to the lowest feasible level, which depending on the function, could be the region, the school system (public or private), the school headmaster or the class-room teacher. This doesn't mean each school invents a curriculum. Woessmann (2002) addresses nicely the issue of which functions should be allocated to which level.

Accountability. Perhaps the key element is a *centralized* system of creating and *disseminating* information to parents/students about *school* performance. This function of the school system needs to be centralized because the schooling system creates a common system of core objectives against which all schools (which receive public funds) are to be gauged and against which progress over time can be measured. And because there are substantial economies of scale in developing a proper way to do this.

Given the *sufficient relevant information* (e.g. about comparative features, costs and expenditures, and performance (adjusted)) there are a variety of *mechanisms* to pressure schools for better performance. Public organizations do that in one way, private non-profit organizations do it another, choice based systems do it another, community management another. The key is that accountability should be based on adequate information. "Choice" without reliable information is not a happy recipe.

Part V: Bringing all opportunities together in an international program: Education for All/Fast Track Initiative (EFA/FTI)

The most prominent intervention on the international (particularly donor) agenda is the Education for All *Fast Track Initiative.* This set of proposals follows up on the Dhakar 'commitments' to Education for All (EFA) and the Millennium Development Goals (MDGs). The basic idea is that donors will provide incremental budget support to countries that provide "credible" plans for meeting the EFA targets (or at least improving performance on the targets). The people behind the EFA/FTI understand the depth of the problem and that meeting the targets is about much more than 'building schools' and that improving quality is about much more than 'increasing spending.' The EFA/FTI is the "cutting edge" approach to the problem of the lack of education. In the descriptions of what the country plans should include it has has attempted to grapple with how to integrate all five 'opportunities' into a coherent package (and particularly a coherent package that could be supported with international financial assistance).

One of the striking and innovative features about the EFA/FTI for instance is that financing gaps are based on a move towards "efficient" expenditures. Based on an empirical analysis of the expenditure allocations of countries which have been successful in reaching the goals of primary completion they establish indicative ratios for teacher wages/GDP and class size. If a country has high unit costs of primary schooling this does not imply a higher financing gap to meet the MDG that donors to meet. Rather, it implies that unit costs should be reduced (or the country should finance the 'excess' expenditure themselves).

As a conclusion, I assess the likely returns in improving basic education from the opportunities considered. Obviously in assessing a credible plan for progress in any given country requires consideration of the existing conditions and the likely outcomes from policy action and the 'fit' of various systemic reform alternatives with existing conditions (including political situations).

Table 12: Conclusion		
	The returns to devoting additional public resources to this are high if:	Frequency of the *possibility* of high payoffs
1--System expansion	Demand under existing conditions exceeds supply as evidenced by: high class sizes (especially in upper grades), students willing to long travel distances. Schools of adequate quality can be provided as evidenced by high retention in existing schools.	Rare
2a) Radial expansion of budgets	Producers are (reasonably) efficient as evidenced by: cost ratios near norms, active management for effectiveness, ability to measure resource use, measurement of quality.	Rare
2b) Selective expansions of inputs	Educational system capable of managing and implementing programs for expansion of high impact of inputs as evidenced by: measurement of quality (to monitor quality improvement), capability of active management, teaching force capable of implementing new techniques/utilizing new inputs.	Rare to uncommon in poorest settings, common in middle income countries
2c) Rigorous evaluation of impact	Knowledge about effectiveness is scarce Some possibility of impact on decisions	Common (since this capability can be 'cocooned')
3) Economic reforms to stimulate demand	Existing returns to education are low	Common in SSA, uncommon elsewhere
	Reforms are possible that will raise economic growth and returns to schooling	Common—feasibility case by case
4) Demand side transfers/ cost reduction	Existing school is of adequate quality (as evidenced by low repetition, high persistence) and preferably by measurement Administrative capability of targeting at reasonable cost Alternative sources to generate revenue	Rare to uncommon in poorest, more common in middle income.
5) System reform	Improve the definition of learning objectives	Everywhere
	Financing	Rare/uncommon
	Expand producer autonomy	Common (at the margins),
	Expansion of availability of information about performance	Common
	Expand accountability in some mode	Common

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer.Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. American Economic Review, 92(5), 1535-1558.

Alderman, H., Orazem P., and Paterno E. 1996. School quality, school cost, and the public/private school choices of low income households in Pakistan. Impact Evaluation of Education Reform Working Paper Series No. 2 PRDPH. Washington DC, The World Bank.

Angrist, Joshua D., and Lavy, Victor. 1999. Using Maimonides' Rule of Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 144, 533-575.

Association for the Evaluation of Educational Achievmenet. 2003. PIRLS 2001 International Report. IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries.

Banerjee, A., S. Cole, E. Duflo and L. Linden. 2003. Remedying Education: Evidence from Two Randomized Experiments in India. [Unpublished Manuscript]

Barro, Robert J. and Jong-wha Lee. 1993. International Comparisons of Educational Attainment. *Journal of Monetary Economics*, 32(3), 363-94.

Behrman JR, Khan S, Ross D, Sabot R . 1997. School quality and cognitive achievement production: A case study for rural Pakistan. Economics of Education Review, 16 (2). 127-142.

Birdsall, Nancy and Francois Orivel. 1996. Demand for Primary Schooling in Rural Mali: Should User Fees be Increased? "*Education Economics*, 4(3).

Bredie, Joseph W. B and Girindre K Beeharry. 1998. School enrollment decline in Sub-Saharan Africa: Beyond the Supply Constraint. Discussion Paper 395, World Bank.

Bruns, Barbara, Alain Mingat, and Ramahatra Rakotomalala. 2003. A Chance for Every Child: Achieving Universal Primary Education by 2015. Washington DC: World Bank.

Burney, Nadeem A. and Mohammad Irfan. 1995. Determinants of child school enrolment: evidence from LDCs using choice-theoretic approach. *International Journal of Social Economics*, 22(1), 24-40.

Canagarajah, Sudharashan and Harold Coulombe. 1997. Child labor and schooling in Ghana. Policy Research Working Paper 1844. World Bank.

Carneiro, P., Heckman J., 2003. Human Capital Policy. National Bureau of Economic Research, Working Paper 9495.

Case, Anne and Angus Deaton. 1999. School Inputs and Educational Outcomes in South Africa. *Quarterly Journal of Economics*, 114(3), 1047-1084.

Castro-Leal, Florencia. 1996. Who benefits from public education spending in Malawi? Results from the recent education reform. Discussion Paper 350, World Bank.

Clemens, Michael A. 2003. The Long Walk to School: Development goals in historical perspectives. Center for Global Development. [Unpublished Manuscript]

Colclough, Christopher and Samer Al-Samarrai. 2000. Achieving schooling for all: Budgetary expenditures on education in Sub-Saharan African and South Asia. *World Development*, 28 (11), 1927-1944.

Comber, L.C., and John P. Keeves. 1973. Science Educational in Nineteen Countries. New York: John Wiley & Sons, Inc.

Crouch, Luis and F. Henry Healy. 1997. Education Reform Support. Volume I: Overview and Bibliography. Office of Sustainable Development Bureau of Africa. U.S. Agency for International Development.

Deininger, Klaus. 2003. Does cost of schooling affect enrollment by the poor? Universal primary education in Uganda. *Economics of Education Review*, 22, 291-305.

Delamonica, Enrique, Santosh Mehrotra, and Jan Vandermoortele. 2001. Is the EFA affordable? Estimating the global minimum cost of 'Education for All.' Innocenti Working Paper 87, Florence: UNICEF Innocenti Research Center.

Devarajan, Shantayanan, Margaret Miller, and Eric Swanson. 2002. Goals for Development: History, Prospects and Costs. Policy Research Working Paper 2819, April, World Bank.

Duflo, Esther. 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795-813.

Elley, Warwick B. 1992. How in the World Do Students Read? Hamburg: International Educational Association.

Feynman, Richard. 1989. *The Feynman Lectures on Physics*. Addison Wesley.

Filmer, Deon and Lant Pritchett. 1999. What Education Production Functions Really Show: A Positive Theory of Education Expenditure. *Economics of Education Review*, 18(2), 223-239.

Filmer, Deon, and Lant Pritchett. 1999. The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries. *Population and Development Review*, 25(1).

Filmer, Deon. 2004. School availability and school participation in 21 developing countries. The World Bank.

Foster, Andrew D, and Rosenzweig, Mark R. 1996. Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution. *The American Economic Review*, 86(4), 931-953.

Fredriksen, Birger. 2002. Education for All African Children by 2015: What will take to keep the promise? Paper prepared for the Annual Bank Conference on Development Economics – Europe, Oslo, June 24-26.

Fuller, B., and Clarke, P. 1994. Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research*, 64(1), 119–157.

Glewwe, Paul and Michael Kremer. 2003. Schools, Teachers, and Education Outcomes in Developing Countries. In *Handbook on the Economics of Education.* Forthcoming.

Glewwe, Paul, and Hanan Jacoby. 1994. Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana. *The Journal of Human Resources*, 29(3), 843-864.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 1997. Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya [Unpublished Manuscript]

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz. 2000. Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya. Working Paper 8018. National Bureau of Economic Research. Forthcoming in Journal of Development Economics.

Glewwe, Paul, Nauman Ilias and Michael Kremer. 2003. Teacher Incentives. Working Paper 9671. National Bureau of Economic Research.

Glewwe, Paul, Nauman Ilias, Michael Kremer. 2003. Teacher Incentives. Working Paper 9671, National Bureau of Economic Research.

Glewwe, Paul. 2002. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. Journal of Economic Literature XL, June. 436-482.

Gradstein M. and M. Justman. 2002. Education, Social Cohesion and Growth. *American Economic Review*, 92 (September), 1192-1204.

Greany, V., Khandker, S., and Alam, M., 1998. Bangladesh: Assessing Basic Learning Skills.

Gundlach, Erich and Ludger Woessmann. 1999. The Fading Productivity of Schooling in East Asia. Working Paper 945, Kiel Institute of World Economics.

Gundlach, Erich, Ludger Woessmann and Jens Gmelin. 1999. The Decline of Schooling Productivity in OECD Countries. Working Paper 926, Kiel Institute of World Economics.

Handa, Sudhanshu. 2002. Raising primary school enrolment in developing countries: The relative importance of supply and demand. *Journal of Development Economics*, 69, 103-128.

Hanushek Eric A. 1999. The Evidence on Class Size. In Susan E. Mayer and Paul Peterson (ed.), Earning and Learning: How Schools Matter. Washington, D.C.: Brookings Institution, 131-168.

Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from other Investigations of Class Size Effects. Educational Evaluation and Policy Analysis 21.143-163.

Hanushek, Eric A. 1986. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24, 1141-1177.

Hanushek, Eric A. 1995. Interpreting Recent Research on Schooling in Developing Countries. World Bank Research Observer 10(August). 227-46.

Hanushek, Eric A. 1996. School Resources and Student Performance. In Gary Burless (ed.) Does Money Matter? The Effects of School Resources on Student Achievement and Adult Success. Washington DC: The Brookings Institute.

Hanushek, Eric and Kim, Dongwook. 1995. Schooling, Labor Force Quality, and Economic Growth. Working Paper 5399, National Bureau of Economic Research.

Hanushek, Eric. 1995. Outcomes, Costs, and Incentives in Schools. In Board on Science, Technology, and Economic Policy, National Research Council, Improving the Performance of America's Schools. Washington, DC: National Academy Press, 28-51. [reprinted in Eric A. Hanushek and Dale W. Jorgenson(ed.).1996. Improving America's Schools: The Role of Incentives, Washington, DC: National Academy Press.]

Hanushek, Eric A, and Dennis D. Kimko. 2000. Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review* 90 (5), 1184-1208.

Harbison, R., & Hanushek, E. A. 1992. Educational performance of the poor: Lessons from northeast Brazil. Oxford, for the World Bank, Washington DC.

Harmon, Colm, Hessel Oosterbeek, and Ian Walker. 2003. The returns to education: Microeconomics. *Journal of Economic Surveys*, 17(2), 115-155.

Hedges, L.V., Laine R.D., Greenwald R. 1994. Does Money Matter? A Meta Analysis of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher* 23(3), 5-14.

Holmes, Jessica. 1999. Measuring the determinants of school completion in Pakistan: analysis of censoring and selection bias. *Economics of Education Review*, 22, 249-264.

Hoxby, Caroline M. 2000. The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics*, 115 (November), 1239-1285.

Jacoby, Hanan. 1994. Borrowing Constraints and Progress Through School: Evidence from Peru. *The Review of Economics and Statistics*, 76(1), 151-160.

Jalan, Jyotsna and Elena Glinskaya. 2003. Improving primary school education in India: An impact assessment of DPEP- Phase I. [Unpublished Manuscript]
James, E., King, E., & Suryadi, A. 1996. Finance, management, and costs of public and private schools in Indonesia. *Economics of Education Review*, 15(4), 387–398.

Jamison, Dean, Barbara Searle, Klaus Galda, and Stephen P. Heyneman. 1981. Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement. *Journal of Educational Psychology*, 73(4), 556-67.

Jimenez, E., & Lockheed M. E. 1995. Public and private secondary education in developing countries. Discussion Paper No. 309. Washington, DC: The World Bank.

King, Elizabeth and Berk Ozler . 1998. What's Decentralization Got To Do With Learning? The Case of Nicaragua's Autonomy Reform. Working Paper Series on Impact Evaluation of Education Reforms, Paper No. 9, April 1998, Development Economics Research Group, The World Bank.

Kremer, Michael. Randomized Evaluations of Education Program in Developing Countries: Some Lessons. Forthcoming in American Economic Review Papers and Proceedings.

Kremer, Michael and Andrei Sarychev, 1998, "Why Do Governments Operate Schools?" mimeo.

Krueger, Alan B. 2000. Economic considerations and class size. Working Paper 447, Princeton University, Industrial Relations Section.

Krueger, Alan B. and Diane M. Whitmore. 2001. The effects of attending a small class in the early graduation college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, 111, 1-28.

Lavy, Victor. 1996. School supply constraints and children's educational outcomes in rural Ghana. *Journal of Development Economics*, 51, 291-314.

Lazear, Edward. 2001. Educational Production. *Quarterly Journal of Economics*, 116(3), 777-803.

Lee, Jong-Wha and Robert J. Barro. 2001. School quality in a cross section of countries. *Economica*, 68, 465-488.

Lockheed, Marliane E., Adriaan M. Verspoor and Associates. 1991. Improving Primary Education in Developing Countries. Oxford University Press for the World Bank.

Medrich, Elliot, and Griffith Jeanne. 1992. International Mathematics and Science Assessment: What Have We Learned? Research and Development Report, January. National Center for Education Statistics.

Miguel, Edward and Michael Kremer. 2001. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. Working Paper 8481. National Bureau of Economic Research (Forthcoming in *Econometrica*).

Morley, Samuel and David Coady. 2003. From Social Assistance to Social Development: Targeted Education Subsidies in Developing Countries.

Munshi, Kaivan and Mark Rosenzweig. 2003. Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. M.I.T. Department of Economics Working Paper No. 03-23.

National Center for Education Statistics. 2000. Pursuing Excellence: Comparison of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999.

Newman, John, Laura Rawlings and Paul Gertler. 1994. Using randomized control design in evaluating social sector programs in developing countries. *The World Bank Research Observer*, 9(2), 181-202.

Nwaerondu, Ndubuisi Goodluck, and Gordon Thompson. 1987. The Use of Educational Radio in Developing Countries: Lessons from the Past. *Journal of Distance Education*, 2(2), 43-54.

OECD. 2001. Reading for Change: Performance and Engagement Across Countries. Results of from PISA 2000.

Pitt, M., M. Rosenzweig, and D. Gibbons. 1993. The determinants and consequences of the placement of Government Programs in Indonesia. *World Bank Economic Review*, 7(3), 319-348.

Pritchett, Lant. 1996. Where has all the education gone? *World Bank Economic Review*, 15(3).

Pritchett, Lant. 2002. Ought Ain't Is: Midnight Thoughts on Education.

Pritchett, Lant. 2002. When will they ever learn": Why all Governments Produce Schooling.

Pritchett, Lant and Michael Woolcock, 2004, "Arraying the Disarray in Development" forthcoming in *World Development*.

Ravallion, Martin and Quentin Wodon. 1999. Does Child Labor Displace Schooling? Evidence on Behavioral Responses to an Enrollment Subsidy. Working Paper 2116, The World Bank.

Robitaille, David and Robert Garden (Eds.). 1989. The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics, Vol. II. Oxford: Pergamon Press.

Roberts, John, 2003, "Poverty Reduction Outcomes in Education and Health: Public Expenditure and Aid," Working Paper #210, ODI: UK.

Scott, James, 1999, *Seeing Like a State.*

Schultz, Paul T. 2001. School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. Center Discussion Paper No. 843. Economic Growth Center, Yale University.

UNESCO. 2002. Education for All: Is the World on Track? EFA Global Monitoring Report.

Urquiola, Miguel. 2001. Identifying class size effects in developing countries: Evidence from rural schools in Bolivia. Policy Research Working Paper 2711. The World Bank.

US Department of Education. 2000. Mathematics and Science in the Eighth Grade: Findings from the Third International Mathematics and Science Study.

Vermeersch, Christel. 2002. School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. [Unpublished Manuscript]

World Bank. 2004. World Development Report 2004: Making Services Work For Poor People. Washington DC: The World Bank. Forthcoming.

Table A.1: The estimated impact of reducing distance to schools in rural areas in 21 countries is typically statistically significant but quantitatively small…

Country	Year	Avg. Distance (in kms)	Average enrollment ages 6-14	Linear Prob. model Coeff. on cluster distance to primary school[a]	Sig.	Scenarios for enrollment based on estimated coefficients: If distance to primary school=0 for all clusters	If distance to primary *and* secondary school is zero for all clusters	If a primary school in each cluster
Chad	98	7.2	27.0%	-0.8%	**	32.8%	40.4%	36.5%
Mali	95/96	6.6	16.1%	-0.8%	**	21.4%	21.4%	28.4%
Senegal	92/93	5.0	18.2%	0.5%	**	15.7%	15.7%	16.8%
CAR	94/95	3.9	46.4%	-0.9%	**	49.9%	45.5%	51.3%
Niger	92	3.4	10.3%	-0.5%	**	12.0%	12.0%	16.0%
Zimbabwe	94	2.5	83.6%	-0.6%	*	85.1%	87.3%	87.4%
Cameroon	91	2.5	62.0%	-1.3%		65.2%	80.2%	66.5%
Niger	98	2.4	14.5%	-0.3%	*	15.2%	18.2%	19.7%
Burkina Faso	92/93	2.4	19.5%	-0.7%	**	21.2%	21.2%	23.9%
Morocco	93/94	2.0	36.4%	0.0%		36.4%	36.4%	39.5%
Haiti	94/95	1.5	66.8%	-1.7%	**	69.4%	69.4%	693%
Uganda	95	1.4	67.9%	-0.5%		68.6%	71.4%	67.8%
Cote d' Iviore	94	1.1	41.7%	-0.1%		41.8%	45.3%	43.3%
Tanzania	91/92	1.0	45.5%	-0.3%	*	45.8%	45.8%	46.5%
Nigeria	99	0.9	59.3%	-0.2%		59.5%	61.8%	59.9%
Benin	96	0.9	33.7%	-1.4%	**	34.9%	44.7%	36.4%
Madagascar	92	0.9	52.2%	-2.3%	**	54.2%	54.2%	55.3%
Bolivia	93/94	0.8	83.1%	-0.8%	+	83.7%	85.5%	85.4%
Dom. Rep.	91	0.5	54.2%	-2.2%	*	55.3%	55.3%	55.6%
Philippines	93	0.4	75.2%	-0.1%		75.2%	75.2%	76.5%
India	98/99	0.4	76.0%	0.0%		76.0%	76.4%	76.6%
India	92/93	0.2	62.9%	-0.1%		62.9%	63.4%	65.0%
Bangladesh	92/93	0.2	73.2%	-3.1%	+	73.8%	74.0%	75.9%
Bangladesh	96/97	0.2	70.3%	-2.2%	*	70.6%	73.1%	73.4%
Average		2.0	49.8%	-0.9%		51.1%	53.1%	53.0%
Median		**1.3**	**53.2%**	**-0.7%**		**54.7%**	**54.7%**	**55.4%**

Source: Filmer, 2003.

a) Based on a linear probability model regression of enrollment for each child in a regression that, in addition to schooling availability included variables for the child (age, age squared, gender), household (dummy for top half of HHs by an asset index, highest schooling of males in the HH, highest schooling of females in the HH, gender of HH head, schooling of the HH head) and cluster (distance to facilities—e.g. post office, local market, bank, cinema).

Table A.2: achievement (positive means smaller class sizes raise achievement) using school fixed effects and instrumental variables to identify the class size effect						
	Mathematics			Science		
Country	Weighted Least Squares	School Fixed Effects	SFE-IV	WLS	SFE	SFE-IV
Australia	-4.33	-4.30	2.08	-3.65	-1.46	0.70
Belgium-Flemish	-2.18	-0.82	-8.09	-1.47	-0.48	-1.08
Belgium-French	-1.51	0.53	-0.80	0.58	1.70	0.67
Canada	-0.76	-0.20	-0.25	-0.09	-0.17	1.22
Czech Rep	-2.37	1.10	-2.67	-1.43	1.82	1.03
France	-2.59	-1.60	2.73	-0.55	-0.10	-0.14
Greece	-0.46	0.88	1.53	-0.29	-0.05	2.41
Hong Kong	-5.47	-4.06	5.22	-5.47	-3.50	12.98
Iceland	-0.16	0.44	2.59	1.01	0.47	0.16
Japan	-3.81	0.29	-0.07	-2.59	0.44	0.26
Korea	0.15	0.21	0.90	-0.19	-0.07	0.42
Portugal	-0.77	-0.85	-1.54	-0.17	-0.07	0.31
Romania	-2.14	-0.30	0.30	-1.43	-0.70	-3.31
Scotland	-2.52	-2.92	4.98	0.66	0.63	-31.58
Singapore	-4.69	-3.10	-0.45	-5.03	-3.47	-0.52
Slovenia	-0.52	0.05	-1.25	0.39	-0.13	-0.29
Spain	-0.17	0.10	0.31	-0.19	-0.10	0.70
United States	0.16	0.00	-20.26	-0.04	-0.15	1.28
Median	-1.82	-0.10	0.12	-0.24	-0.10	0.36
Fraction positive	11.1%	44.4%	55.6%	22.2%	27.8%	72.2%
Number positive, reject zero			2			2
Number fail to reject 0, reject effect as large as 3			10			9
Number fail to reject either zero or large as 3			6			6