Schools, Teachers, and Education Outcomes in Developing Countries

Paul Glewwe
University of Minnesota

Michael Kremer
Harvard University, Brookings Institution, Center for Global Development, and NBER

Second draft of chapter for *Handbook on the Economics of Education*

April 2005

We would like to thank Patricia Cortes, Hai-Anh Hoang Dang, Lynn Johnson, Muthoni Ngatia, Courtney Umberger, Heidi Williams, and Meng Zhao for outstanding research assistance. Some of the material in this chapter draws on Glewwe (2002), Kremer (2003), and Duflo and Kremer (2005). We thank Rachel Glennerster, Eric Hanushek and Kathleen Lynch for comments.

I. Introduction

Eight out of 10 of the world's children live in developing countries (World Bank, 2003). For economists working on education, the study of developing countries offers both policy questions of fundamental importance and a rich set of experiences to examine.

The important policy questions stem from the potential role of education in improving the welfare of the 5 billion people living in developing countries. Many macroeconomists have emphasized the impact of education on economic growth (Lucas, 1988; Barro, 1991; Mankiw, Romer and Weil, 1992) (although some others have raised questions about the causal relationship between education and economic growth).[1] Among microeconomists, both an older literature using ordinary least squares (OLS) regressions (Psacharopoulos, 1985 and 1994) and a newer literature using natural experiments and instrumental variable techniques (Duflo, 2001) estimate that both the private and social rates of return to education are high in developing countries. Education has also been found to play a crucial role in the adoption of new agricultural technologies in those countries (Foster and Rosenzweig, 1996). Finally, education is also seen as a means to improve health and reduce fertility (Schultz, 1997 and 2002; Strauss and Thomas, 1995) and is seen as an intrinsic good in itself (Sen, 1999, pp.292-97). Behrman (1999), Glewwe (2002), and Huffman (2001) provide recent reviews of the microeconomic literature on the impact of education on income and other outcomes in developing countries.

This support for education among economists is matched by equal or greater enthusiasm among development policymakers (UNDP, 1990; World Bank, 2001a). As discussed in Section II, developing countries have massively expanded their education systems in the last 40 years.[2] One example demonstrating the focus policy makers have placed on education is that two of the eight Millennium Development Goals (MDGs) adopted at the United Nations Millennium Summit in September 2000 focus on education: first, for all children to complete primary school by 2015, and second, to achieve gender equality at all levels of education by 2015.

The rich set of experiences worth examining includes wide variation in input levels and education systems across developing countries and, in recent years, dramatic policy changes and reforms in many developing countries. In addition, in the last 10 years randomized evaluations of education policies (which are rare in developed countries) have been undertaken in several developing countries. All of this makes the study of education in developing countries a potentially fruitful area of research.

In view of the widespread consensus on the importance of education and the existence of several reviews on the impact of education on income and other outcomes, this chapter focuses not on those impacts but rather on issues pertaining to the provision of education: namely, how education programs and systems affect the quantity and

[1] Pritchett (1996) argues that it is difficult to find a relationship between education and growth. Others, such as Krueger and Lindahl (2001), argue otherwise. Bils and Klenow (2000) argue the evidence favors a dominant role for growth impacting schooling. See also Levine and Renalt (1992, table 5) and Easterly (2001, pp. 71-85).

[2] Behind average figures on the remarkable expansion of schooling in developing countries lie educational miracles like Nepal, which increased primary enrollment from 10 percent in 1960 to 80 percent in 1990.

quality of education obtained by children in developing countries.[3] This focus implies that we do not examine the impact on schooling of factors outside the education system such as economic crises (Frankenberg *et al.*, 1999) orphan status (Case, Paxson and Ableidinger, 2003; Evans and Miguel 2004), or early childhood nutritional status (Glewwe, 2005).

Despite the tremendous progress in expanding enrollment and increasing years of schooling since 1960, 113 million children of primary school age are still not enrolled in school (UNDP, 2003), 94 percent of whom live in developing countries (UNESCO, 2002). In addition, the quality of schooling in developing countries is often very low. Grade repetition and leaving school at an early age are common, teachers are often absent from classrooms, and many children learn much less than the learning objectives set in the official curriculum (Lockheed and Verspoor, 1991; Harbison and Hanushek, 1992; Hanushek, 1995; Glewwe, 1999). Visitors from developed countries are often shocked at the conditions in many (but not all) schools in developing countries. Many schools lack the most basic equipment and school supplies—textbooks, blackboards, desks, benches, and sometimes even classrooms (in which case classes meet outside and are cancelled when it rains). In rural areas of Vietnam's Northern Uplands region in 1998, 39 percent of primary school classrooms did not have blackboards. In India in 1987, more than 8 percent of schools did not have a building in which to meet (World Bank, 1997). Teacher quality and availability is also a common problem. In rural areas of Northeast Brazil in the early 1980s, 60 percent of primary school teachers had not even completed primary education (Harbison and Hanushek, 1992). Shortages of teachers and school buildings can result in double shifts (which shorten the school day for individual pupils) or very large class sizes. In Vietnam, more than 90 percent of children in rural areas attend schools with two or more shifts, resulting in an average class time of only 3 hours and 10 minutes per day (Glewwe, 2004). In districts with low literacy rates in the Indian State of Tamil Nadu, the average class size in primary school was 78 students (World Bank, 1997). Teachers often have weak incentives and little supervision, and their absenteeism runs high. Chaudhury *et al.* (forthcoming) reports that when enumerators made surprise visits to primary schools in six developing countries, on average (across these countries) about 19 percent of teachers were absent. Beyond absence, many "present" teachers were found to not be actually teaching; for example, in India one quarter of government primary school teachers were absent from school, but only about half of the teachers were actually teaching in their classrooms when enumerators arrived at the schools.

The research discussed in this chapter suggests a number of conclusions, both substantive and methodological. First, additional children can be attracted to school at relatively low cost, either by reducing the cost of schooling and providing incentives for school attendance or by addressing basic health problems. Second, the evidence is mixed concerning the impact on learning of providing more educational inputs. Earlier surveys based on retrospective studies suggest that providing additional resources (given the

[3] Due to the dearth of rigorous research on post-secondary education in developing countries, we focus here on primary and secondary education. In addition, this chapter excludes the transition economies of Central and Eastern Europe, the Balkans, and the Commonwealth of Independent States (CIS); on labor markets and the impact of education on wages in those countries see Svejnar (1999).

existing education systems in developing countries) may have little impact on learning. More recent evidence from natural experiments and randomized evaluations paints a more mixed, but far from uniformly positive, picture. Third, education systems in developing countries are weak: education finance systems lead to budget distortions, incentives for teachers are weak or nonexistent, and curriculums are often inappropriate. Decentralization and school choice programs offer some promise, but their impact depends on the details of implementation. Finally, we offer methodological suggestions regarding the study of education in developing countries. In particular, we argue there is scope for increasing the use of randomized evaluations in assessing the impact of education programs in developing countries.

With regard to education initiatives in developing countries more broadly, some observers emphasize that schools need more money while others emphasize the weaknesses of the schools systems and the need for reform. While these two views are often placed at odds with each other, we argue they are not mutually exclusive; in fact, both may be true. By definition, highly distorted systems are such that marginal products have not been equalized across all expenditure categories. Thus, in settings with highly distorted education systems some types of spending will have low marginal product while others will have high marginal product.

The remainder of this chapter is organized as follows. Section II provides a general context for the chapter by giving background on primary and secondary education in developing countries. Section III outlines an analytical framework we will use in interpreting the studies discussed in this chapter. We then review selected empirical work: Section IV examines the factors influencing the quantity of education obtained; Section V focuses on education quality by examining the determinants of skills obtained while in school; and Section VI examines distortions in education systems, the political economy of education, and school reform initiatives. Finally, Section VII reviews methodological lessons and provides recommendations for future research on education in developing countries.

II. Education in Developing Countries

Almost every chapter in this *Handbook* focuses on education issues in developed countries. There are many differences between the education systems of developed and developing countries, so this section provides basic information on education in developing countries. Subsection A discusses trends in the quantity of education provided, subsection B discusses the persistent problems of school quality, and subsection C provides background on the more general issues of education finance, school organization, and education management policies.

A. Trends in the Quantity of Education: Enrollment, Years of Schooling, and Literacy

School enrollment rates and adults' years of schooling have increased dramatically in almost all developing countries since 1960 (the earliest year for which reliable data are available), but despite significant progress toward universal primary education and rapid increases in secondary school enrollment, there is still much room for improvement. In

2000, about 850 million adults (age 15 or older) in developing countries – 1 out of every 4 – were illiterate (UNESCO, 2002). This is in part because a sizable percentage of the adult population in these countries never attended school. This subsection examines some basic data on the quantity of schooling attained in developing countries and discusses current patterns by income levels, geographic region, and gender. In particular, it examines statistics on gross and net enrollment rates, rates of completion of 4 years of schooling, average years of schooling of the adult population, and adult literacy rates.

The most cited and most widely available indicator of the education quantity is the *gross enrollment rate,* defined as the number of children enrolled in a particular level of education, regardless of age, as a percentage of the population in the age group associated with that level. The age range for primary school is usually 6 to11 years. In 1960, primary school gross enrollment rates were 65 percent in low-income countries, 83 percent in middle-income countries, and over 100 percent in high-income countries (table 1).[4] By 2000, enrollment rates had reached or exceeded 100 percent in both low and middle income countries in all regions except Sub-Saharan Africa, where gross enrollment rates peaked at 80 percent in 1980 and then declined slightly.

Gross enrollment rates above 100 percent do *not* imply all school-age children are in school. First, grade repetition raises gross enrollment rates. For example, in a school system with 6 years of primary education, a 100 percent gross enrollment rate is consistent with 75 percent of children taking 8 years to complete primary school (because each child repeats two grades) and 25 percent of children never attending school. Second, gross enrollment rates are typically computed by comparing census data on the school-age population with Ministry of Education data on school enrollment, obtained from school principal's reports. In many countries, principals and teachers have incentives to exaggerate the number of students enrolled (PROBE Team, 1999). An example of this is from India; the official primary gross enrollment rate in 1993 was 104.5 percent, but household survey data for 1993 show a rate of 95.9 percent (World Bank, 1997). Both overreporting and grade repetition can cause reported gross enrollment rates to reach or exceed 100 percent even when many children never enroll in school.

An alternative measure of progress toward universal primary education is *net enrollment rates,* the number of children enrolled in a particular level of schooling *who are of the age associated with that level of schooling,* divided by all children of the age associated with that level of schooling. Net enrollment rates can never exceed 100 percent, and they remove the upward bias in gross enrollment rates cause by the enrollment of "overage" children in a given level (due to repetition or delayed enrollment). They do not, however, address overreporting in official data. Net enrollment rates are much lower than gross enrollment rates for low- and middle-income countries (table 2). The lower net rates for low- and middle-income countries reflect higher repetition of grades in those countries (table 2, column 3) and late school-starting age in many developing countries (table 2, column 4).

[4] This classification of countries is defined by per capita income in 1960. *Low-income countries* are those with a per capita income below $200 per year, *middle-income countries* are those with an income between $200 and $450, and *high-income countries* are those with an income greater than $450. These cutoff points, while arbitrary, yield about the same number of countries in each group.

Statistics on the percentage of children who have completed 4 years of schooling (table 2, column 5) are the most appropriate for assessing whether universal primary education has been achieved. Although the gross enrollment rates in 2000 were over 100 percent in both low- and middle-income countries, universal completion of primary school has not been attained in either group of countries. In 1999 only 80 percent of children in low-income countries and 88 percent of children in middle-income countries had completed 4 years of primary school.

Over the past 40 years, enrollment has increased dramatically at both the primary and secondary levels (table 3). However, progress in secondary enrollment has slowed in the past two decades. In both low- and middle-income countries the secondary gross enrollment rate increased by about 150 percent from 1960 to 1980, while the increase from 1980 to 2000 was 59 percent in low-income countries and about 51 percent in middle-income countries. Another way to see this is to note that from 1970 to 1980 middle-income countries increased their secondary enrollment ratio from 33 percent to 51 percent in only one decade, while low-income countries took 20 years (1980 to 2000) to increase from 34 percent to 54 percent. Middle-income countries progress slowed down sharply in the 1980s, increasing by only eight percentage points (51 percent to 59 percent) in that decade, although the increase was stronger in the 1990s (from 59 percent to 77 percent).

Trends in secondary gross enrollment rates from 1960 to 2000 differ substantially by region. For example, secondary school rates in South Asia, Latin America and the Middle East and North Africa were similar in 1960 (10 percent, 14 percent, and 13 percent, respectively), but by 2000 the rate in Latin America (86 percent) was much higher than in South Asia (47 percent) and the Middle East and North Africa (66 percent). Sub-Saharan Africa's performance over time has been slower than that of other regions. A final interesting comparison is between Latin America and East Asia. East Asia had a higher secondary enrollment rate than Latin America in 1960 (20 percent vs. 14 percent), but the rates in Latin American countries surged in the 1990s, so that the average rate in 2000 was 86 percent, compared to 67 percent in East Asia.

In low, middle, and high income countries, average years of schooling increased by about 3 years between 1960 and 2000 (table 4).[5] (See Pritchett, this volume, for further discussion of this issue.) If the 1.7-year increase in schooling from 1980 to 2000 in middle-income countries continues from 2000 to 2020, middle-income countries will reach a level of 7.6 years of education in 2020, slightly above the level of high-income countries in 1960. Thus middle-income countries are about 60 years behind high-income countries in the level of schooling of their adult population. Similarly, low-income countries are 10 to 20 years behind middle-income countries (their schooling level of 5.4 in 2000 was reached sometime between 1980 and 1990 in middle-income countries), or about 70 to 80 years behind high-income countries.

Literacy rates show similar trends (table 5): low-income countries are about 30 years behind middle-income countries, which are about 60 years behind developed countries (assuming the literacy rate for middle-income countries will increase from 85 percent in 2000 to 95 percent in 2020). There are some notable regional differences in the

[5] The increase for low-income countries is 3.6, but this comparison is biased because data are not available for China in 1960 and 1970. When China is excluded in 2000, adult years of schooling is 4.5, which implies a change of 2.9 years.

trends for adult years of schooling and literacy rates. In 1960, Sub-Saharan Africa, the Middle East and North Africa, and South Asia were similar in their years of adult schooling (about 1.5) and literacy rates (between 24 percent and 33 percent). By 2000, the Middle East and North Africa region had an average of 5.4 years of education, while South Asia had 4.6 years and Sub-Saharan Africa only 3.4 years. Yet in terms of literacy rates the ranking is different: in 2000 South Asia has a lower literacy rate than Sub-Saharan Africa. This apparent contradictory pattern most likely reflects greater inequality in the distribution of education in South Asia: 46 percent of adults 15 years and older in South Asia have no formal education, while 2 percent have completed some form of higher education; the analogous figures for Sub-Saharan Africa are 44 percent and 0.8 percent.

In many countries, gender disparities in access to education are significant. About 56 percent of the 113 million school-age children not in school are girls (UNESCO, 2002). In low-income countries, primary gross enrollment rates are 107 percent for boys and 98 percent for girls; this gender gap is wider at the secondary level, 60 percent for boys and 47 percent for girls (table 6). In middle-income countries, the primary-school enrollment gap between boys and girls is smaller (only 4 percentage points), and in secondary school girls have a slightly higher rate than boys. In high-income countries, there is almost no difference in primary enrollment rates, and girls have a slightly higher rate at the secondary level.

Major differences in gender gaps emerge across different regions of the world. In Latin America, East Asia, and Eastern Europe/Former Soviet Union and in the countries in the Organization for Economic Cooperation and Development (OECD), there is almost no gender gap at the primary level, although East Asian countries have a gender gap at the secondary level. In contrast, in Sub-Saharan Africa and Middle East/North Africa, gender gaps are sizable at both the primary and secondary levels. The largest gender gaps at both the primary and the secondary levels are in South Asia.

B. The Quality of Education: Resources and Academic Achievement

The focus thus far has been on *quantity* of education; however, the *quality* of education in many developing countries is low in the sense that children learn much less in school than the curriculum states they should learn (Lockheed and Verspoor, 1991; Harbison and Hanushek, 1992; Hanushek, 1995; Glewwe, 1999). This low quality is not entirely surprising because the rapid expansion of primary and secondary education in developing countries has strained those countries' financial and human resources.

Comparisons of education quality across countries require internationally comparable data on academic performance. The two main sources of such data are the TIMSS (Third International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) projects administered by the International Association for the Evaluation of Educational Achievement (IAEEA) and the PISA (Programme for International Student Assessment) project managed by the OECD.[6] The TIMSS, PIRLS and PISA data are primarily from developed countries, but they include a few, mostly middle-income, developing countries.

[6] The first and second studies that were precursors to TIMSS were undertaken between 1964 and 1984; the results are not comparable with those of the TIMSS, and very few developing countries were included.

The scores of students in grades 7 and 8 on the 1999 TIMSS mathematics test are shown in the first two columns of table 7. The two developed countries, Japan and the United States, have scores of 579 and 502, respectively. South Korean students scored even higher (587), and Malaysian students also performed well (519). Scores were generally considerably lower in other developing countries, ranging from 275 in South Africa to 467 in Thailand. In fact, the gap between these developing countries and the developed countries is underestimated because of the low secondary school enrollment rates in those countries (ranging from 40 percent in Morocco to 85 percent in Chile). Assuming that more academically talented students are more likely to remain in school, the scores from those developing countries are for students of above average talent.

Reading results for grade 4 students in 2001 are shown in the last column of table 7. All seven of the participating developing countries (Argentina, Belize, Colombia, Iran, Kuwait, Morocco and Turkey) have much lower performance than the three developed countries shown (France, the United Kingdom, and the United States). The PISA tests in mathematics and reading, which were administered to 15-year-old students, tell a similar story (table 8). South Korea outperforms all four developed countries in reading, and almost all in mathematics (the exception being Japan), but the other seven developing countries lagged far behind.[7] The percentage of students with very low reading skills was much higher in these seven countries than in the developed countries (ranging from 2.7 percent to 6.4 percent).[8] Again, the gap is probably underestimated because secondary school enrollment is well below 100 percent in almost all of these countries (except for South Korea).

A clear regional difference exists among the developing countries tested: two of the three East Asian countries (the exception being Indonesia) have test score means exceeding those of each of the five Latin American countries. Although Indonesia has lower scores than do most Latin American countries, one must bear in mind that in 2000 Indonesia's per capita income was about $730, while per capita incomes in five other Latin American countries ranged from $2,080 (Peru) to $7,690 (Argentina). This regional pattern, together with the small difference in adult years of schooling and adult literacy seen in the previous section, suggests that, if education played a role in East Asia's "economic miracle," it may have been as much due to the *quality* of education as to the quantity. (See Hanushek and Kimko, 2000, for a detailed examination of this role of school quality.)

Internationally comparable data are not available for very low-income countries, but the performance of students on achievement tests administered within many of these countries suggests that academic achievement is often very low. In Bangladesh, for example, Greaney, Khandker and Alam (1999) found that 58 percent of a sample of rural children age 11 and older failed to identify seven of eight presented letters, and 59 percent correctly answered only five or fewer of eight tasks requiring recognition of one-

[7] The PISA was administered to 15-year-old students enrolled in any kind of educational institution (including vocational and technical education). Many developed countries participated in 2000 (including all OECD countries), as did three developing countries (Brazil, Mexico, and South Korea). Six new developing countries (Argentina, Chile, Hong Kong, Indonesia, Peru, and Thailand) participated in 2002.
[8] In fact, the Brazil scores may be lower, because the 16 percent of Brazil's 15-year-old students, those who were in or below grade 6, were excluded from the assessment.

and two-digit numbers, writing one-digit numbers and recognizing basic geometric shapes. In Ghana, the mean score of grade 6 students on a very simple multiple-choice reading test was 25 percent, the score one would expect from random guessing (Glewwe, 1999). In India, 36 percent of grade 6 students were unable to understand and correctly answer the following question: "The dog is black with a white spot on his back and one white leg. The color of the dog is mostly: (a) black, (b) brown, or (c) grey" (Lockheed and Verspoor, 1991).

In summary, primary and secondary school students in most (but not all) developing countries learn less than their counterparts in developed countries. Moreover, these gaps are significant: mathematics (TIMSS) score disparities are equivalent to about a 3-year education gap between developed and developing countries.[9] These large gaps could reflect differences in family characteristics, but they almost certainly also reflect low school quality in developing countries.

C. School Finance and Education Systems

Government spending on education as a percentage of total gross domestic product (GDP) is similar across different groups of countries (table 9). The percentages are larger in high-income countries than in low-income countries but not remarkably so. Neither are the differences dramatic across regions. Yet since school age children are a much larger percentage of the population in developing countries, educational resources per child are typically lower in developing countries relative to GDP per capita. In low-income countries, spending per primary student is about 7 percent of per capita GDP, and this figure increases to 13.3 percent and 18.8 percent for middle- and high-income countries, respectively (table 9, column 3). In contrast, spending per secondary student as a percent of per capita GDP is much more similar (ranging from 15.5 percent to 21.5 percent).

Table 10 presents expenditures per pupil in U.S. dollars using two different methods, both revealing significant disparities due to large differences in per capita income. Using current exchange rates, middle-income countries outspend low-income countries by a ratio of 12 to 1 for primary education and about 8 to 1 for secondary education. Expenditures in high-income countries exceed those in low-income countries by a ratio of about 70 to 1 for primary education and about 50 to 1 for secondary education. Since expenditure on education is on nontraded goods and services (*e.g.*, teacher salaries), a better method to obtain comparable figures across countries is to convert local currencies to purchasing power parity (PPP) dollars, which account for price differences in nontraded goods and services across countries. This reduces the gaps somewhat. In primary education, middle-income countries spend 4 times more, and high-

[9] The 1995 TIMSS results show 30 to 40 point differences between seventh and eighth grade students in France, Japan, and the United States, suggesting that the 100 point gaps commonly found between developed and developing countries are equivalent to about 3 years of education. Yet it is worth noting that when the mean scores of some developing countries on the TIMSS, PIRLS or PISA assessments are two or three standard deviations below the mean in developed countries it may be difficult to use those scores to measure precisely the gaps between those countries and developed countries, since the tests were not designed to measure outcomes precisely at the extremes of the distribution found in developed countries.

income countries nearly 15 times more, than low-income countries. For secondary education, the analogous figures are 3 and 10.

Teacher salaries are by far the largest component of government expenditures on education in developing countries. According to a study of 55 low-income countries, on average, teacher salaries and benefits account for 74 percent of government recurrent expenditures on education (Bruns *et al.*, 2003). One reason for the high proportion of teacher salaries in education spending in developing countries is that low-income countries typically pay high teacher salaries, relative to GDP per capita, partly due to the scarcity of skilled workers in poor countries but also partly due to political economy factors. Countries respond to this high cost of teachers by maintaining large class sizes (table 11, columns 1 and 2). Sub-Saharan Africa and South Asia have the highest pupil teacher ratios. As a country develops, teachers' relative salaries decrease. (See Lakdawalla, 2001, for an analysis of the evolution of teachers' salaries and class size in the United States over the twentieth century.)

Developing countries also respond to the scarcity of trained teachers by hiring more untrained teachers. Whereas almost all teachers in developed countries are trained, in low-income countries, only 90 percent of primary school teachers and 69 percent of secondary school teachers are trained (table 11). The amount of training required for certification as a teacher varies, but requirements in poor countries are typically lower than in more affluent countries. The two regions with the smallest percentage of trained teachers at the primary level (data at the secondary level are less reliable) are Sub-Saharan Africa and South Asia, also the regions with the highest pupil-teacher ratios. These two regions simply have too few teachers to accommodate their rapid expansion in school enrollment. This is not surprising, since they also had the lowest years of schooling of the adult populations in 1990 and 2000 (table 4).

Most countries spend more per tertiary (post-secondary) student than per primary and secondary students, but the gap is much larger in developing countries (table 9, column 5). On average, governments in low-income countries spend 34 times more on a student in tertiary education than they spend on a student in primary education and 14 times more than on a student in secondary education. The analogous figures for high-income countries are 1.8 and 1.4. Since the poorest children rarely reach high levels of schooling, greater per student spending at higher (rather than lower) levels of education is likely to be regressive.

This low spending on primary and (to a lesser extent) secondary education in developing countries often implies that households bear much of the cost of that education. Thus parents, rather than the school or ministry of education, are responsible for providing many basic school inputs such as textbooks, chairs, and even the school building itself. Some of these costs are the collective responsibility of parents, but some are passed on to parents through official or unofficial school fees or by requiring parents to purchase uniforms and textbooks for their children. Data on such costs are not available for many countries, but a few examples are worth considering, although it is worth bearing in mind that they may not be representative. In Jamaica, government expenditures per primary school student are US$221 while private expenditures are $178 (Planning Institute of Jamaica, 1992). In the Philippines, the analogous figures are $110 and $309 (Asian Development Bank, 1999), and for Vietnam they are $23 and $14 (World Bank, 1997b). These figures include private school costs.

Aside from differences in education finance, education systems in developing countries differ in other ways from those in developed countries. In many developing countries, school systems are highly centralized and teachers' unions are strong. Teachers often have weak incentives and little supervision, and absence rates are high (table 12). A team of researchers who visited schools in India (PROBE Team, 1999) found some teachers who kept schools closed or nonfunctional for weeks or months at a time, drunken teachers, and a headmaster who expected the students to perform domestic chores and babysit. Sexual abuse of female students by teachers is a problem in several countries. To the extent that teachers do have incentives, these incentives are often focused on exam scores. Teachers often instruct by rote, sometimes copying from textbooks onto the blackboard and having students copy from the blackboard onto notebooks or slates.

The lack of teacher accountability in many developing countries may reflect the colonial legacy, the hierarchical nature of many developing societies, and the large gaps in education and social status between teachers and their pupils' families. In many countries, teachers offer, and pressure parents to pay for, "extra lessons" after school or on weekends to prepare students for important examinations (Bray, 1999). In such situations, increased teaching effort at school could reduce the demand for extra lessons, and thus teacher income.

Another unusual characteristic of many developing countries is that students are taught in a language that is not their mother tongue. This primarily reflects the fact that almost all developing countries were once colonies of developed countries, and their school systems still embody many elements of the systems developed under colonial rule. Many Sub-Saharan African countries use English or French as their national language, and most of India's 1 billion inhabits are not native speakers of either of the two official national languages (Hindi and English).

Given the heterogeneity in educational background, school quality, and language within many developing countries, designing a single curriculum appropriate for all students is difficult for any country. Yet most developing countries have a single centrally set curriculum, often geared to the needs of relatively elite students, which leaves many other students behind. This contributes to the poor performance of a significant percentage of students on national examinations and to high dropout and repetition rates. For example, in Tanzania between 1997 and 2001, only 22 percent of the students who attempted were able to pass the primary education final examination, and only 28 percent of those who attempted passed the certificate of secondary education exam (Tanzania Media Monitoring, 2002).

In response to the high cost and low quality of some centralized school systems, alternative, locally controlled systems have been established in some countries. These include nonformal education (NFEs) centers in India and the EDUCO schools in El Salvador. NFEs in India hire locally and pay a tiny fraction of regular salaries. Most teachers are not officially qualified. EDUCO schools in El Salvador allow local education committees to monitor teacher performance, hire and fire teachers, and manage school equipment and maintenance (Jimenez and Sawada, 1999).

The potential for competition among schools and for Tiebout sorting in developing countries is limited, since substantial proportions of the populations in developing countries reside in rural areas—68 percent in low-income countries, 50

percent in middle-income countries (World Bank, 2002). Rural areas are often characterized by low residential mobility, land markets subject to major transaction costs, and limited transportation networks. Nonetheless, some rural areas have sufficiently dense populations to allow for significant competition between schools. For example, one out of every four households in a rural area of Kenya sends their child to a school that is not the closest to their house (Miguel and Gugerty, 2002). Among middle school students in Ghana, at least 26 percent of those living at home do not attend the closest middle school (Glewwe and Jacoby, 1994).

Policies toward private schools in developing countries vary widely, from outright prohibition (Cuba, Sri Lanka) to heavy subsidization (Chile). Consequently, in some countries (Algeria, Mongolia, Tanzania), less than 1 percent of primary school students are enrolled in private schools. In other countries (Chile, Pakistan, Zimbabwe), nearly one half or more are enrolled in private primary schools.

In summary, in recent years, education systems in developing countries have rapidly expanded from a very low base, but there is still room for improvement in enrollment rates (especially net enrollment rates). In general, school quality in developing countries is low (in the sense that students in these schools do not learn as much as their counterparts in more developed countries), and per-pupil expenditures are often quite low as compared to high- or middle-income countries, even after adjusting for price differences. Finally, although schools in developing countries vary from country to country, many of these education systems are highly centralized and have weak teacher incentives.

III. Methodological Issues

A substantial and rapidly growing literature attempts to estimate the causal relationships underlying education outcomes in developing countries, and to formulate policy recommendations based on those estimates. To evaluate this body of literature, a methodological framework is needed to clarify the different types of causal relationships that one might try to estimate and to judge the credibility of the estimation methods used. This section provides such a general framework and discusses its implications for estimation. Subsection A outlines the framework that will be used to interpret the research discussed in later sections. Subsection B discusses estimation using retrospective data, and Subsection C discusses estimation using randomized trials and natural experiments.

A. Behavioral Models and Causal Relationships

To understand the impact of education policies on years of schooling and skills learned, a useful assumption for economists is that each household (in particular, the parents of the child) maximizes, subject to constraints, a (life-cycle) utility function. The main arguments in the utility function are consumption of goods and services (including leisure) at different points in time, and each child's years of schooling and learning. The constraints faced are the production function for learning, the impacts of years of schooling and of skills obtained on the future labor incomes of children, a life-cycle budget constraint, and perhaps some credit constraints or an agricultural production function (for which child labor is one

possible input), or both. The production function for learning is a structural relationship that can be depicted as:

$$A = a(S, \mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{I}) \qquad (1)$$

where A is skills learned *(achievement)*, S is years of schooling, **Q** is a vector of school and teacher characteristics *(quality)*, **C** is a vector of child characteristics (including "innate ability"), **H** is a vector of household characteristics, and **I** is a vector of school inputs under the control of parents, such as children's daily attendance and purchases of textbooks and other school supplies. Although children may acquire different skills in school, suggesting (1) should have multiple outputs and A should be a vector, for the purposes of this chapter little is lost, and some simplicity is gained, by treating A as a scalar.

Assume that all elements in the vectors **C** and **H** are exogenous. Examples of such variables are credit constraints, parental tastes for schooling, parental education, and children's genetic endowments of "ability." Some child characteristics affecting education outcomes (such as child health) could be endogenous; such variables can be treated as elements of **I**, all of which are endogenous.[10] Another important set of variables to introduce in this framework is prices related to schooling, denoted by the vector **P**. These prices can include school fees, prices for school supplies purchased by parents, and even wages paid for child labor. **P** does not appear in equation (1) because it has no direct effect on learning; its effect works through decisions made for the endogenous variables S and **I**.

In the simplest scenario, assume that only one school is available to each household and that nothing parents can do will change the characteristics of that school. Thus all variables in **Q** and **P** are exogenous to the household. Parents choose S and **I** (subject to the above-mentioned constraints) to maximize household utility, which implies that years of schooling S and schooling inputs **I** can be expressed as general functions of the four vectors of exogenous variables:

$$S = f(\mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{P}) \qquad (2)$$
$$\mathbf{I} = g(\mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{P}) \qquad (3)$$

Inserting (2) and (3) into (1) gives the reduced form equation for (A):

$$A = h(\mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{P}) \qquad (4)$$

This reduced form equation is a causal relationship, but it is not a production function because it reflects household preferences and includes prices among its arguments.

The more realistic assumption that households can choose from more than one school implies that **Q** and **P** are endogenous even if they are fixed for any given school. In this scenario, households maximize utility with respect to each schooling choice, and then choose the school that leads to the highest utility. Conditional on choosing that school, they choose S and **I**, as in the case where there is only one school from which to choose.

Policymakers are primarily concerned with the impact of education policies on years of schooling, S, and eventual academic achievement, A. For example, reducing class

[10] For a similar exposition that focuses on the role of child health, see Glewwe (2005).

size can be seen as a change in one element of **Q**, and changing tuition fees can be seen as altering one component of **P**. Equations (2) and (4) show how such changes would affect S and A. Assuming the cost of such changes is not difficult to calculate, the benefits in terms of increases in S and **A** can be compared to those costs. Of course, the costs should include costs borne by households from the policy change, so changes in **I**, as expressed in equation (3), and in household leisure must be included in the overall cost figure.

Consider a change in one element of **Q**, call it Q_i. Equation (1) shows how changes in Q_i affect A when all other explanatory variable are held constant, and thus provides the *partial* derivative of A with respect to Q_i. In contrast, equation (4) provides the *total* derivative of A with respect to Q_i because it allows for changes in S and **I** in response to the change in Q_i. Parents may respond to better teaching by increasing their provision of educational inputs such as textbooks. (Alternatively, if they consider better teaching a substitute for those inputs, they may decrease those inputs.) For example, Das and others suggest that household educational expenditures and governmental non-salary cash grants to schools are substitutes, and that households cut back on expenditures when the government provides grants to schools (Das *et al.*, 2003). In general, the partial and total derivatives could be quite different, and researchers should (but often do not) always clarify which relationship they are estimating. One possible reason (but not the only one) why different studies obtain different estimates of the factors that affect learning is that they are estimating different relationships.

When examining the impact of policies on academic skills, A, should policymakers look at equation (1), or equation (4)? Equation (4) is of interest because it shows what will actually happen to A after a change in one or more element in **Q** or **P**. In contrast, equation (1) will not show this because it does not account for changes in S and **I** in response to changes in **Q** and **P**. Although the total derivative obtained from equation (4) is of clear interest to policymakers, the partial derivative from (1) is also of interest because it may better capture overall welfare effects. Intuitively, if parents respond to an increase in Q_i by (for example) reducing purchases of inputs **I,** they will be able to raise household welfare by purchasing more of some consumer good. The reduced form impact (total derivative) reflects the drop in A due to the reduction in **I**, but it does not account for the increase in household welfare from the increased purchase of consumer goods. In contrast, the structural impact measured in equation (1) ignores both effects. Since these two effects have opposing effects on household welfare, they tend to cancel each other out, so the overall welfare effect is reasonably approximated by the change in A measured in equation (1). This is explained more formally in Glewwe and others (2004).

Results from randomized evaluations provide reduced form estimates of the impacts of changes in **P** and **Q**, and these reduced form parameters are total derivatives that reflect both the partial derivatives and agents' optimizing responses. For example, suppose school quality increases in some way. One possible response of parents is to reduce the time they spend helping their children with schoolwork. An education production function would not include this response, but a reduced form estimate (*e.g.*, by a randomized trial) would include both responses. Thus, if a researcher conducting a randomized trial wants to measure welfare, he or she should measure not only the program impact on the outcome variable, but also its impact on all other inputs. By combining these data with price data, a measure of the program's impact on welfare could be obtained.

This framework can be extended to examine policies that do not directly change **P** and **Q** but instead change the way schools are organized such as decentralization, promoting competition by removing restrictions on private schools, or developing incentive schemes that link teacher pay to student performance. In principle, these types of policies affect schooling outcomes by changing what happens in the classroom. For example, increased competition may change the behavior of teachers, and these behaviors can be included as components of the vector **Q**. Formally, education policies, denoted by **EP**, may interact with local community characteristics, denoted by **L**, to determine the quality of a school and even the prices of educational inputs in some cases (*e.g.* policies that allow communities to set school fees):

$$\mathbf{Q} = q(\mathbf{L}, \mathbf{EP}) \qquad (5)$$
$$\mathbf{P} = p(\mathbf{L}, \mathbf{EP}) \qquad (6)$$

Estimating equations (5) and (6) would require very detailed data on what happens in schools such as the many dimensions of teacher behavior. An alternative is to substitute (5) and (6) into (2) and (4) to obtain the reduced form relationships:

$$S = j(\mathbf{C}, \mathbf{H}, \mathbf{L}, \mathbf{EP}) \qquad (7)$$
$$A = k(\mathbf{C}, \mathbf{H}, \mathbf{L}, \mathbf{EP}) \qquad (8)$$

Knowledge of these functions would directly link education policies to the main outcomes of interest to policymakers.

The methodological framework presented in this subsection, while simple, is a useful guide for evaluating empirical work on education in developing countries. Yet it does have two limitations. First, it assumes a unitary household model and thus abstracts from bargaining among household members regarding education decisions. Indeed, common sense suggests that education decisions can be affected by household bargaining both between men and women and between parents and children. For example, Miguel and Kremer (2003) find that child social networks are as or more important than adult social networks in influencing take-up of school-based deworming programs. The framework presented above could be adapted to situations where adults disagree or children disagree with adults. In particular, the reduced form demands in equation (2), (3), (4), (7) and (8) can be supplemented with **C** and **H** variables that reflect the relative power of different household members such as individual wealth or income sources.

A second limitation of the methodological framework is that it abstracts from the general equilibrium effects of education policies. Changes in education inputs or education policy may eventually change the supply of educated adults and thus change the returns to education and thereby the demand for education. These relatively long-run impacts are, for the most part, ignored in the rest of this chapter.

B. Estimation Using Retrospective Data

Most empirical studies of the determinants of years of schooling and learning in both developed and developing countries are *retrospective studies*, based on data generated by ordinary (nonexperimental) variation across schools and households. This subsection

discusses the feasibility of using such data to estimate the relationships of interest discussed above, especially equations (1), (2), (4), (7), and (8). As we will see, there are formidable estimation problems even for this relatively simple scenario, and prospects dim further when more complicated scenarios are considered.

Consider estimation of equation (2), the (reduced form) determinants of years of schooling (S). For simplicity, assume that school quality (**Q**) and prices (**P**) are exogenous, the policies of interest can be adequately described by changes in the elements of **Q** and **P**, and there is only one school from which to choose (a relatively remote rural area, for example). Since **C** and **H** are also considered to be exogenous, OLS estimates of (2) will provide unbiased estimates of the causal parameters associated with each variable as long as one has (retrospective) data on S and on *all* the elements in the vectors **Q**, **C**, **H,** and **P**. In practice, it is neither necessary nor possible to have data on all elements in these four vectors. Data are not needed for any unobserved elements that are unlikely to be correlated with the variables in the four vectors for which one has data, so all such elements can be combined to form the error term in the regression equation.

Unfortunately, if any of these unobserved elements that are part of the error term are correlated with the variables for which one does have data, that correlation will lead to omitted variable bias in OLS estimates of the relationship being estimated. Such omitted variable bias is very likely: no retrospective data set will have data on all the elements in the vectors **Q**, **C**, **H,** and **P**, and it is very common for many of the unobserved elements to be correlated with some of the variables that are observed. Examples of variables that are almost impossible to observe (with the vectors they pertain to) are: the child's innate ability (**C**) and motivation (**C**), parents' willingness (**H**) and capacity (**H**) to help their children with schoolwork, teachers' interpersonal skills (**Q**) and motivation (**Q**), and the management skills of school principals (**Q**). When such data are missing from estimates of equation (2), OLS parameter estimates are likely to be biased because these variables are likely to be correlated with some of the observed variables in the regression. For example, schools that are "high quality" are likely to be high quality in many dimensions, both observed and unobserved. This will produce positive correlation between the error term and the observed school and teacher quality variables, leading to overestimation of the impact of observed school variables. Another example is parental tastes for their children's education, which is rarely observed and is likely to be positively correlated with parental education, leading to overestimation of the impact of parental education. When this type of bias occurs, it affects the estimated parameters not only for the observed variables that are correlated with the error term but also for the observed variables that are uncorrelated with the error term.

Researchers sometimes try to measure variables that they think are the most important omitted variables. For example, they may use an IQ test as a measure of innate ability or use parental schooling to indicate parents' ability to assist their children, but even here there are problems. It is not clear that innate ability can be measured: any test that claims to do so (in the sense of measuring a genetic endowment) almost always reflects environmental factors (American Psychological Association, 1995). One may be able to address this problem by using data on twins (*e.g.*, Behrman *et al.*, 1994), but such data from developing countries are very rare.

Measurement error in observed explanatory variables is another very difficult estimation problem. Anyone who has seen how household or school survey data are

collected in developing countries understands that even the best surveys contain a substantial amount of error. Data on school characteristics (including fees and prices of educational inputs) may be inaccurate or out of date. Indeed, they are often averages across grades and across classes within grades and thus do not match the experience of any particular child attending the school. Child and household variables can also be measured with a substantial amount of error in developing countries, including data on the age of the child, the distance to the nearest school, the education of the parents, and household resources (*e.g.*, land owned). Random measurement error typically leads to underestimation of the true underlying impacts, while nonrandom measurement error could lead to biases in either direction.

Even when parents cannot alter school quality, quality could be correlated with the error term if governments provide better schools to areas with unobserved education problems (Pitt *et al.*, 1993). On the other hand, governments are just as likely (and some observers would argue much more likely) to provide more school inputs in areas that already have good education outcomes, since these areas may have disproportionate political influence in both autocratic and democratic systems, may pay more taxes, and may put higher weight on education than other areas when choosing how to spend the resources they receive from the central government (World Bank, 2001a). Whatever the direction, correcting for this "endogenous program placement" bias is difficult.

One approach toward addressing the problems of omitted variables, measurement error, and endogenous program placement is instrumental variables. Unfortunately, it is often difficult to find plausible instruments – that is, variables correlated with the observed variables that are not orthogonal to the error term but uncorrelated with the error term. Some examples of this will be discussed in Section IV.

Now consider estimation of equations (1) and (4), the structural and reduced form determinants of learning, respectively. All of the above problems apply to these equations as well, and there is another problem: attrition bias. Communities with high-quality schools will keep children in school longer, leading to a student population with lower average innate ability (more "low ability" children stay in school). This will lead to underestimation of the impact of observed school quality on learning if no variable accurately measures innate ability.

Further complications arise for equations (1), (2), and (4) when allowing for endogeneity of school quality (and prices) in the sense that parents can choose from among more than one school, although they have no influence on the quality of any given school. Parents in remote rural areas may have little choice, so that all school characteristics are exogenous—but this is doubtful. First, parents may send their children to live with relatives (allowing them to attend a nonlocal school) or to a boarding school. About 19 percent of secondary students in rural Peru live away from their families (Gertler and Glewwe, 1990), and the same holds for 27 percent of middle school students in Ghana (Glewwe and Jacoby, 1994). Second, families with stronger tastes for educated children may migrate to areas with better schools, a common occurrence in the United States.

When parents can alter school quality through school choice, selection bias is possible if unobserved characteristics of children and households that affect test scores and years of schooling are correlated with unobserved factors that determine school choice. If data are available on some of the school choices, including schools not chosen, standard selection correction methods can be used (Heckman, 1979; see also Pagan and Ullah, 1999,

chapter 8). In particular, exclusion restrictions can be used to identify the generated selection correction term, namely the characteristics of the schools not chosen. In practice, however, modeling school choice in a tractable way may be difficult when many schooling options exist.

A final approach to consider is to abandon attempts to estimate equations (1), (2), and (4) because of the impossibility of collecting all the price and school characteristic variables in **P** and **Q** and instead estimate equations (7) and (8). An example of this, which will be examined in Section VI, is from Nicaragua, where some schools follow the "old" education policies and others follow the policies of the EDUCO program. In this case, all one needs is a dummy variable indicating which schools are EDUCO schools. This approach may be attractive if data on education policies and local characteristics, **EP** and **L**, are of lower dimension and therefore easier to collect than data on **P** and **Q**.

While this approach appears promising, it can still suffer from omitted variable bias if unobserved child, household, or community characteristics are correlated with the **EP** variables. In practice, retrospective estimates of equations (7) and (8) face many problems. First, in many programs, procedures that are supposed to be followed as part of a particular education policy are often not followed. Second, it is quite possible for unobserved child, household and community variables to be correlated with the new education policies, since the location of the programs may be affected by household actions (omitted variable bias) or government choices (endogenous program placement).

In summary, uncritical application of simple OLS regressions using retrospective data can lead to biased estimates of the impact of the determinants of learning and years of schooling. Some problems underestimate the impacts, others overestimate them, and some could go either way. These difficulties are so daunting that some economists doubt that they can be overcome (Hanushek, 1995). One response to these problems is to turn to randomized trials and natural experiments. Next, we review estimation issues that arise using these approaches.

C. Natural Experiments and Randomized Trials

Suppose equations (1) through (4), (7), and (8) could not be estimated using retrospective data, due to the problems raised above. An alternative approach is to exploit natural experiments generated by idiosyncratic details of policies that create instrumental variables for program participation that are plausibly uncorrelated with the error term for schooling outcomes (on natural experiments, see Campbell, 1969; Meyer, 1995; Rosenweig and Wolpin, 2000). Randomized trials are a third approach. Such trials are very common in medicine and are increasingly common in labor economics (Heckman *et al*., 1999; Manski and Garfinkel, 1992; and the special issue of the *Journal of Labor Economics*, 1993).

A few clarifications are in order regarding the use of randomized evaluations to estimate program effects. First, a distinction can be made about what exactly the evaluation is attempting to estimate. Randomized evaluations can be used to estimate the effect of a treatment on either the entire population subject to the randomization or on a subset defined by predetermined characteristics. In contrast, instrumental variable techniques estimate local average treatment effects, which are the effects on the population whose participation in the treatment was strongly influenced by the instrumental variable (Imbens and Angrist, 1994; Heckman *et al.*, 1997; Heckman *et al.*,

1998; Heckman *et al.*, 1999). In some settings, for example where enrollment is 100 percent, this distinction does not exist. In general, studies should clarify which type of treatment effect they are attempting to estimate. Second, randomized evaluations estimate partial equilibrium treatment effects, which may differ from general equilibrium treatment effects (Heckman, Lochner, and Taber, 1998). If some educational programs were implemented on a large scale, the programs might affect the functioning of the school system and thus have a different impact.

The basic idea of randomized evaluations of any kind is to compare two groups of observations that have no systematic differences other than one group received the treatment and the other did not. The simplest method is to sample a population of interest and randomly divide the sample into treatment and comparison groups. Under certain assumptions, discussed below, differences in the variables of interest across the two groups are unbiased estimates of the (reduced form) effect of the treatment.

Randomized trials and natural experiments typically do not estimate an education production function, that is equation (1), but they can provide reduced form estimates (total derivatives) of the impacts of changes in **C**, **H**, **P**, and **Q** on S, **I,** and A, as in equations (2), (3), and (4) or equations (7) and (8).[11] To the extent that some inputs provided by other actors, in particular the variables in **I** in equation (1), can be adjusted over different time horizons, the total derivative measured by these studies may be different in the short and the long run. For example, if treatment schools are provided with abundant supplies of a particular input, parents may reduce their efforts to supply these inputs. However, parents may not immediately throw out existing inputs such as parent-provided textbooks, and hence the stock of these inputs may decline only gradually over time. These studies can examine the reduced form impact of the program at various time horizons, and it can also measure the inputs provided by parents, but it does not directly measure the (partial derivative) impact of public provision of textbooks, holding parental provision constant.

Randomized trials can avoid, or reduce, some of the problems that arise with estimates based on retrospective data (discussed in subsection III.B). In particular, random assignment of observations into treatment and control groups implies that both observed and unobserved characteristics of those observations are uncorrelated in expectation with treatment status. Another problem that randomized studies should resolve is measurement error: in any well-managed study, treatment status should be measured without error.

Yet randomized evaluations do not address all of the statistical issues associated with retrospective analysis. Problems of selection and attrition bias remain, and randomized evaluations may generate new problems to the extent that people change their behavior because they know they are taking part in an evaluation. As we discuss in Section VII, a number of techniques can be used to address these issues. However, rather than discuss these issues in the abstract, first we review some examples.

[11] Technically speaking, equations (2), (3), and (4) show the relationship for all possible values of the variables in the vectors **C**, **H**, **P,** and **Q**, while a series of randomized trials can at most show a finite number of points on the "surface" of these relationships. The same point applies to equations (7) and (8).

IV. Factors Influencing the Quantity of Education Attained

The MDGs adopted in 2000 call for universal primary education by 2015, yet there is little consensus on how best to achieve this goal or on how much it would cost. One view holds that attracting additional children to school will be difficult, since most children not in school in developing countries are earning income their families need. Another view is that the potential contribution of children of primary-school age to family income is very small, which implies that modest incentives or improvements in school quality **Q** could significantly increase enrollment. Neither is there agreement on the role of school fees (elements of **P**). Some observers see fees as crucial for ensuring accountability in schools and as only a slight barrier to school enrollment; others contend that reducing school fees would greatly increase enrollment.

This section reviews the recent evidence on these issues. Subsection A discusses two general measurement problems that often arise when examining issues concerning the quantity of schooling in developing countries. Subsection B considers the tradeoff between investing in the construction of additional schools (making schools more accessible to students by increasing capacity and reducing distance) and investing in improving the quality of existing schools (making them more attractive). Subsection C examines the impact of reducing the cost of school or even compensating students for school attendance (changes in **P**), either implicitly (*e.g.* offering school meals) or explicitly. Because poor health may also limit school participation, Subsection D reviews recent work on school-based health programs (see Glewwe, 2005, for a review of recent work on the impact of health and nutrition in early childhood on education outcomes). We then discuss lessons from this work concerning the differential sensitivity of girls' and boys' schooling decisions (Subsection E) and the cost-effectiveness of various interventions to increase school participation (Subsection F). As we will see throughout this section, the evidence suggests that there are several promising avenues to increase the quantity of education attained by children.

Several of the studies discussed in this section examine both quantity of schooling and determinants of students' academic performance; in those cases we consider the findings with respect to the quantity of schooling in this section and report the findings on academic performance in the next section. Similarly, since grade repetition primarily reflects academic performance it will also be addressed in Section V.

A. Two Measurement Issues

This subsection discusses two measurement issues that often arise in research on the quantity of schooling in developing countries. First, defining what "quantity" should be measured can be difficult. Second, difficulties often arise when attempting to match the current and historical data on school and individual characteristics that are needed to investigate the factors affecting the quantity of schooling.

Measuring current school participation

The framework presented in Section III defined the quantity of schooling (S) as years of completed schooling, but in practice researchers look sometimes at the determinants of completed schooling and sometimes at measures of current schooling such as the

completion of a given level of schooling or the decision to drop out or continue to the next grade, both of which are incomplete measures of eventual years of schooling completed. One issue is whether educational inputs and education policies that increase the probability of staying in school or completing a given level of schooling will also increase years of schooling eventually completed, rather than simply creating intertemporal substitution in the timing of education.

Another issue is that in developing countries many pupils attend school erratically and the line between a "frequently absent pupil" and a "dropout" is often unclear. Attendance rates can vary dramatically among individuals. Thus large differences in the quantity of schooling would be overlooked by considering only years of schooling. One attractive way to incorporate wide variation in attendance when measuring the quantity of schooling is to focus on a more comprehensive measure of schooling often called "participation." For any child, *participation* is defined as the proportion of days that he or she is present in school for a given number of days that the school is open (*e.g.* Miguel and Kremer, 2003, 2004). This can be applied to a child's schooling over one or more years, or just for a few days for which reliable data are available. Participation differs from attendance in that it includes all children in the appropriate age range while *attendance* is usually defined only for children officially enrolled in school. Throughout the rest of this chapter, we use the terms *quantity of schooling* and *school participation* interchangeably. Both can be thought of as total time in school, which is imperfectly measured by years of schooling.

Classroom attendance registers are often very inaccurate in developing countries. One solution is to organize independent data collection in which unannounced observers visit schools a few days a year to record which children are actually in class.

Examining determinants of completed schooling

The other general measurement issue is that any individual's completed years of schooling (or some other measure of completed time in school) is known only many years after he or she first enrolled in school, which implies that data on years of schooling must be collected several years after data are collected on household and school characteristics. Thus cross-sectional data sets covering a relatively young population will include many children still in school, for whom the years of completed schooling variable is (right) censored. Alternatively, if cross-sectional data are collected from an older cohort for whom years of completed schooling is known, examining the impact of school characteristics on educational attainment requires historical data on school and household characteristics data. For example, consider a student who began school at age 6, left at age 16, and is surveyed at age 18: the relevant school characteristic data refer to a time period from 2 to 12 years before the time of the survey. Finding historical data on school quality in developing countries is often quite challenging, and matching it to individual students who attended those schools during those years is more difficult still. One common approach is to collect current school quality data and assume school characteristics have changed very little in the past 5 to 10 years, but, if this assumption is incorrect, the consequent measurement error could introduce serious biases into any econometric estimates.

B. Building Additional Schools versus Improving the Quality of Existing Schools

Many students in developing countries must travel long distances to attend school, so one policy option is to construct new schools in communities that have none. However, an inherent tradeoff exists between investing in the construction of new schools and investing in improvements in the quality of existing schools, which would make these schools more attractive to students. For example, the PROBE report (1999), based on in-depth surveys in five Indian states, argues that a key factor in low school participation is the low quality (unmotivated teacher incentives, weak curriculums, inadequate physical facilities) of available schools.

Several retrospective studies examine the impact of both distance to school (often measured by travel time) and school quality on the quantity of schooling. A number of concerns, particularly omitted variable biases, provide reason for caution in interpreting the results of these retrospective studies, but we here present the studies and some caveats which should be applied in interpreting their results. We then present results from natural experiments and randomized evaluations which likely offer more credible evidence on the relative impacts of distance to school and school quality.

A retrospective study in Ghana by Glewwe and Jacoby (1994) presents evidence on the impact of distance and school quality on the years of schooling of individuals aged 11 to 20, using data collected in 1988-89 on household, school, and teacher characteristics.[12] To estimate the impact of school characteristics and other factors on years of schooling attained, an ordered probit specification was used that allows for right censoring. According to the study findings, years of schooling was reasonably responsive to school quality. The estimates indicate that years of completed schooling could increase by 2 to 2.5 years by raising average teacher experience (from 2 years to 10 years), repairing leaking roofs, reducing travel time (from 2 hours to a few minutes), or providing blackboards to schools without them. Since repairing roofs and providing blackboards is much less expensive than building new schools, these results suggest that repairing classrooms in Ghana is a more cost-effective means of increasing the quantity of schooling than building new schools to reduce travel time.

Although the results from the Ghana study appear plausible, the estimates could be biased for a number of reasons. The data had 18 school and teacher variables, but schools can differ in many more ways, which raises the problem of omitted variable bias. Measurement error in these variables is also a potential problem, either because the assumption that they change little over time is false or because errors were made in collecting the data. Finally, no attempt was made to avoid bias due to endogenous program placement.

In a retrospective study in Tanzania, Bommier and Lambert (2000) found that distance to school had a significantly negative effect on years of schooling, while the quality of Swahili teaching had a positive effect. However, the authors note some

[12] Lavy (1996) used the same data to study the impact of secondary school characteristics (particularly distance) on primary school attainment. He found that the secondary school distance variables had significantly negative impacts. The school quality variables were almost always insignificant, perhaps because they were aggregated up to 33 regions, which reduced their variation and introduced random measurement error.

problems with measurement error in these variables. For example, many households reported implausible distances to the nearest primary schools, sometimes more than 100 kilometers. Moreover, since school characteristics were averaged over responses given by households, there could be systematic bias. For example, parents may "justify" a decision to allow a child to drop out of school by claiming that the local school was of low quality (in the Ghana study school quality variables were collected from schools, not households). Finally, given that there are only four school quality variables, there are serious concerns of omitted variable bias.

Using retrospective data from India, Drèze and Kingdon (2001) found that several school quality variables had statistically significant effects on years of primary school attained: both provision of a mid-day meal and "waterproof" classrooms had no effect on boys but strong positive effects on girls; teacher absences due to nonteaching duties had a negative effect on boys but no effect on girls; a parent-teacher cooperation index had a positive effect on both sexes; and class size had a negative effect on both sexes. Though plausible, these results should be interpreted very cautiously. Omitted variable biases are likely. Indeed, the authors suggest that the strong and significant effect of "waterproof" classrooms could also be interpreted as representing the general state of the school building. They also suggest that the (unobserved) motivation of school principals, parents, or both may be the real reason for both higher quality of schooling and the associated higher quantity of schooling.

Another strand in the literature looks at "natural experiments." Case and Deaton (1999) examined education outcomes in South Africa using data collected in 1993, when government funding for schools was highly centralized and blacks (people of African descent) had virtually no political representation of any kind. The authors argue that blacks did not control the funds provided to their children's schools and that tight migration controls limited their ability to migrate to areas with better schools. They show that pupil-teacher ratios varied widely across black schools, and argue that this variation, combined with migration barriers and black South Africans' lack of control over their schools, generates a kind of natural experiment.

Case and Deaton's estimates indicate that raising school resources (as measured by student-teacher ratios) increases years of completed schooling and enrollment rates for blacks but not for whites. Since blacks had much larger class sizes than whites, this is consistent with the idea that there are diminishing returns to reductions in class size. They estimate large effects from reducing class size at black schools: decreasing the student teacher ratio from 40 to 20 (the approximate means in black and white schools, respectively) increases grade attainment by 1.5 to 2.5 years.

Several issues raise concerns about the interpretation of these results. A key point is that, even if blacks could not influence class size in their children's schools, someone, presumably some government officials, made decisions that influenced class sizes in South Africa's black schools. If these decisions were influenced by education outcomes in those schools, or were merely correlated with such outcomes for some reason other than the causal impact of class size, they could yield biased estimates of the impact of class size (and, more generally, school resources) on those outcomes. This is the problem of endogenous program placement discussed in Section III. Another issue is that the children tested were not a random sample of household members, and data on student-teacher ratios

from the Ministry of Education are not highly correlated—an R^2 coefficient of 0.15—with the authors' community data for that variable.

In another natural experiment, Duflo (2001) took advantage of a rapid school expansion program in Indonesia to estimate the impact of building schools on years of schooling attained (as well as on subsequent wages, which is beyond the scope of this chapter). In 1973, the Indonesian government decided to use a portion of its oil revenues to build more schools. The allocation rule for the schools was known (more schools were built in places with low initial enrollment rates), and cohorts participating in the program are easily identified (children 12 years or older when the program started did not participate in the program). Duflo found that the school construction policy was effective in increasing the quantity of education and calculates that each school built for every 1,000 children led to an average increase of 0.12 years of education. Trends across regions were parallel before the program and shifted clearly for the first cohort exposed to the program, which raises confidence in the identification assumption.

Chin (2002) also takes a natural experiment approach to estimate the impact of placing additional teachers in Indian schools, an investment in school quality. "Operation Blackboard," a recent major policy initiative in India, addressed low primary school enrollment rates by mandating the provision of a second teacher to all primary schools with a single teacher. Chin (2002) evaluated the second teacher placement program and found that the program helped girls but had no effect on boys: girls' primary school completion increased by 3 to 4 percentage points, and the girls' literacy rate increased by 2 to 3 percentage points. Identification is again based on the fact that cohorts participating in the program are easily identified (only children attending primary school after 1987 were exposed).

A third strand of literature is based on randomized evaluations. As discussed in subsection E, Banerjee and others (2000) find that provision of additional teachers (usually female) in Indian nonformal education centers increased school participation by girls. A number of randomized evaluations recently done in Kenya (most of them discussed in Section V) found that programs designed to improve school quality, for example, by providing inputs like textbooks, had no detectable effect on school participation, and limited effects on test scores (see, for example, Glewwe, Kremer and Moulin 2002). Programs that reduced the cost of schooling or provided incentives to attend school had a much greater impact on school participation, as discussed below.

C. Reducing the Cost of Education

In many developing countries, parents face significant private costs of education for school fees and required inputs such as uniforms. For example, in Kenya parents have historically been required to purchase uniforms that cost about $6—a substantial expense in a country with a per capita income of $340. One simple way to increase the quantity of schooling would be to remove financial barriers by reducing the cost of school or paying students to attend. However, the desirability of school fees is much debated. Proponents argue that fees are necessary to finance inputs, that they increase parental participation in school governance, and that the price elasticity of demand for schooling is low (Jimenez and Lockheed1995). Opponents argue that school fees prevent many students from attending school and cite dramatic estimates from Sub-Saharan Africa: when free

schooling was introduced in Uganda in 1997, primary school enrollment reportedly doubled from 2.6 million to 5.2 million children (UNICEF 1999); when primary school fees were eliminated in Tanzania in 2002, an estimated 1.5 million students (primarily girls) began attending primary school almost immediately (Coalition for Health and Education Rights, 2002); and when Kenyan President Mwai Kibaki eliminated primary school fees in late 2002, a massive influx of new students reportedly overwhelmed school systems in certain districts.[13] While there can be little doubt that eliminating school fees generated a large enrollment response, the magnitudes cited in these accounts should be taken with a grain of salt. The data on which they are based are often unclear, and free schooling is sometimes announced simultaneously with other policy initiatives and often accompanied by programs that replace school fees with per-pupil grants from the central government, which create incentives for schools to over report enrollment.

Several recent randomized evaluations examine the impact of reducing costs on the quantity of schooling. Kremer and others (2002) conducted a randomized trial in rural Kenya to evaluate a program in which a nongovernmental organization (NGO), Internationaal Christelijk Steunfonds Africa (ICS), provided uniforms and textbooks and built classrooms for 7 schools randomly selected from a pool of 14 poorly performing schools. Dropout rates fell considerably in the 7 schools selected for participation, and after 5 years pupils in those schools had completed about 15 percent more years of schooling. In addition, many students from nearby schools transferred into program schools, raising class size by 50 percent. This suggests that students and parents were willing to trade off much larger class sizes for the benefit of free uniforms, textbooks, and improved classrooms. The authors argue that the main reason for the increase in years of schooling is most likely the financial benefit of free uniforms. A randomized trial of textbook provision in Kenya, discussed in the next subsection, showed almost no impact of textbooks on the quantity of schooling, and while the new classroom construction may have had an impact, the first new classrooms were not built until the second year of the program, while dropout rates fell dramatically in the first year. Anticipation of later classroom construction may have influenced these results, but the authors doubt it, because effects were present for students in the upper grades who would have finished school by the time the classrooms were built.

Several programs have gone beyond simply reducing school fees by actually paying students to attend school, in the form of either cash grants or school meals. Perhaps the best known randomized evaluation is the PROGRESA program in Mexico, which was designed to increase school enrollment and academic performance by paying cash grants to mothers conditional on their children's school attendance and participation in preventative health measures (nutrition supplementation, health care visits, and health education programs). When the program was launched in 1998, 506 communities participated, half of them randomly selected. Schultz (2004) finds a 3.4 percent increase in enrollment, on average, for all students in grades 1 through 8. The largest increase, 14.8 percent, was among girls who had completed grade 6. Using a difference-in-difference estimator, Schultz finds that PROGRESA increases educational attainment for the poor by 0.66 years, with a particularly large impact on the enrollment in the transition year to junior secondary school (20 percent for girls and 10 percent for boys). In part

[13] Lacey, Marc (2003) "Primary Schools in Kenya, Fees Abolished, Are Filled to Overflowing," *The New York Time*s, 7 January: A8.

because these evaluations clearly documented the program's success, PROGRESA was subsequently expanded to urban communities and, with support from the World Bank, similar programs are being implemented in several neighboring Latin American countries (*e.g.*, the PRAF program in Honduras).[14] Schultz (2004) estimates that if the current neighboring urban wage differentials approximate what PROGRESA program beneficiaries can expect to earn from their schooling in terms of future percentage increases in ages, the internal rate of return to the educational grants provided by PROGRESA is 8 percent per year in real terms (adjusted for inflation).

Conditional transfers such as those awarded through the PROGRESA program leave open one potential problem that, in some contexts, the people administering the program may not enforce the conditionality (Sen, 2002). Linden and Shastry (2005) provide evidence that teachers mis-represented student attendance in response to a program which provided grain to students who regularly attended school. In these circumstances, school meals may provide a stronger incentive to attend school, because children must go to school to receive the rations. Government-subsidized school meals have been provided in India, Bangladesh, Brazil, Swaziland, and Jamaica to increase both enrollment and attendance (World Food Program, 2002). Proponents of school meals also claim that school meals can increase both the quantity of schooling and academic performance by improving child nutrition. Others argue that families may reduce resource allocation to children who receive school meals. However, school meals would nonetheless serve as an incentive for families to send children to school. Moreover, Jacoby (2002) presents evidence from the Philippines that parents do not reduce food provided at home in response to school feeding programs (see also Long, 1991 and Powell *et al*., 1983). The Drèze and Kingdon (2001) study discussed in subsection B examined, among other variables, the impact of providing mid-day meals, which increased years in primary school for girls but not for boys.

Vermeersch and Kremer (2004) conducted a randomized evaluation of the impact of school meals on participation in Kenyan preschools, and found that school participation was 30 percent greater in the 25 Kenyan preschools where a free breakfast was introduced than in the 25 comparison schools. There was some evidence the provision of meals cut into instruction time. In schools where the teacher was relatively well trained prior to the program, the meals program led to higher test scores (0.4 of a standard deviation) on academic tests. There were no effects on tests of general cognitive skills, implying the school meals program did not improve children's nutritional status and that the academic test score increases were likely due to more time spent in school.

D. School-Based Health Programs

Poor health may also limit school participation: for example, intestinal helminths (*e.g.*, hookworm, roundworm, whipworm, and schistosomiasis) affect a quarter of the world's population, and are particularly prevalent among school-age children. Miguel and Kremer (2004) used randomized methods to evaluate a program of twice-yearly school-based

[14] Morley and Coady (2003) review and evaluate several Conditional Transfer for Education programs (CTEs) that have been implemented in developing countries, including PROGRESA, and conclude that CTEs are effective instruments for reducing poverty and increasing school enrollments.

mass treatment with inexpensive deworming drugs in Kenya (where the prevalence of intestinal worms among children is very high).They found that child health and school participation (as defined in subsection A) improved not only for treated students but also for untreated students at treatment schools (22 percent of pupils in treatment schools did not receive deworming medicine) and untreated students at nearby nontreatment schools due to reduced disease transmission. The authors used two approaches to address identification issues that arise in the presence of these disease-reduction externalities. First, randomization at school level allows them to estimate the overall effect of deworming on a school even if there are treatment externalities among pupils within treatment schools. (The authors use non-experimental means to decompose the overall effect on treatment schools into a direct effect and a within-school externality effect.) Second, cross-school externalities—the impact of deworming for pupils in schools located near treatment schools—are identified using exogenous variation in the local density of treatment school pupils generated by the school-level randomization. The authors find that absenteeism in treatment schools was 25 percent (7 percentage points) lower than in comparison schools. This reflects both the direct effect of deworming and any within-school externalities. Including the cross-school externalities, they find that deworming increased schooling by 0.15 years per pupil treated: decomposed into an effect of the treatment on the students treated and a spillover effect, school participation on average increased by 7.5 percent among pupils in treatment schools and by 2 percent among pupils in comparison schools.

Bleakley (2002) provides retrospective estimates of the effects of deworming from the United States. He finds that areas in the US South with higher hookworm infection levels prior to the 1910-1920 Rockefeller Sanitary Commission deworming campaign experienced greater increases in school attendance after the intervention, and estimates that each case of hookworm reduced the number of children attending schooling by 0.23 (similar to the estimates of Miguel and Kremer 2004). Although it is difficult to fully rule out omitted variable bias using a non-experimental approach, an important strength of Bleakley's work is that the Rockefeller campaign was introduced throughout a large geographic area, and thus the estimates are not subject to the biases faced by medical studies that randomize treatment at the individual level.

Bobonis, Miguel and Sharma (2004) find evidence from a randomized evaluation conducted in India that health programs can raise preschool attendance in urban areas. While in the Kenyan sample 92 percent of surveyed primary school pupils had at least one helminth infection (Miguel and Kremer, 2004), in this sample of Indian preschoolers "only" 30 percent were found to have worm infections but 69 percent of children were found to have moderate to severe anemia. The program therefore provided both iron supplementation and deworming medicine to these preschool students. After five months of treatment, the authors found large weight gains and a 5.8 percent reduction in absenteeism among 4 to 6 year olds (but not for younger children).

These findings that school health programs can increase the quantity of schooling raise the question of how best to implement such programs in developing countries. One view is that reliance on external financing of medicine is not sustainable and instead advocates health education, water and sanitation improvements, or financing the provision of medicine through local cost sharing. Kremer and Miguel (2003) analyzed several deworming interventions, including numerous "sustainable" approaches such as

cost sharing, health education, verbal commitments (a mobilization technique), and improvements in sanitation (all but the sanitation efforts were examined with randomized evaluations). Overall, their results suggest that there may be no alternative to continued subsidies for deworming. The "sustainable" public health strategies of health education, community mobilization, and cost recovery were ineffective, while provision of free deworming drugs led to high drug take-up and large reductions in the incidence of serious worm infections. A related paper (Miguel and Kremer, 2003) examines data on social networks to explore the effects of variation in social contacts' program exposure on individuals' adoption decisions. The authors found that children with (randomly) more social links to early treatment schools are themselves significantly less likely to take deworming drugs (perhaps because they learn that the drugs work for only a few months and seek to free ride on others' use of the drugs).

E. Gender and School Participation

There is some evidence that the elasticity of demand for schooling may be higher for girls than for boys, so that even policies and programs that do not specifically target girls may result in greater increases in school participation for girls than for boys. Many of the studies described above support this hypothesis: Chin (2002), regarding placement of additional teachers in schools; Drèze and Kingdon (2001), on the provision of mid-day meals; and both Schultz (2004) and Morley and Coady (2003), in their evaluations of PROGRESA.

Subsection C discussed several types of programs that reduced households' cost of schooling for both boys and girls, but an alternative is to implement programs that specifically target girls. For example, research in several countries suggests that one way to increase girls' school enrollment may be to hire female teachers (World Bank, 2001b; Herz *et al*., 1991; Rugh, 2000). However, it is very difficult to assess causality without conducting a randomized evaluation since in regions that are more open toward women's education, more women will obtain the education needed to become teachers.

Banerjee and others (2000) used a randomized evaluation to examine the impact of a program in India that attempted to raise school quality by hiring additional teachers, especially female teachers. An Indian NGO, Seva Mandir, runs nonformal schools that teach basic numeracy and literacy skills to children who do not attend formal schools and, in the medium term, attempts to "mainstream" these children into the regular school system. These schools are plagued by high teacher and child absenteeism, so the NGO decided to evaluate the impact of hiring a second teacher (where possible, a woman) in the hope of increasing the number of days the school was open, increasing student attendance, improving performance through individualized attention to students, and making school more attractive to girls. The program reduced the number of days a school was closed (one-teacher schools were closed 44 percent of the time, whereas two-teacher schools were closed 39 percent of the time), and girls' attendance increased by 50 percent. However, the program had no significant effect on the attendance of boys. One possible interpretation is that more girls are at the margin of choosing between some schooling and no schooling and that they would have been attracted to school by additional teachers independent of the teachers' gender. Another interpretation is that girls were attracted by hiring female teachers. Some weak support for the latter

hypothesis is provided by the fact that the effect on girls' enrollment was smaller when the original teacher was female. This is consistent with the possibility that the presence of at least one female teacher is important in providing a role model for girls but that the addition of a second female teacher has a comparatively minor role-model effect. There is no clear evidence of the program impacting test scores either positively or negatively.

Research on girls' scholarship programs is limited but suggests that scholarships can have major impacts on girls' enrollment rates. Research on a small fellowship program in Pakistan that subsidized girls' primary education in private schools was shown to be successful in urban areas but a failure in rural areas (Kim, Alderman, and Orazem, 1999; Alderman, Kim, and Orazem, 2003).[15] A national scholarship program for girls in rural Bangladesh increased girls' enrollment rates even after controlling for other measurable influences (World Bank, 2001b). Because with economic development enrollment of girls usually rises (and the gender gap between boys' and girls' enrollments narrows), it is potentially very problematic to draw conclusions from before and after comparisons of girls' enrollment rates. This difficulty highlights the need for randomized evaluations of such programs.

Kremer, Miguel, and Thornton (2004) conducted a randomized evaluation of the Girl's Scholarship Program (GSP), introduced in rural Kenya in late 2001 to enhance girls' education. Out of a set of 128 schools, half were randomly chosen as schools eligible for the program. The program consisted of a merit-based scholarship awarded to girls in two districts of Western Kenya who scored in the top 15 percent on tests administered by the Kenyan government. One portion of the scholarship was paid directly to the school for school fees, the other portion to the family for school supplies and uniforms. Girls eligible for the scholarship had significantly higher school attendance rates (as well as significantly higher test scores, average gains of 0.12-0.19 standard deviations). Schools offering the scholarship also had significantly higher teacher attendance after the program was introduced, and there is evidence of positive program externalities on boys (who were ineligible for the awards) as well as on girls with low pre-test scores (who were unlikely to win awards).

F. Summary

The studies discussed in this chapter provide mixed evidence on the extent to which school participation responds to school quality but suggest that it is fairly responsive to incentives.

Anecdotal evidence from East Africa, as well as randomized studies in Kenya and Mexico, shows sizeable impacts on school participation from reducing the cost of schooling (including subsidies conditional on school attendance). Randomized studies in Kenya and India demonstrate that school health programs can also increase the quantity

[15] The authors note several reasons for the relative success of the program in urban schools in contrast to rural schools. First, the latent demand for girls' schooling was higher in the urban areas. Second, urban parents were able to pay more than rural parents. Third, urban schools could take advantage of economies of scale to reduce costs per pupil. Finally, urban schools found it much easier than rural schools to attract good teachers (especially female teachers).

of schooling. Finally, several retrospective and randomized studies provide evidence that girls' school attendance is particularly elastic.

Many of the studies based on natural experiments and randomized evaluations are limited in that a central policy concern for developing countries is the *relative cost-effectiveness* of various interventions to increase school participation. Evaluations of cost-effectiveness require knowledge of a program's costs as well as its impact, and comparability across studies requires some common environment. Comparing the impact of PROGRESA's cash transfers and school meals in Kenya is difficult, since it is unclear whether the resulting differences are associated with the type of program or the larger environment. Policymakers are usually left with an unappealing choice between retrospective studies, which allow comparisons of different factors affecting school participation but may yield biased estimates, and randomized evaluations, which yield credible estimates but only for a single programs.

One exception to our general inability to compare cost-effectiveness of credible estimates is the recent set of studies conducted in Kenya. Because the Kenyan programs discussed in this section were conducted in similar environments, cost-effectiveness estimates from these randomized evaluations can be readily compared (see Poverty Action Lab 2005). Deworming was found to be extraordinarily cost-effective at only $3.50 per additional year of schooling (Miguel and Kremer, 2004). In contrast, even under optimistic assumptions, provision of free uniforms would cost $99 per additional year of school participation induced (Kremer *et al.,* 2002). The school meals program, which targeted preschoolers rather than primary school age children, cost $36 per additional year of schooling induced (Vermeersch and Kremer, 2004). This suggests that school health programs may be one cost-effective way of increasing school participation, and the Bobonis and others results for India suggest that this conclusion may be relevant in low-income countries outside Sub-Saharan Africa. More research on school-based health programs in developing countries is needed to confirm this.

V. Empirical Results on Quality: Factors Affecting Skills Obtained in School

Increases in the quantity of education in developing countries could be jeopardized by weaknesses in the quality of education. The success since 1960 in expanding the quantity of education in most developing countries (see Section II) has shifted attention to education quality, especially as measured by student performance on academic tests. This section examines recent research that attempts to identify the impact of school and teacher characteristics (**Q**) on learning in primary and secondary schools in developing countries. The first subsection reviews retrospective studies, and the second subsection examines "natural experiments" and randomized trials. The final subsection concludes.

A. Retrospective Studies

Many researchers, both economists and other social scientists, have used retrospective data to investigate the impact of school and teacher characteristics on learning. Hanushek's (1995) review of the evidence up to the mid-1990s draws the pessimistic conclusion that there is little empirical evidence that commonly used educational inputs

raise students' test scores in developing countries. To support this claim, he presents evidence from 96 studies, summarizing the findings for six educational inputs: teacher-pupil ratio, teacher's education, teacher's experience, teacher's salary, expenditure per pupil, and physical facilities (table 13). Based on the results in table 13, Hanushek concludes that, except for physical facilities, measured resources are not systematically related to student performance. Kremer (1995) points out that an alternative interpretation of the studies in table 13 is that almost all of the school inputs in the table raise test scores (the exception being the teacher-pupil ratio) because the probability is very small that several studies will find a statistically positive coefficient when the real coefficient is zero or negative. Even so, Kremer notes that improvements in student performance may be modest for some inputs and thus may not be worth the costs.

A third and perhaps most reasonable interpretation is that the simultaneous presence of so many significantly positive and negative coefficients suggests that either the studies do not measure the same parameter or the estimates are biased. This would be the case, for example, if there were omitted variable bias in many of these estimates, with some of the estimates having a positive omitted variable bias and some having a negative omitted variable bias. Drawing any definite conclusions from these data is difficult without knowing more precisely what the parameters represent (including whether they are structural production functions or reduced form relationships) and what biases may be present in the estimates.

Since the mid-1990s, the most significant recent retrospective studies of the determinants of learning in developing countries are: the research of Ghanaian middle schools by Glewwe and Jacoby (1994); the study of Jamaican primary schools by Glewwe and others (1995); the investigation of grade 8 students in India by Kingdon (1996); and the paper on Philippines primary schools by Tan and others (1997).[16] We first review the results of these retrospective studies and then provide a summary of critiques and concerns over why the results of these and other retrospective studies need to be interpreted very cautiously. In the following sub-sections we will then review evidence from natural experiments and randomized evaluations which allow for more credible estimation of factors impacting school quality.

The study by Glewwe and Jacoby (1994) on Ghana discussed in Section IV also examined student achievement in 1988-89, using scores on reading (English) and mathematics in Ghanaian middle schools (grades 7 to 10). Eighteen school and teacher variables were examined, but most estimated effects were small and statistically insignificant. The only statistically significant teacher variable was teaching experience, but its effect was indirect: it raised children's grade attainment, which increased both reading and mathematics test scores. In contrast, school facilities had larger impacts. The estimated impact (direct plus indirect) of repairing leaking classrooms was an increase of 2.0 standard deviations in reading scores and 2.2 in math scores; this impact seems to operate by reducing school closings due to rain. Blackboards also had large estimated impacts (direct

[16] A very recent study by Bedi and Marshall (2002) presents regressions on the factors that determine reading (Spanish) and mathematics scores of Honduran primary school students, but the impacts of the teacher and school variables on the scores of second grade students are so different from the impacts on fourth grade students that the authors conclude that they "are unable to identify clear-cut, policy-relevant variables that influence educational achievement (p.147)" despite very large samples of more than 7,000 second grade students and more than 5,000 fourth grade students.

plus indirect), raising reading scores by 1.9 standard deviations and mathematics scores by 1.8. Adding a library led to smaller increases, 0.3 standard deviations for reading and 1.2 for mathematics scores.

A study by Glewwe and others (1995) used Jamaican data collected in 1990 to examine the performance of primary school students in reading (English) and mathematics. More than 40 school and teacher characteristics were examined, including pedagogical processes and management structure. Most variables had statistically insignificant effects. The school variables with significantly positive impacts were administration of eye examinations (reading only), teacher training within the past 3 years (mathematics), routine academic testing of students (reading and mathematics), and the use of textbooks in class (reading). Class time devoted to written assignments had a significantly negative impact in both subjects. The size of these estimated impacts (in standard deviations of the test score variable) were lower than those for Ghana. The largest impact is a change from never using textbooks in instruction to using them in almost every lesson, which raises reading scores by 1.6 standard deviations. The smallest is from teacher training: a school in which all teachers were trained is estimated to have mathematics scores 0.7 standard deviations higher than an otherwise identical school with untrained teachers.

Kingdon's (1996) study of India is based on data collected in 1991. Tests in reading (Hindi and English) and mathematics were given to students in "class 8" (grade 8). Kingdon examined five teacher variables (years of general education, years of teacher training, marks received on official teacher exams, years of teaching experience, and salary) and three school variables (class size, hours per week of academic instruction, and an index of 17 physical characteristics). The teacher variables with significant effects were teacher exam marks, which had significantly positive impacts on both mathematics and reading scores, and teachers' years of education, which had a significantly positive impact on reading scores. Two of the three school variables, the physical characteristics index and time in academic instruction, had significantly positive effects on both reading and mathematics scores. Larger class size was not significantly correlated with mathematics scores, and was correlated *positively* and significantly with reading scores. The impact of the teacher's exam marks is not robust to attempts to control for selection into schools (an issue further discussed below). These impacts are not particularly large. An additional year of teacher's education raises reading scores by 0.13 standard deviations. Going from zero to all 17 physical facilities (which would be quite costly since this includes toilets, computers, and musical instruments) increases mathematics scores by 0.7 standard deviations and reading scores by 1.0 standard deviations. Adding another hour per week of instructional time raises mathematics and reading scores by only 0.04 and 0.02 standard deviations, respectively.

Tan, Lane, and Coustère (1997), using data from 1990 and 1991, investigate the impact of school and teacher variables on the mathematics and reading scores of 2,293 first graders in the Philippines. The five teacher variables examined were academic qualifications (master's degree or not), abstract reasoning ability, scores on subject-based tests, years of teaching experience, and the teacher's attitude toward "innovation in learning." The eight school variables included whether the classroom had sufficient furniture (as judged by the teacher), the pupil-teacher ratio, the value of pedagogical materials received from a government program (PRODED), the availability of textbooks and workbooks per pupil, and four variables on the attitudes and practices of the school

head. Of the teacher variables, the score on the subject knowledge test in reading had a positive impact on students' reading scores: a one standard deviation increase in the teacher's score raised student learning by 0.12 standard deviations. The same is true for mathematics scores: a one standard deviation increase in the teacher's score raised student learning by 0.10 standard deviations. Turning to school characteristics, the impact of textbooks was unstable for both subjects, in some cases significantly negative. More plausibly, the workbook-pupil ratio had significantly positive coefficients for both subjects, so that providing a workbook for each student in schools that have none increases math and reading scores by 0.22 and 0.21 standard deviations, respectively. The only other school variable significant at the 5 percent level was the lack of adequate furniture, which was associated with a drop of –0.32 standard deviations in math and –0.29 standard deviations in reading.

In all four studies, most school and teacher variables were not significantly different from zero, although this could reflect both low sample sizes (163 students in Ghana and 355 in Jamaica) and high correlation among many of these variables.[17] While each study did find that one or more teacher variable had statistically significant impacts, they differed widely across the studies. Similarly, three of the four studies find significant impacts of physical inputs (the exception being the Jamaica study), but again the specific inputs vary across the different studies. Part of this variation could reflect differences in the variables available in the data, and part could reflect large socioeconomic differences across countries (*e.g.*, Jamaica has a much higher income than Ghana and India) but, whatever the reason for this variation, the conclusion is that there are no general results regarding which teacher and school variables raise learning in developing countries.

The summary of the results in the previous paragraphs assumes that the estimated impacts of these four retrospective studies are accurate, but the discussion in Section III provides many reasons to worry about biases in such estimates. One potential source of bias is that unobserved components of a child's innate ability and motivation, as well as parents' motivation, may be positively correlated with school quality because high-ability children tend to enroll in higher quality schools (see Glewwe, 2002, for a simple behavioral model that demonstrates this point). This leads to upwardly biased estimates of the impact of school quality variables. The Ghana and India studies used data from an "intelligence" test, the Raven's Coloured Progressive Matrices test, to control for innate ability. The Ghana study concedes that this test not only measures innate ability (however defined) but also reflects environmental influences, including time in school. It used a simple "family fixed effects" procedure to extract what is probably a cleaner estimate of innate ability from the Raven's test, but this method relies on several simplistic assumptions. The India study used the Raven's test score directly, without any refinement, and the Jamaica and Philippines studies had no variables to control for child innate ability. One of the four studies, the one on India, attempted to control for child motivation as a factor distinct from innate ability. Regarding parents' motivation and ability to help their children, none of these studies goes beyond the common practice of using mother's and father's years of education. Three of the four studies (the exception being the Philippines study) use standard selectivity correction methods (primarily to

[17] Although the sample size in the India study was larger, with 902 students, students are concentrated in 30 schools, which limits variation in school characteristics. The Philippines study has by far the largest sample, with 2293 students in 110 schools.

account for choices among different types of schools). Although this may reduce bias caused by a variety of unobserved variables, including innate ability, these methods may be sensitive to functional form assumptions. They may also yield misleading results if factors assumed to influence choice of school such as distance from school interact with factors that can affect learning such as child ability or household income.

Another potential problem is bias due to omitted school and teacher quality variables. If unobserved school and teacher variables are positively correlated with observed school and teacher variables, the estimated impacts on the observed variables will tend to be biased upward. At first glance, all four studies seem to minimize this problem by including large numbers of school and teacher variables. The original Ghana study used 18 school variables (Glewwe and Jacoby, 1992), and the Jamaica study had 42, including variables on pedagogical techniques and "school organization, climate and control." The India study used data on about 24 variables, although 17 were aggregated into a single index, and the Philippines study used 13 variables. Yet some variables, such as teacher motivation, are inherently difficult to measure and thus are not used in any of these studies (unless the variable on teacher "attitude toward innovation in teaching" used in the Philippine study reflects teacher motivation). Thus, the large number of school variables used does not necessarily avoid bias due to omitted school and teacher characteristics.

A third potential problem is sample selection bias. In many developing countries, some children never attend school, grade repetition is common, and a substantial fraction of children drop out of school after only a few years. As explained in Section III, if weak students are less likely to drop out of high quality schools, the impact of school quality could be underestimated (unless student ability is adequately measured). Biases can also arise due to the choices parents make regarding the schools their children attend and actions parents may take to change those schools, since this may also cause child and household variables to be correlated with unobserved components of school quality. Each of these studies attempted to address at least some of these problems**.** Although the sample size in the India study was larger, with 902 students, students are concentrated in 30 schools, which limits variation in school characteristics. The Philippines study has by far the largest sample: 2,293 students in 110 schools. The India study appeals to the Ghana study for evidence that selection of students (in terms of "survival" to higher grades) does not matter. It does address selection into public and private schools but without explaining how the selection term is identified. The efforts to deal with selection bias are better in the Ghana and Jamaica studies. Both explain the identification strategy (the identifying variables are characteristics of the school not chosen), and the Ghana study accounts for sample selection effects due to delayed enrollment and dropping out (using a similar identification strategy). In both cases, controlling for sample selection has little impact on the results, but this is not the case in the India study. While bias due to school selection is small in two of the three studies, results from more countries are needed before concluding that this problem is not serious.

A fourth potentially serious problem is measurement error in the regressors. Only one of the four studies, the Philippines study, addresses this issue; the other three do not mention it. The Philippines study found evidence of measurement error in the textbook and workbook variables and used, as instruments, textbooks and workbooks in other subjects. A potential downside of this approach is that books in one subject could affect

the scores in other subjects, violating the exclusion restrictions. Measurement error in other school and teacher variables is assumed to be unimportant. Yet a plausible case can be made that measurement error is a serious problem: most such errors are probably random, so that the true effects are likely underestimated. This may explain why in each study most of the teacher and school variables were insignificant.

A final potential problem is that school and teacher characteristics could be correlated with the error term in estimates of equations (1), (4), or (8) if governments build schools or allocate resources to schools based on unobserved community characteristics. This is the problem of endogenous program placement, discussed in Section III. None of these four retrospective studies explicitly addresses this issue, although, arguably, the selection correction methods for school choice decisions may reduce such biases.

This review of conventional studies leads to several conclusions. First, many studies suffer from multiple estimation problems. Second, recent studies have made some progress, but many problems remain. In particular, they use more sophisticated econometric methods, or show an awareness of many of the potential estimation problems, but they have not overcome all of them. Third, three problems are difficult to resolve in conventional studies that attempt to estimate the impact of school characteristics on student achievement: omitted school characteristics, unobserved characteristics of children and their households, and measurement error in school variables. Regarding the first problem, although the Ghana, Jamaica, India, and Philippines studies included large numbers of school characteristic variables in their regressions, other important, but hard to observe, characteristics such as teacher motivation may be highly correlated with the variables that are observed. This will lead to biased estimates. Some results seem counterintuitive; for example, the most important single school characteristic in the Ghana study was leaking roofs. Perhaps the underlying relationship is that more motivated teachers, principals, and parents were more likely to keep the building in good repair. The inability to observe certain child and household characteristics such as the child's innate ability and parental tastes for education also leaves lingering doubts. Finally, it is likely that school variables are measured with a large amount of error—examples have been presented for Tanzania (distance to schools) and the Philippines (books per pupil). Random measurement error could explain why these variables are often statistically insignificant.

In the past few years, two new approaches have been used to investigate how school characteristics affect student achievement, natural experiments and randomized evaluations.

B. Natural Experiments

In this subsection we examine studies that use "natural" variation in a school characteristic that is plausibly uncorrelated with other determinants of child learning to assess the impact of school quality on performance. [18]

[18] See Rosenzweig and Wolpin (2000) for a thorough discussion of "natural" natural experiments, *i.e.* natural experiments whose parameters of interest are identified by date of birth, twin births, gender of newborn child or siblings, and weather. The issues raised in that paper also apply to "less natural" experiments, and many are discussed below.

Before asking which characteristics of schools affect learning, the first question is "Do schools affect learning at all?" Gould, Lavy, and Paserman (2003) shed light on this question using data on Ethiopian Jews brought to Israel on an overnight airlift ("Operation Solomon") in 1991. Gould and his coauthors argue that sorting the refugees into absorption centers and initial schooling environment was random and can be considered exogenous to both family background and parental decisions. According to the authors, this creates a natural experiment that can be used to study the impact of primary school environments on secondary school outcomes. They find that attending an elementary school with a good mathematics program (as measured by grade 4 and 5 standardized test scores prior to the arrival of the Ethiopian emigrants) reduced students' probability of dropping out of high school from 10 percent to 4.9 percent and increased passing rates on high school matriculation exams by 26 percent. The authors note that attending elementary schools with good verbal programs (also measured by grades 4 and 5 standardized test scores) did not improve most high school outcomes. They conjecture that this was because the Ethiopian students were learning Hebrew in separate classes with inexperienced teachers. It is important to note that although the authors control for observed community characteristics, it is difficult to isolate the effect of the quality of elementary school from other potentially unobservable characteristics of the students and parents in the community.

Since teachers account for the bulk of school spending, understanding the impact of class size on learning is critical. The Case and Deaton (1999) analysis of South Africa, discussed above, also examined test scores. They found that decreasing the student-teacher ratio from 40 to 20 raises students' reading test scores (conditional on years of school attendance) by an amount equivalent to the impact of two additional years of schooling. In contrast, there was no significant impact on mathematics scores. However, in interpreting the results of this study one must keep in mind the caveats discussed above in subsection IV.B.

A recent study of class size based on a natural experiment is that of Angrist and Lavy (1999), who examine the impact of class size on student academic performance in Israel. A rule proposed by Moses Maimonides, a twelfth century Talmudic scholar, stipulates that class size should not exceed 40 students, and a form of this rule is used in Israel today. The limits on class size determined by this rule lead to actual class sizes that vary non-monotonically with total enrollment in a given grade, providing an unusually credible instrumental variable to get around the problem that class size may be correlated with unobserved determinates of student learning. The authors use data from the early 1990s on a national test for Israeli third, fourth, and fifth graders. Most of the data, and the analysis, are at the classroom level. For each grade, the sample is approximately 2,000 classrooms from about 1,000 schools.

The only explanatory variables used by Angrist and Lavy are class size, the percentage of disadvantaged students in the school (averaged over all grades), and total enrollment for the grade. Maimonides' rule generates a zigzag relationship between class size and total school enrollment. In grades with an enrollment of 40 or less, class size will equal total enrollment. When total enrollment hits 41, the class must be split into two, so that class size falls abruptly—class size is half of total enrollment for grades with 41 to 80 students. When total enrollment hits 81 a third teacher must be hired, and class size falls

again. This zigzag relationship between total enrollment and class size allows the authors to create an instrument for class size that is not highly correlated with total enrollment, so they can include total enrollment and its square as additional regressors.[19]

A potential problem with the estimation strategy is that some parents may know how Maimonides' rule is applied, and those with high tastes for child education may transfer their children out of schools where that rule leads to large classes. This could cause correlation between unobserved parental tastes for child education and the instrumental variable used to predict class size. The authors argue that this bias should be small since most Israeli parents would not want to transfer their child into another school or switch the child from a secular to a religious school to take advantage of smaller class sizes. Angrist and Lavy find a significantly negative impact of class size on the reading and mathematics scores of fifth graders. The estimated effects of a one standard deviation decrease in class size (a reduction of 6.5 pupils) are increases in reading scores of 0.2 to 0.5 standard deviations and in mathematics, scores of 0.1 to 0.3 standard deviations (the range reflects differences in the sample and in the other covariates). The effects on fourth graders are less precisely estimated. Sometimes they are significantly negative for reading scores, but for mathematics scores the effects are all insignificant. For third graders, all estimated impacts are insignificant; the authors suggest that this may reflect difficulty in measuring a presumably cumulative effect at lower grades. They also point out that testing conditions for the third graders were different from those for fourth and fifth graders.

A final recent paper on education that could be interpreted as a natural experiment on school inputs is by the same authors, also on Israel, Angrist and Lavy (2002). The authors investigate whether providing computers for pedagogical use in classrooms, *computer-aided instruction* (CAI), increased learning in reading (Hebrew) and mathematics among fourth and eighth graders in Israel. The data are from about 200 schools in 1996, 2 years after the introduction of a program that gave Israeli schools funds to purchase computers. The full data include 4,779 fourth graders for math (but only 3,689 for Hebrew) and about 3,200 eighth graders for both subjects.

The authors first present OLS and 2SLS results that are not based on any natural experiment. The OLS results are mostly insignificant, although for some specifications the use of computers appears to have a negative impact on grade 8 math scores. The 2SLS estimates use funding from the Tomorrow-98 program, begun in 1994, under which two thirds of the schools received funding for computers. This instrument had explanatory power for use of computers only for fourth grade students, so 2SLS estimates could not be done for grade 8. The results for fourth grade students show small reductions in the reading and math scores from the use of computers, but only the math score effects are statistically significant (and only for some specifications).

The use of funding from the Tomorrow-98 program as an instrument may be problematic, because local communities (towns) had to apply to the program to receive funding, and those that did apply were required to submit a priority ranking for the schools in their community. Thus, if schools that performed poorly were given higher

[19] Some estimates are restricted to students in schools whose total enrollments are either slightly above or slightly below these "break points." This smaller sample is known as the discontinuity sample, and the results are similar to those for the full sample.

priority, and other regressors in the second stage estimates do not account for all of this poor performance, this instrument will be correlated with the error term in the second stage equation. The authors present evidence that this is not the case, but they also devise an estimation method that uses information on how the ranking affected the probability of receiving program funding. The assumption behind this method is that the ranking variable can be used in two ways. It is assumed that any correlation between the ranking variable and the error term in the second regression is adequately controlled for by including a quadratic specification of that variable in the second stage equation, while the relation between the ranking variable and receipt of funds from the Tomorrow-98 program is sufficiently nonlinear and irregular that the prediction of receipt of funding based on the ranking variable is not completely collinear with the quadratic specification of that variable (and thus can serve as an instrument for use of computers). The validity of this method is debatable since it essentially achieves identification using functional form assumptions. The paper shows that the same results are found with this estimator; there is no evidence that computers improve learning, and in one specification they appear to reduce learning.

Angrist and Lavy (1997) study the effect of changing the language of instruction on test scores and returns to schooling in Morocco.[20] To reaffirm independence from colonial rule and promote nationalism, the language of instruction of Moroccan sixth graders was switched from French to Arabic in 1983. The authors use the sharp change in the language of instruction as a natural experiment to identify the relation between language skills and earnings. They find that the Arabization program reduced returns to secondary education by 20 percent and that the main mechanism was a significant decline in French writing skills. However, the results of this study should be interpreted with caution since the evaluation was done soon after the change, so the results may partially reflect a temporary process of adjustment as workplaces were caught with older cohorts educated in one system and younger cohorts educated in another. Moreover, teachers may have had trouble adjusting to the change in language.

C. Randomized Evaluations

Jamison and others (1981) conducted a randomized trial in Nicaragua in which 48 first-grade classrooms received radio mathematics instruction, 20 received mathematics workbooks, and 20 served as a comparison group. After 1 year, on mathematics tests the radio students scored more than one standard deviation higher, and the workbook students about a third of a standard deviation higher, than students in the control group. Both differences were highly statistically significant.

Three of the Kenya studies discussed in Section IV also examine student academic achievement. As noted in Section IV, a package of assistance including uniforms, textbooks, and school construction led to a tremendous increase in class size as students were attracted from neighboring schools and dropout rates fell. There is no evidence that

[20] A recent paper on Tanzania and Kenya by Miguel (2003) provides one example of how public schools affect the cohesiveness of a nation. Miguel finds that nation-building policies (including the adoption of Swahili as a national language) have allowed ethnically diverse areas in Tanzania to achieve considerably better local public good outcomes, including primary school funding, than their counterparts in Kenya.

the combination package of increased class size and more nonteacher inputs led to a change in test scores. These data are consistent with several hypotheses. One hypothesis is that textbooks have a strong positive impact on learning, but this was offset because of the increase in class size. Another is that neither textbooks nor class size had much impact on test scores. However, as seen below, provision of textbooks in the same area of Kenya had very little effect on test scores, suggesting that the change in class size brought about by this program also had little effect.

Glewwe, Kremer, and Sylvie Moulin (2003) find no evidence that provision of official Kenyan government textbooks increased scores for the typical student. However, they do find evidence that textbooks led to higher test scores for the subset of students who scored well on a pretest. The authors note that English, the medium of instruction in Kenyan schools and the language in which textbooks were written, was the third language for most pupils, and cite evidence that many pupils had difficulty reading the books. As discussed further below, there is reason to think that the Kenyan curriculum is not appropriate for the typical student in rural areas.

The third Kenya study discussed in Section IV, Miguel and Kremer (2004), examined the impact of deworming medicine not only on the quantity of schooling but also on test scores. A priori, the impact on learning may be small because this intervention raised attendance rates by about 5 percentage points for 2 years, which implies attending school only 20 additional days over 2 years. Moreover, the impact on learning per day in school may also be small because very few cases of severe infection were reported. Indeed, the authors find that the deworming treatment had no effect at all on students' test scores after 2 years.

A third Kenyan study, not discussed above, is Glewwe and others (2004). It examined flip charts: large poster-sized charts with instructional material that can be mounted on walls or placed on easels. This intervention, which was not examined in Section IV because it did not evaluate the impact of flip charts on any indicators of the quantity of schooling, covered 178 primary schools, half of them randomly selected to receive flip charts covering science, mathematics, geography, and health. Despite a large sample size and 2 years of follow-up data, the estimated impact of flip charts on student test scores is very close to zero and completely insignificant. In contrast, several conventional OLS estimates, which may suffer from many of the problems described in subsection III.B, show impacts as large as 0.2 standard deviations, 5 to 10 times larger than the estimates based on randomized trials.

A remedial education program in urban India, focused on improving the learning environment in public schools, appears to have increased test scores at a low cost. Banerjee and others (2004) conducted a randomized evaluation of a 2-year remedial education program in Mumbai and Vadodara, India. The remedial education program is run by a collaboration between a local NGO and the Indian government, and hires (at a yearly cost of only US$5 per child) young women from the community to teach basic literacy and numeracy skills to children who reach grade 3 or 4 without mastery of some basic competencies. On average, the program increased test scores by 0.14 standard deviations in the first year and 0.28 in the second year. The gains were largest for children at the bottom of the distribution, which is unusual for educational programs. Results were similar in both grade levels and in two different cities. The authors note that this program would be several times more cost-effective than hiring new teachers. The

success of this program suggests that students were being poorly served by the existing education system.

Finally, Banerjee and others (2004) recently conducted a randomized evaluation of a computer-assisted learning program in India and found much more positive results than those from the computer-assisted learning program in Israel (Angrist and Lavy, 2002). The idea of using computers in schools seems particularly promising in areas where both the number of qualified teachers and the quality of employed teachers is notoriously poor. The Indian CAL program took advantage of a donation by the state government of four computers to each municipal primary school in Vadodara and gave each child in the fourth standard 2 hours of shared computer time to play educational games that reinforced mathematical concepts (ranging from standard 1 to standard 3 levels). The program was found to be quite effective, with average mathematics score increases of 0.36 standard deviations in the first year and 0.51 standard deviations in the second year. The program was equally effective across student ability levels.

D. Summary

Policymakers are keen to know the likely impacts on student academic achievement of various policy interventions, but retrospective studies offer only limited guidance. Even the best retrospective studies suffer from serious estimation problems, the most serious being omitted variable bias with respect to school and teacher characteristics, unobserved child and household characteristics that are correlated with observed school and teacher variables, and measurement error in school and teacher data. This has turned attention in recent years to many studies based on natural experiments and randomized trials.

Evidence from recent natural experiments in middle-income countries suggests that increases in school resources (as measured by the student-teacher ratio) raise academic achievement on reading tests (but not math tests) among black students in South Africa. Studies using Israeli data indicate that reducing class size raises reading scores and (less often) math scores and that providing computers has no effect on academic performance.

Finally recent randomized trials offer evidence from some relatively poor developing countries. In Nicaragua, workbooks and radio instruction had significant impacts on pupils' math scores, and the impact of radio education was particularly high. (Ironically, radio education was never implemented in Nicaragua after this study demonstrated its effectiveness.) Provision of textbooks raised performance on academic tests in the Philippines, but in Kenya the only effect of textbooks was among the better students (most likely because the textbooks were too difficult for many students). Evidence from Kenya also suggests little impact on test scores of reductions in class size, flip charts and deworming medicine, although school meals were found to have positive impacts on test scores as long as teachers were well trained. A remedial education program in urban India, focused on improving the learning environment in public schools, appears to have increased test scores at a low cost. Finally, a computer-assisted learning program in India suggests that such programs have potential in developing countries. The findings on radio education in Nicaragua and computer instruction in India suggest that technologies that help substitute for weak teachers may be particularly helpful.

While these natural experiments and randomized trials are beginning to build a database of results that are less likely to suffer from the estimation problems that plague retrospective studies, a much larger set of results is needed before general conclusions can be drawn for policymakers. However, one interpretation of these results is that in many developing countries, the most effective means of improving school quality may be through addressing the problem of weak teaching. The remedial education program in urban India, the radio mathematics program in Nicaragua, and the computer instruction program in India all provided inputs which addressed the problem of weak teaching, whereas programs which provided inputs that were dependent on use by the teachers themselves (such as the flipcharts and, to some extent, the textbook program in Kenya) were less effective.

Below we discuss the problem of incentives and education systems more broadly.

VI. Education Systems, the Political Economy of Education, and Reform Initiatives

The studies reviewed in Sections IV and V considered education policies that consisted primarily of direct changes in educational inputs available in the classroom such as textbooks, blackboards and other physical supplies, new schools and repairs of existing ones, and more and better trained teachers. In terms of the analytical framework presented in Section III, these policies directly change the characteristics of schools (**Q**), the prices of educational inputs (**P**), or both. Thus the studies discussed in Section IV were attempts to estimate equation (2) and the studies in Section V were attempts to estimate equations (1), (4), or both. Yet many education policies do not directly attempt to change the classroom environment but instead change the fundamental institutional arrangements in the education system such as incentives for teachers and financing arrangements. These changes should affect what happens in the classroom and, through this, learning. The impacts of such policies on the quantity and the quality of schooling are depicted in equations (7) and (8), Section III.

A. Institutional Issues and Problems

Education systems in developing countries face many challenges. In some cases, resources intended for education are diverted for other purposes. Teachers may be paid but nonetheless absent from their classrooms, and while funds may be budgeted for inputs such as textbooks those textbooks may never reach the students. Second, financing distortions may imply the funds spent on education are often allocated inefficiently. For example, spending on salaries relative to nonsalary inputs is inefficiently high, and many local communities are not in control of their own budget and thus cannot reallocate resources to fit local needs. Third, the curriculum used in many schools is inappropriate for the typical child due to an elite orientation of many curricula.

In examining the education finance system, it cannot be assumed that resources are being used, and personnel deployed, in accordance with official budgets. Reinikka and Svensson (2004) examine a program launched in Uganda in1991 that provided a per-student grant to cover schools' nonwage expenditures, using district education offices as distribution channels. The 250 schools surveyed over a 5-year period (1991–95) received

on average only 13 percent of the grants, based on the authors' comparison of the disbursed flows from the central government and the schools' records of resources received. It is not clear whether the funds were stolen, used for other purposes within or outside the education system, or simply not disbursed. Reinikka and Svensson argue that they were probably used to finance the local political machinery. The extent to which such diversion of education funds occurs in other developing countries is unknown. The program was new at the time of the original study, and Reinikka and Svensson find considerable improvement after the introduction of improved budgetary procedures and steps to publicize the program to local schools.

Most educational spending is on teacher salaries. Teachers are usually in a strong position to ensure that these funds reach them, but this does not necessarily mean that teachers are in the classroom. A recent study by Chaudhury and others (forthcoming) reports survey results in which enumerators made surprise visits to primary schools in Bangladesh, Ecuador, India, Indonesia, Peru, and Uganda and recorded whether teachers were present (table 12). Averaging across the countries, about 19 percent of teachers were absent. The authors found that absence is not typically concentrated among a small number of frequently absent providers, but seems rather to be fairly widespread.

Absence rates across Indian states varied from 15 percent in Maharashtra to 42 percent in Jharkhand. Both cross-country (Chaudhury *et al.* forthcoming) and cross-state within India (Kremer *et al.* forthcoming) analyses suggest absence rates are generally higher in poorer regions: doubling national- or state-level per-capita income (PPP-adjusted) is associated with absence rates that are 5.8 percentage points lower. Proxies for salary levels and intensity of community monitoring are not robust predictors of absence. Higher-ranking and more powerful providers, such as headmasters and doctors, are absent more often than lower-ranking ones. The relationship between absence and contractual terms for teachers seems more complicated than often hypothesized. Community managed schools and schools managed by the central ministry have similar absence rates. Contract teachers' absence rates are typically similar to those of regular civil servants, and sometimes considerably higher. In India, where the authors examined absence rates in private schools, the study found that they were similar to those in public schools, but considerably lower than those of public schools in the same village. However, private school teachers and contract teachers are often paid much less than civil servants, which will enter into any judgment about the cost effectiveness of these teachers.

While high absence rates in some developing countries may reflect a variety of factors, including the prevalence of infectious diseases such as malaria and AIDS, these unavoidable absences are unlikely to account for all absences. Glewwe, Ilias, and Kremer (2004) found that in a region of Kenya with 20 percent teacher absenteeism, staff at a nonprofit organization working in the same area had absence rates of only 6.3 percent.

A second basic institutional problem is that even when the allocated funds are spent on education, they may be used inefficiently. Pritchett and Filmer (1999) argue that policymakers do not choose inputs solely to maximize the production of educational outputs but also try to provide rents to teachers. The authors argue that several studies have found that the marginal product per dollar of inputs not directly valued by teachers (such as textbooks and infrastructure) are 10 to 100 times higher than that of inputs valued by teachers such as salaries and class size (World Bank, 1996; Harbison and

Hanushek, 1992). They conduct a meta-analysis of educational studies (taken from Fuller and Clark, 1994), showing that nonteacher inputs have a much higher probability of being statistically significant and of the expected sign than inputs they argue are more likely to appear directly in teacher's utility functions. Though suggestive, the underlying studies may suffer from many of the biases discussed in Section III. One could easily imagine that coefficients on individual non-teacher inputs are picking up a much broader set of omitted inputs. Another caveat is that, in the absence of direct evidence, it is not clear that teachers care more about class size than infrastructure or textbooks—they may also like having electricity and school buildings that do not leak.

A third issue is that in many developing countries, educational systems are oriented towards elites. As we discussed in Section II, per-pupil expenditures in most developing countries are much higher for tertiary (post-secondary) students than for primary and secondary students. Another manifestation of this elite orientation is that in many developing countries there is a mismatch between the curriculum and the typical student. Many educational systems in developing countries are highly centralized (certainly compared to the educational system in the US), and to the extent that policymakers are often members of elite groups, it is not surprising that the chosen curricula are often much more suitable for advanced students than for the typical student. For example, Glewwe *et al.* (2002) provide evidence from Kenya that increasing availability of official textbooks raised test scores for the top two quintiles of students (as measured by initial academic achievement) but had no effect on either test scores or dropout and repetition rates of average and below average students. Indeed, the authors found that the typical median student in grades 3, 4, and 5 could not read the English textbooks designed for those grades. When curricula are set too far beyond the level of the average student, too many students fall behind, and lose the ability to follow. The results of the evaluation of a remedial education program in India (as discussed in Section V) suggest that the school system is not taking advantage of opportunities to serve students at the bottom of the distribution there, either (Banerjee *et al*. 2004). The remedial education program was found to have substantial positive impacts on test scores – gains which were largest for children at the bottom of the distribution. There is likely much more heterogeneity in a variety of factors – including student school attendance, teacher absence, educational backgrounds, *etc.* – among students in less developed countries than there is among students in developed countries. This heterogeneity implies it is difficult to devise a single curriculum suitable for the entire population.

Recognition of the institutional weaknesses of education systems in developing countries has led both policymakers and researchers to shift their focus to policy reforms that attempt to reduce distortions and inefficiencies in the institutional arrangements of education systems. Reform initiatives range from policies that preserve current education governance structures but seek to strengthen links between teachers' pay and students' performance, to decentralizing budget authority so that local communities have more power to manage their resources, to introducing vouchers and other methods to increase school choice. Now, we turn to empirical evidence on each of these reforms.

B. Teacher Incentives

According to advocates of incentive pay for teachers, teachers in many developed and developing countries face weak incentives, with pay determined almost entirely by educational attainment, training, and experience instead of by performance. In some developing countries, incentives are extremely weak, with no effective sanctions for behavior that would invite disciplinary action in developed countries. Some observers see linking teachers' pay to students' performance as a way to increase teacher effort.

In developed countries, opponents of teacher incentives based on students' test scores argue that, since teachers' tasks are multidimensional and only some aspects are measured by test scores, linking compensation to test scores could cause teaches to sacrifice promoting curiosity and creative thinking in order to teach the skills tested on standardized exams (Holmstrom and Milgrom, 1991; Hannaway, 1992). Another concern is that linking pay to individual teachers' performance could undermine cooperation among teachers. Education experts are therefore generally less sympathetic to individual-based incentives than to school-based incentives, which they feel are more conducive to cooperation among teachers.

The extremely weak teacher-supervision systems in many developing countries raises both the potential benefits and costs of teacher incentives. On one hand, it could be argued that teachers in many developing countries are already teaching to the test, that the first order problem is to get teachers to show up to work, and hence that teacher incentives are particularly appropriate for developing countries. On the other hand, developing countries with weak systems of teacher accountability may be more prone to attempts by teachers to game incentive systems. In particular, teachers could try to force weak students to drop out so as to avoid bringing down average scores, or they could make it difficult for weak students to enroll in the first place. Empirical evidence on the effectiveness of monetary teacher incentives is scarce, particularly in developing countries. Nevertheless, two recent studies from Israel and Kenya provide some initial, and intriguing, evidence. The first study, by Lavy (2002), evaluates a program in Israel that offered teachers monetary incentives based on their students' achievements in three dimensions: the average number of credits per student, the proportion of students receiving a matriculation certificate (required for college admission), and the school dropout rate. Awards were given at the school level, so that all teachers in a school shared the same award. The program was implemented in 62 nonrandomly selected secondary schools starting in 1995. The incentives took the form of awards on a rank order tournament: only the top three schools, ranked by relative improvement, received a prize.

Lavy's identification strategy is based on the program's selection criteria, which limited participation to schools that were the only school of their kind in a community (religious girls' and boys' Jewish schools, secular Jewish schools, and Arab schools). He compares the results of program schools with control group schools where there are more than one kind of school in the same community. Using a fixed effects estimation procedure, Lavy finds that, after 2 years, the program had a positive and significant effect on two of the three student outcomes evaluated: average credits were 0.7 units higher and the proportion of students sitting for the matriculation exam increased by 2.1 percent. He then interacts the treatment dummy with mother's education and finds that the program mainly affected weaker students.

The findings from Israel are consistent both with the conjecture that incentive pay affects teacher effort and the claim that incentive pay causes teachers to teach more

strictly to the test. To distinguish between these two hypotheses not only the effect of the program on test scores must be considered, but also the channels through which this effect occurs. Glewwe, Ilias and Kremer (2004) do this, examining a randomized evaluation of the impact of a teacher incentives program in Kenya on both teacher behavior and test scores. They consider a model in which teachers can invest both in efforts to promote long-run learning and in short-run manipulation of test scores. Data were collected on many types of teacher effort—attendance, homework assignment, pedagogical techniques, and holding extra exam-preparation sessions—and on scores after the end of the program.

The teacher incentive program in Kenya offered teachers prizes based on their schools' average scores on district-wide exams. The program penalized teachers for dropouts by assigning low scores to students who did not take the exam. During the two years the program was in place, student scores increased significantly in treatment schools (0.14 standard deviations above the control group). However, analysis of the Kenyan data suggests that this improvement did not necessarily occur through the channels intended. Teacher attendance and student dropout and repetition rates did not improve, and no changes were found in either homework assignment or pedagogy. Instead, teachers were more likely to conduct test-preparation sessions outside of normal class hours. Data from the year after the program ended show no lasting test score gains, which suggests that the teachers' effort was concentrated in improving short-run outcomes, rather than stimulating long-run learning. The test-score effect was strongest for subject tests on geography, history, and Christian religion, arguably the subjects involving the most memorization. Also consistent with this hypothesis, the program had no impact on dropout rates, but exam participation rose (presumably because teachers wanted to avoid penalties for no-shows at exams).

C. Decentralization and Local Community Participation

In response to the failures of centralized school systems, many observers advocate decentralization and community participation (World Bank, 2004). Local communities arguably have the best knowledge about the needs of their children, strong incentives to monitor the performance of teachers and headmasters, and a comparative advantage in conducting this monitoring. Decentralization reforms are increasingly being adopted. At this point, however, rigorous empirical evidence on their impact is scarce.

The EDUCO program in El Salvador is often cited as an example of the benefits of decentralization. Under the program, school committees are responsible for contracting and removing teachers and closely monitoring their performance and for equipping and maintaining the schools. All of their resources come from the central government and international organizations. An evaluation by Jimenez and Sawada (1999) finds that the program successfully expanded education in poor rural areas (its main objective) and also reduced student absences by 3 to 4 days in a 4-week period. No effect was found on student achievement. However, credibly identifying the impact of the EDUCO program is very difficult because the program was not implemented in any randomized way. The authors use standard selection correction techniques, and the selection correction term is identified primarily by functional form assumptions. The only variables in the selection equation excluded from the equations of interest are district dummies variables, and the

theoretical justification for this exclusion restriction is unclear. Overall, the results are intriguing and intuitively plausible, but more research is needed before making policy recommendations.

Reinnika and Svensson (2003) examine the effect of local community empowerment through an information campaign on delivery of nonwage funds from the central government to schools in Uganda. Using a survey similar to Reinikka and Svensson (2002), the authors calculate that the percentage of the funds from the central government that actually reached the schools increased from 20 percent in 1995 to 80 percent in 2001. The authors argue that the improvement was mainly the result of better monitoring of local officials' handling of resources by the schools, stimulated by an information campaign launched by the government after the results of the 1991–95 survey came out. Under the campaign, data on monthly capitation grant transfers to districts was published in major newspapers and broadcast on the radio. Exploiting differential access to newspapers across schools, the authors argue that schools with access to newspapers increased their funding on average by 12 percentage points more than schools with no access to newspapers, despite the fact that the two groups had similar funding levels in 1995. Monitoring from the center of the districts was also strengthened. While the fixed effects control for time-independent determinants of funding, this identification strategy cannot rule out the possibility that other features of schools, correlated with newspaper access, could have had an effect on funding in the later period but not earlier. For example, economic development was uneven across Uganda during this period and could have been correlated with newspaper access. It is unclear what caused the large reported increase in funds reaching schools: the authors argue this information campaign was successful through a "bottom-up" approach, but it is unclear whether this or more of a "top down" approach deserves more emphasis, since Uganda had an authoritarian leadership that was strongly committed to reform, and since international donors also played an important role in Uganda at the time in promoting both the original grant program and the survey designed to determine whether funds were reaching schools. The grant program was also relatively new at the time of the original survey, and part of the increase in funds reaching schools may reflect implementation over time.

Overall, more work on the impacts of informational campaigns would be useful. Partly in response to a desire among policymakers to improve the accountability of social services, several ongoing randomized evaluations being led by researchers at the Poverty Action Lab at MIT are studying the impact of information on the quality of education services. In rural Uttar Pradesh, India, Abhijit Banerjee and others are evaluating various strategies designed to empower villages to demand better quality education; one strategy is to provide villagers with information such as the names of local (village) officials responsible for education, the funds available for education, and the number of children in the village who are unable to read. In Sierra Leone, Rachel Glennerster and Edward Miguel are examining how providing communities with information about how many textbooks the communities are meant to receive from the government influences the actual number of government-provided textbooks these communities receive.

Miguel and Gugerty (2004) suggest that the impact of decentralization can vary with the local environment. As mentioned in the previous subsection, in Kenya local school committees must raise funds to build schools and provide nonwage inputs, and

they do so through school fees and local fundraisers called *harambees*. Miguel and Gugerty argue that communities with high ethnic diversity have major difficulties overcoming free-rider problems in collective action, such as imposing and enforcing sanctions, and therefore have less local school funding and lower quality school facilities than homogeneous ones. Using data on 100 rural primary schools, the authors find that moving from complete ethnic homogeneity to average school-level ethnic diversity is associated with a drop in funding of 20 percent of average local funding.

Kremer, Moulin, and Namunyu (2002) examine Kenya's mix of centralized and decentralized control over different aspects of education. They argue that the system creates incentives for misallocation and then test for empirical evidence of misallocation. At independence Kenya adopted an education finance system in which local communities were responsible for raising the resources to build schools, while the central government assigned teachers to schools and paid their salaries once the schools were built. Local communities had to provide nonteacher inputs such as textbooks and chalk, which they typically did by levying school fees. The system blended substantial centralization with elements of local control and school choice. The authors argue that the system suited the interests of the ruling coalition at independence, which drew support from some of the country's more educationally advanced and politically organized regions. The system allocated resources disproportionately to these regions, since they were best placed to build schools. At the same time, it retained central control over teachers, thus avoiding the possibility that local hiring would lead to discrimination against outsiders (which would have hurt the constituents of the ruling coalition.) However, the education finance system created an interlocking set of distortions. Local communities had strong incentives to build new schools, because once they had built one, the central government provided the teachers, which absorbed more than 90 percent of the present discounted cost of operating the school. Thus, many small schools, with small classes, were built close together. In the districts studied in the paper, the median distance between primary schools was 1.5 kilometers, and Kenya's average pupil-teacher ratio of 28 in 1998 is much lower than the average of 43 for Sub-Saharan Africa in 2000 (table 11). The system led to excessive spending on teachers relative to nonteacher inputs. For example, a Ministry of Education survey showed that, on average, 17 primary school pupils in Kenya shared one textbook.

The education system also created incentives for schools to set high fees and other attendance requirements, which kept many children out of school. Typically, increasing enrollment did not bring any more resources from the central government, because a new teacher was assigned only when class size surpassed 55, and most classes were substantially smaller, at least in the upper grades, given the large number of schools that had been constructed. Setting fees that lead marginal students to drop out eases teacher workload and could potentially help increase the school's average score on the national exams, the main criterion used to judge schools and headmasters.

Empirical evidence of distortions in education systems is provided by the evaluation of an NGO program. In 1994, the NGO selected 14 poor schools and divided them randomly into treatment and comparison groups. For the next 5 years, treatment schools were provided with uniforms, textbooks, and new classrooms. The free uniforms represented a substantial reduction in the cost of schooling. The program schools attracted a large influx of pupils from neighboring schools, increasing average class size

by 8.9 students. The combination of larger classes, more nonteacher inputs, and lower schooling costs led to a large expansion in the quantity of education, and no apparent reduction in quality. Students in the seven treatment schools remained enrolled an average of 0.5 years longer and advanced an average of 0.3 grades further than their counterparts in the seven comparison schools. There is no evidence that the combination of larger class sizes and more nonteacher inputs led to different test scores among students originally enrolled in treatment schools than among those originally enrolled in comparison schools. The revealed preferences of the households that transferred their children from other schools into the treatment schools suggest that they were willing to accept an increase in class size of at least 8.9 students in exchange for lower costs and extra nonteacher inputs. The inefficiencies of the current education system are apparent from the fact that the Kenyan government could have financed this package of textbooks, classroom construction, and uniforms using the savings that could be generated from much, much smaller increases in class size than those associated with the program. Overall, this evidence suggests that the details of decentralization are critical. The results do not imply that decentralization is ineffective but suggest that inefficiencies arose in Kenya from a mismatch between decision-making power and financial responsibilities. Local communities had authority to start new schools while covering only a fraction of the cost.

D. Vouchers and School Choice Programs

Perhaps the most fundamental policy reforms are voucher and school choice programs, which provide government funds that students can use to enroll in either public or private schools. A number of studies in the 1980s and 1990s argue that private schools are much more efficient than public ones. However, the econometric difficulties surrounding such comparisons are formidable.[21] Vouchers have been implemented in two Latin American countries, Chile and Colombia, on a much larger scale than voucher programs in the United States.

Angrist and others (2002) examine the effects on education outcomes of Colombia's voucher program, which offered vouchers to attend private secondary schools to more than 125,000 students from poor urban neighborhoods. In most communities, the demand for vouchers exceeded the supply, so voucher eligibility was determined by a lottery, generating a natural experiment. Data were collected from 1,600 applicants for the vouchers (primarily from Bogota) three years after they had started high school. The sample was stratified so that half those sampled were lottery winners and half were lottery losers. Angrist and his coauthors find that lottery winners were between 15 percent and 20 percent more likely to be in private schools, 10 percent more

[21] The studies typically regress children's test scores, and in some cases school expenditures per pupil, on a private-public dummy variable and attempt to correct for selection bias using observed variation in child characteristics and a selection term from prior estimation of the choice between public and private schools. Cox and Jimenez (1991) find that private secondary schools in Colombia and Tanzania have robust advantage over public schools in test scores and unit costs. However, their results may be sensitive to their identification strategy, which relies on the exclusion restriction that family background and the child's ability do not enter into the test score regression. Using a similar method, and therefore subject to the same caveats, Kingdon (1996) finds that private unaided schools in India have a ratio of cost over test score that is only about half of the corresponding ratios for public schools and private-aided schools.

likely to complete eighth grade, and scored 0.2 standard deviations higher on standardized tests, equivalent to a full grade level. A number of channels could account for the impact of the vouchers. First, lottery winners were more likely to have attended participating private schools, and these schools may be better than public schools. Second, vouchers allowed some pupils who would have attended private schools anyway to attend more expensive schools. Finally, because voucher recipients who failed a grade risked losing vouchers, lottery winners had an incentive to devote more effort to school, and the schools they attended had an incentive not to fail them. The authors also find that vouchers affected non-education outcomes: winners spent less time working in the labor market than losers and were less likely to marry or cohabit as teenagers. Analysis of the economic returns to the additional schooling attained by winners after 3 years of participating in the program suggests that the benefits likely greatly exceeded the $24 per winner additional cost to the government of supplying vouchers instead of public school places.

Work by Angrist, Bettinger, and Kremer (2004) suggests that the vouchers not only had significant effects on the short-run outcomes of their recipients, but that their impact persisted over time. Using administrative records on registration and test scores on a centralized college entrance examination, the authors find the lottery program increased secondary school completion rates by 15-20 percent. Correcting for the greater percentage of lottery winners taking college admissions tests, the program increased test scores by two-tenths of a standard deviation in the distribution of potential test scores. Boys, who have lower scores than girls in this population, show larger test score gains, especially in math.

The analysis of school vouchers in Colombia by Angrist and others (2002), discussed above, is based on random assignment, and therefore addresses many omitted variable bias concerns. It is important to note that their estimates capture the overall effect of the voucher program, rather than simply the effect of moving pupils from public to private schools. Because voucher recipients who failed a grade risked losing the vouchers, lottery winners also had increased incentives to devote more effort to school, and the private schools they attended had an incentive not to fail them.

While Angrist and others (2002) examine the effect of vouchers on participants in voucher programs, such programs may also affect children who do not participate and instead stay in public schools. If more advantaged public school students switch to private schools, and if these students generate positive externalities for their public school peers, the students who remain in public schools might be hurt by vouchers. However, competition from private schools might improve public schools, as argued by Hoxby (2000). The overall effect is therefore an empirical question.

Hsieh and Urquiola (2000) address this question by looking at Chile, which in 1981 began a nationwide school voucher program that gave a fixed per student voucher payment to any participating school, public or private. The main features of the program remain in today's school system; the 20-year program has substantially changed the education market in Chile. When the program started, 22 percent of all students were in private schools; by 1990 this number had risen to 41 percent. These numbers hide a wide cross-sectional variation on the change in private enrollment, however, with highly urbanized, educated, and densely populated areas experiencing a much larger expansion of private schooling. Using fixed effects, Hsieh and Urquiola (2002) argue that higher

private enrollment rates negatively affect the relative test scores, repetition rates, and socioeconomic status of students in public schools. They also find that higher private enrollment rates did not affect the average outcomes of municipalities. They interpret these results as evidence that the program merely increased sorting rather than adding value to education. However, their identification is problematic, because their fixed effect estimation does not control for time-varying unobserved characteristics and idiosyncratic shocks to schools and municipalities that may be related to private enrollment trends. In particular, it is plausible that people in areas experiencing negative shocks to public schools turned to private schools in response. This would produce the correlations found by Hsieh and Urquiola.

More research is needed to provide a firmer assessment of the impact of voucher programs on nonparticipants. One way to shed light on this would be through randomization evaluations at the level of local communities, which could allow estimation of the total program effect.

Another strand of the literature examines the political economy of school choice. School systems not only teach students skills but also shape their preferences and ideology. In theory, in a public school system, the median voter determines the ideology taught. Under a voucher system, parents might choose to educate their children in schools with an ideology similar to their own, leading to potentially conflict-generating ideological and cultural segregation (Kremer and Sarychev, 2000) These issues may be particularly important in countries with ethnic diversity.

One reform that may be worth considering is allowing increased choice within the public school system and allocating resources to public schools based on enrollment. This would create some competition among schools for students and is also likely to be more equitable than current school finance systems which often allocate not budgets, but teacher slots, in proportion to the number of pupils. Since more experienced and better qualified teachers are likely to wind up in better-off areas, allocating teacher slots on a per pupil basis provides more funding for students in richer regions.

E. Summary

Many education systems in developing countries are subject to major distortions. The evidence presented in this section suggests that schools in these countries face significant challenges: distortions in educational budgets often lead to inefficient allocation and spending of funds; weak teacher incentives lead to problems such as high rates of teacher absenteeism; and curriculums are often focused excessively on the strongest students and are not well-matched with the typical student, especially considering the high rates of teacher and student absenteeism.

Numerous school reform initiatives have been proposed, ranging from programs designed to strengthen links between teacher pay and performance, to reforms to decentralize budget authority, to voucher and school choice programs. Although the evidence is scarce on teacher incentive programs in developing countries, results from Israel suggest that teacher incentives positively and significantly affected student education outcomes (and mainly for weaker students). Results from Kenya suggest that teacher incentives increased teachers' efforts on short-run outcomes (test scores) but not on stimulating long-run learning (through changes in teacher attendance, student dropout

rates, or pedagogy). Decentralization programs appear promising, but the results of decentralization policies appear to be very heavily dependent on the details of implementation. Finally, a school choice program in Colombia yielded dramatic benefits for participants, but evidence from voucher programs in Chile as well as numerous developed countries suggests that more research is needed to gauge the generalizability of such program impacts.

VII. Conclusions and Directions for Future Research

This section summarizes some of what research has taught us about education in developing countries and then discusses ways that research can help shed light on some of the open questions. In particular it discusses the potential of randomized evaluations to improve knowledge about education in developing countries.

A. Conclusions Regarding the Determinants of Education Outcomes in Developing Countries

Despite rapid progress in expanding school enrollment since 1960, many children are still not in school, the quality of education is often low in developing countries, and many education systems are dysfunctional.

As discussed in Section IV, a number of techniques can be used to expand school participation fairly easily. To what extent investing in school quality attracts children to school, however, is unclear. Programs that reduce the cost of schooling or provide attendance incentives (either implicitly through school meals, or explicitly) have sizable impacts on school participation. Randomized evaluations of school-based health programs suggest that, in some situations, these programs can be an extraordinarily cost-effective means of increasing the quantity of schooling attained in developing countries.

Evidence concerning the impact of additional educational inputs is more mixed. In general, retrospective studies suggest that educational inputs have limited impact on improving the quality of schooling in developing countries. However, since even the best retrospective studies suffer from serious estimation problems, attention has turned in recent years to studies based on natural experiments and randomized trials, both of which paint more mixed pictures of the impact of educational inputs. Evidence from recent natural experiments in middle-income countries suggests that reducing class size can raise academic achievement but that providing computers has little effect. Recent randomized trials conducted in low-income countries provide a more mixed picture.

The evidence suggests that the most effective forms of spending are likely to be those that respond to inefficiencies in schooling systems. Providing textbooks written with atypical students in mind will benefit only atypical students, whereas remedial education may be extremely effective in an environment in which many students fall behind and are no longer able to follow teachers' lessons. Providing radio mathematics education or computer-based education may be effective when teachers attend irregularly.

Schools in developing countries face significant institutional problems: distortions in education budgets often result in inefficient allocation and spending of funds; weak teacher incentives lead to problems such as high rates of teacher absenteeism; and, given

the difficulties faced by these school systems, curriculums are often inappropriately matched with the level of the typical student. Yet reform initiatives can easily have unintended consequences. The details of these programs are critical for their incentive effects. Governance reforms and allowing school choice appear to hold more promise than simply providing monetary incentives to teachers based on test scores, but much more empirical evidence is needed on the impact of these reforms as well.

As noted in the introduction, sometimes a false dichotomy is constructed regarding education initiatives in developing countries. Some observers argue that these schools need more money; others emphasize the weaknesses of the school systems and the need for reform. These two views are not, however, mutually exclusive; in fact, both may be true. In settings with highly distorted education systems, some types of spending will have low marginal product while others will have high marginal product. Hence, carefully targeted spending can be extremely productive in such settings.

B. Methodological Conclusions

We have learned some things about education, but much remains to be learned. This section presents some methodological lessons for future research, drawing on the examples discussed in this chapter.[22]

1. Estimates from randomized evaluations can be quite different from those drawn from retrospective evaluations.

As seen in the studies of textbooks and flip charts in Kenya, estimates from prospective randomized evaluations can often be quite different from estimated effects in a retrospective framework, suggesting that omitted variable bias is a serious concern (Glewwe and others, 2004). Similar disparities between retrospective and prospective randomized estimates arise in studies of the impact of deworming in Kenya (Miguel and Kremer, 2004a) and the impact of social networks on take-up of deworming drugs (Miguel and Kremer, 2004b). This is consistent with the findings of Glazerman, Levy, and Meyers (2002), who assessed both prospective (experimental) and retrospective (nonexperimental) methods in studies of welfare, job training, and employment service programs in the United States, synthesizing the results of 12 design replication studies. They found that retrospective estimators often produce results dramatically different from randomized evaluations, that the estimated bias is often large, and that they were unable to identify any strategy that could consistently remove bias and still answer a well-defined question.[23] We are not aware of any systematic review of similar studies in developing countries. Future research along these lines would be valuable, since

[22] This section draws upon the discussion in Duflo and Kremer (2005).

[23] A recent study by Buddlemeyer and Skoufias (2003) is not included in the analysis of Glazerman, Levy, and Meyers (2002). Buddlemeyer and Skoufias use randomized evaluation results as a benchmark to examine the performance of regression discontinuity design (a type of natural experiment) for evaluating the impact of the PROGRESA program on child health and school attendance. In this case, they found the performance of regression discontinuity design to be good.

comparative studies can be used to assess the size and prevalence of biases in retrospective estimates. However, when the comparison group for the retrospective portions of these comparative studies is selected ex post, the evaluator may be able to pick from a variety of plausible comparison groups, some of which may have results that match experimental estimates and some of which may not. To address these concerns, future researchers should conduct retrospective evaluations before the results of randomized evaluations are released or conduct blind retrospective evaluations without knowledge of the results of randomized evaluations or other retrospective studies.

2. Publication bias appears to be substantial with retrospective studies. Randomized evaluations can help address publication bias problems, particularly if institutions are put in place to compile the study results systematically.

There is a natural tendency for positive results to receive a large amount of publicity: agencies that implement programs seek publicity for their successful projects, and academics are much more interested in publishing and more able to publish positive results than modest or insignificant results. However, many programs are failures, and publication bias will be substantial if positive results are much more likely to be published. In particular, if comparison groups are defined ex post, as in retrospective studies, researchers who obtain negative results using one potential comparison group may simply try other comparison groups instead. There is evidence of strong publication bias (DeLong and Lang, 1992). Instrumental variable estimates may be particularly subject to such bias because such estimates tend to have larger standard errors. Ashenfelter, Harmon, and Oosterbeek (2000) show strong evidence of publication bias of instrumental variables estimates of the returns to education: on average, estimates with larger standard errors also tend to be larger. This accounts for most of the oft-cited result that instrumental estimates of the returns to education are higher than ordinary least squares estimates.

Randomized evaluations are likely less subject to publication bias because they require committing considerable resources in advance to a particular comparison group: once the evaluation is done the results are usually documented and published even if the results suggest quite modest effects or even no effects at all.

However, it is also important to put institutions in place to ensure that negative results are disseminated. Such a system is in place for medical trial results, and creating a similar system for documenting evaluations of social programs would help alleviate the problem of publication bias. For example, donors could require programs to submit the results of their evaluations to a database. Such a database would ideally be readily searchable and would contain numerous types of information that could be useful in interpreting the results (*e.g.*, estimates, sample size, region and time, type of project, cost, cost-benefit analysis, caveats). Over time, such a database could become a basic reference for organizations and governments as they seek project funding.

3. Randomized evaluations are feasible and can be conducted successfully, although they are labor-intensive. Non-governmental organizations are well-suited to conduct randomized evaluations but will require outside technical assistance and financing.

As is clear from the examples discussed in this chapter, a number of randomized evaluations have been conducted successfully in developing countries. Randomized evaluations are labor-intensive and costly, but no more so than other data collection activities. As the example of the initial PROGRESA program indicates, governments can sometimes conduct randomized evaluations successfully. However, political constraints often make it difficult for governments to randomize their programs, especially as governments are expected to serve their entire populations. For example, "Opportunidades," the urban version of PROGRESA, did not start with a randomized evaluation because of opposition to delaying access to the program for any randomly chosen control group.

Nongovernmental organizations in developing countries may be very well placed to conduct randomized evaluations. Unlike governments, NGOs are not expected to serve entire populations. Also unlike governments, financial and administrative constraints often lead NGOs to phase in programs over time, and randomization will often be the fairest way to of determining the phase-in order. In contrast to developed countries, where NGOs typically do not have sufficient resources to conduct large programs that could serve as a model for public policy, this is not the case in developing countries. Since many NGOs exist and they frequently seek out new projects, NGOs willing to conduct randomized evaluations can often be found. For example, the set of recent studies conducted in Kenya have been carried out through a collaboration with the Kenyan NGO Internationaal Christelijk Steunfonds (ICS) Africa. ICS was keenly interested in using randomized evaluations to see the impact of its programs as well in sharing credible evaluation results with other stakeholders and policymakers. A second example is the collaboration between the Indian NGO Pratham and researchers from the Massachusetts Institute of Technology that led to the evaluations of the remedial education (Banerjee *et al.*, 2000) and computer-assisted learning programs (Banerjee *et al.*, 2004). However, while NGOs are well placed to conduct randomized evaluations, expecting them to finance the research is less reasonable, as the results are global public goods. The evaluations of the ICS deworming programs were made possible by financial support from the World Bank, the Partnership for Child Development, the U.S. National Institutes of Health, and the MacArthur Foundation. In the case of the Indian educational programs, Pratham found a corporate sponsor, India's second-largest bank, ICICI Bank, which was keenly interested in evaluating the impact of the program and helped finance part of the evaluation.

4. Costs can be reduced and comparability enhanced by conducting a series of evaluations in the same area.

Once evaluation staffs are trained, they can work on multiple projects. Since data collection is the most costly element of these evaluations, cross-cutting the sample can also dramatically reduce costs. For example, many of the programs to increase school participation and learning were implemented in the same area, by the same organization. The teacher incentives (Glewwe, Ilias, and Kremer, 2004) and textbook (Glewwe, Kremer, and Moulin, 2002) programs were evaluated in the same 100 schools: one group had textbooks only, one had textbooks and incentives, one had incentives only, and one had neither. The effect of the incentive program should thus be interpreted as the effect of an incentive program conditional on half the schools having extra textbooks. Likewise, a

computer-assisted learning program was implemented in Vadodara, India, in the same set of schools where the remedial education study was conducted. This approach must consider potential interactions between programs (which can be estimated if the sample is large enough), and may be inappropriate if one program makes the schools atypical. Finally, as discussed in Section IV, another advantage is that conducting a series of studies in the same area (such as the set of studies recently conducted in Kenya) enhances comparability by allowing researchers to compare the cost-effectiveness estimates of different interventions in the same setting.

5. Randomized evaluations have limitations, but many of those limitations also apply to non-randomized studies.

Sample selection bias, attrition bias, and spillover effects can affect both randomized and retrospective evaluations. When conducting randomized evaluations, correcting for these limitations is often easier than when conducting retrospective studies.

Sample selection problems could arise if factors other than random assignment influence program allocation. For example, parents may attempt to move their children from a class (or a school) without the program to a class with the program. Conversely, individuals allocated to a treatment group may not receive the treatment (for example, because they decide not to take up the program). Even if randomized methods have been employed and the intended allocation of the program was random, the actual allocation may not be. This problem can be addressed through intention to treat methods or by using random assignment as an instrumental variable for actual assignment. It is much harder to address in retrospective studies, since it is often difficult to find factors that plausibly effect exposure to the program that would not affect education outcomes through other channels.

A second issue affecting both randomized and retrospective evaluations is differential attrition in the treatment and the comparison groups: program participants may be less likely to move or otherwise drop out of the sample than nonparticipants. However, at a minimum, randomized evaluations can use statistical techniques to bound the potential bias and can attempt to track down individuals who drop out of the sample (*e.g.*, administer tests to students who have dropped out of school), which is often not possible with retrospective evaluations.

Third, programs may create spillover effects on untreated people. These spillovers may be physical, as found for the Kenyan deworming program. Deworming interferes with disease transmission and thus makes children in treatment schools (and in schools near treatment schools) less likely to have worms even if they were not themselves given the medicine. Spillovers may also operate through prices. Vermeersch and Kremer (2004) found that provision of meals in some schools leads other schools to reduce school fees. Finally, there might also be learning and imitation effects (Duflo and Saez, 2004; Miguel and Kremer, 2004b). If such spillovers are global (*e.g.*, due to changes in world prices), identification of total program impacts will be difficult with any methodology. However, if such spillovers are local, randomization at the group level can allow estimation of the total program effect within groups and can generate sufficient variation in local treatment density to measure spillovers across groups. For example, the solution in the case of the deworming study was to choose the *school* (rather than the pupils within a school) as the

unit of randomization and to look at the number of treatment and comparison schools within neighborhoods. Of course, this requires a larger sample size.

One limitation of randomized evaluations is that the evaluation itself may cause the treatment group to change its behavior (the Hawthorne effect) or the comparison group to change its behavior (the John Henry effect). The Hawthorne and John Henry effects are specific concerns for randomized evaluations, but similar effects can occur in other settings. For example, the provision of inputs could temporarily increase morale among students and teachers, which could improve performance. While this would create problems for randomized evaluations, it would also create problems for fixed-effect or difference-in-difference estimates.

A final issue is that the program may generate behavioral responses that would not occur if the program were generalized. For example, children may switch into a school receiving additional inputs. This may affect the original pupils by increasing class size (if class size affects the outcome of interest). This would not be part of a reduced form effect because a nationwide adoption of the policy would not have this effect.

In summary, while randomized evaluation is not a bulletproof strategy, but the potential sources of bias are well known and can often be corrected. This stands in contrast to biases of most other types of studies, where the bias due to the nonrandom treatment assignments often cannot be signed or estimated.

A challenge for the future is to integrate randomized evaluations with theory to shed light on issues of more general interest. In particular, evaluating various school reform initiatives is likely to shed light on more general issues of incentive and political economy.

References

Alderman, Harold, Jooseop Kim, and Peter Orazem (2003) "Design, Evaluation, and Sustainability of Private Schools for the Poor: the Pakistan Urban and Rural Fellowship School Experiments," *Economics of Education Review*, 22: 265-274.

American Psychological Association (1995) *Intelligence: Knowns and Unknowns*, Washington, DC.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King and Michael Kremer (2002) "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *American Economic Review*, 92(5): 1535-1558.

Angrist, Joshua, Eric Bettinger, and Michael Kremer (2004) "Long-term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," BREAD (Bureau for Research in Economic Analysis of Development) Working Paper #79.

Angrist, Joshua and Victor Lavy (2002) "New Evidence on Classroom Computers and Pupil Learning," *The Economic Journal*, 112(482): 735-786.

Angrist, Joshua and Victor Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement," *Quarterly Journal of Economics*, 114(2): 533-576.

Angrist, Joshua and Victor Lavy (1997) "The Effect of a Change in the Language of Instruction on the Return to Schooling in Morocco," *Journal of Labor Economics*, 15(1): S48-S76.

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek (2000) "A Review of Estimates of Schooling/Earnings Relationship, with Tests for Publication Bias," NBER working paper #7457.

Asian Development Bank (1999) "Education Costs and Financing in the Philippines," Technical Background Paper #2 for the 1998 Philippines Education Sector Study, Asian Development Bank, Manila.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden (2004) "Remedying Education: Evidence from Two Randomized Experiments in India," mimeo, MIT.

Banerjee, Abhijit, Suraj Jacob, and Michael Kremer with Jenny Lanjouw and Peter Lanjouw (2000) "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials," mimeo, MIT.

Barro, Robert (1991) "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*, 106(2): 407-443.

Barro, Robert and Jong-Wha Lee (2001) “International Data on Educational Attainment: Updates and Implications,” *Oxford Economic Papers*, 53(3): 541-563.

Bedi, Arjun S. and Jeffery H. Marshall (2002) “Primary School Attendance in Honduras,” *Journal of Development Economics*, 69(1): 129-153.

Behrman, Jere (1999) “Labor Markets in Developing Countries,” in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics: Vol. 3B*, Elsevier.

Behrman, Jere, Mark Rosenzweig, and Paul Taubman (1994) “Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment,” *Journal of Political Economy*, 102(6): 1131-1174.

Bils, Mark and Peter Klenow (2000) “Does Schooling Cause Growth?” *American Economic Review*, 90(5): 1160-1183.

Bleakley, Hoyt (2002) “Disease and Development: Evidence from Hookworm Eradication in the American South,” mimeo, MIT.

Bobonis, Gustavo, Edward Miguel, and Charu Sharma (2004) “Iron Deficiency Anemia and School Participation,” mimeo, University of California, Berkeley.

Bommier, Antoine, and Sylvie Lambert (2000) “Education Demand and Age at School Enrollment in Tanzania,” *Journal of Human Resources*, 35(1): 177-203.

Bray, Mark (1999) *The Shadow Education System: Private Tutoring and Its Implications for Planners*, International Institute of Educational for Educational Planning, Paris.

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

Buddlemeyer, Hielke and Emmanuel Skofias (2003) "An Evaluation on the Performance of Regression Discontinuity Design on PROGRESA," Institute for Study of Labor, Discussion Paper #827.

Campbell, Donald T. (1969) “Reforms as Experiments,” *American Psychologist*, 24: 407-429.

Case, Anne and Angus Deaton (1999) “School Inputs and Educational Outcomes in South Africa,” *Quarterly Journal of Economics*, 114(3): 1047-1085.

Case, Anne, Christina Paxson, and Joseph Ableidinger (2003) “The Education of African Orphans,” mimeo, Princeton University.

Chaudhury, Nazmul, Jeff Hammer, Michael Kremer, Karthik Muraldhiran, and Halsey Rogers (forthcoming) “Missing in Action: Teacher and Health Worker Absence in Developing Countries,” *Journal of Economic Perspectives*.

Chin, Aimee (2002) “The Returns to School Quality When School Quality is Very Low: Evidence from Operation Blackboard in India,” mimeo, University of Houston.

Coalition for Health and Education Rights (2002) “User Fees: The Right to Education and Health Denied,” New York.

Cox, D. and E. Jimenez (1991) “The Relative Effectiveness of Private and Public Schools: Evidence from Two Developing Countries,” *Journal of Development Economics*, 34: 99-121.

Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan (2004) "When Can School Inputs Improve Test Scores?" Policy Research Working Paper No. 3217 The World Bank. Washington, DC.

DeLong, J. Bradford and Kevin Lang (1992) “Are All Economic Hypotheses False?” *Journal of Political Economy*, 100(6): 1257-72.

Drèze, Jean and Geeta Kingdon (2001) “School Participation in Rural India,” *Review of Development Economics*, 5(1): 1-33.

Duflo, Esther (2001) “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, 91(4): 795-814.

Duflo, Esther and Michael Kremer (2005) “Use of Randomization in the Evaluation of Development Effectiveness,” in George Pitman, Osvaldo Feinstein, and Gregory Ingram (editors), *Evaluating Development Effectiveness*, New Brunswick, NJ: Transaction Publishers.

Duflo, Esther and Emmanuel Saez (2004) “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment,” *Quarterly Journal of Economics*, 118: 815-842.

Easterly, William (2001) *The Elusive Quest for Growth: Economists’ Adventures and Misadventures in the Tropics*, MIT Press.

Evans, David and Edward Miguel (2004) “Orphans and Schooling in Africa: A Longitudinal Analysis,” Harvard Center for International Development, working paper #056.

Foster A.D. and M.R. Rosenzweig, (1996) "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103 (6): 1176-1209, 1995.

Frankenberg, E., K. Beegle, D. Thomas and W. Suriastini (1999) "Health, Education, and the Economic Crisis in Indonesia," unpublished.

Fuller, B. and P. Clarke (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.

Gertler, Paul and Paul Glewwe (1990) "The Willingness to Pay for Education in Developing Countries: Evidence from Rural Peru," *Journal of Public Economics*, 42: 251-275.

Glazerman, Steven, Dan Levy, and David Meyers (2002) "Nonexperimental Replications of Social Experiments: A Systematic Review," Mathematica Policy Research, Inc. Interim Report/Discussion Paper.

Glewwe, Paul (1999) *The Economics of School Quality Investments in Developing Countries*, St. Martin's Press, New York.

Glewwe, Paul (2002) "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes," *Journal of Economic Literature*, 40(2): 436-482.

Glewwe, Paul (2004) "An Investigation of the Determinants of School Progress and Academic Achievement in Vietnam," in P. Glewwe, D. Dollar and N Agrawal, eds., *Economic Growth, Poverty, and Household Welfare in Vietnam*. The World Bank. Washington, DC.

Glewwe, Paul (2005) "The Impact of Child Health and Nutrition on Education in Developing Countries: Theory, Econometric Issues and Recent Empirical Evidence." *Food and Nutritional Bulletin* (forthcoming).

Glewwe, Paul, Margaret Grosh, Hanan Jacoby, and Marlaine Lockheed (1995) "An Eclectic Approach to Estimating the Determinants of Achievement in Jamaican Primary Education," *World Bank Economic Review*, 9(2): 231-258.

Glewwe, Paul, and Hanan Jacoby. 1992. "Estimating the Determinants of Cognitive Achievement in Low-income Countries: The Case of Ghana." Living Standards Measurement Study Working Paper No.91. The World Bank. Washington, D.C.

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

Glewwe, Paul, Nauman Ilias, and Michael Kremer (2004) "Teacher Incentives," mimeo, Harvard University.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin (2002) "Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya," Development Research Group, World Bank, Washington, DC.

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz (2004) "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics*, 74: 251-268

Greaney, Vincent, Shahidur R. Khandker, and Mahmudul Alam (1999) *Bangladesh: Assessing Basic Learning Skills*, University Press, Dhaka, Bangladesh.

Gould, Eric, Victor Lavy, and Daniele Paserman (2004) "Immigrating to Opportunity: The Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *Quarterly Journal of Economics* 119(2):489-526.

Habyarimana, James, Jishnu Das, Stefan Dercon, and Pramila Krishnan (2004) "Sense and Absence: Absenteeism and Learning in Zambian Schools," World Bank, Washington DC.

Hannaway, Jane (1992) "Higher Order Thinking, Job Design, and Incentives: An Analysis and Proposal," *American Education Research Journal*, 29(1): 3-21.

Hanushek, Eric (1995) "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer*, 10(2): 227-246.

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

Harbison, Ralph, and Eric Hanushek (1992) *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*, Oxford University Press.

Heckman, James (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1): 153-160.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd (1998) "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5): 1017-1098.

Heckman, James, Hidehiko Ichimura, and Petra Todd (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 64(4): 605-654.

Heckman, James, Robert LaLonde, and Jeffrey Smith (1999) "The Economics and

Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics: Vol. 3B,* Elsevier.

Heckman, James, Lance Lochner, and Christopher Taber (1998) "General Equilibrium Treatment Effects: A Study of Tuition Policy," NBER working paper #6426.

Herz, Barbara, K. Subbarao, Masooma Habib, and Laura Raney (1991) *Letting Girls Learn: Promising Approaches in Primary and Secondary Education*, World Bank Discussion Paper #133.

Holmstrom, Bengt and Paul Milgrom (1991) "Multi-Task Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7: 24-52.

Hoxby, Caroline (2000) "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90(5): 1209-1238.

Hsieh, C. and M. Urquiola (2002) "When Schools Compete, How Do They Compete?" mimeo, Princeton University.

Huffman, Wallace (2001) "Human Capital: Education and Agriculture," in B. Gardner and G. Rausser, editors, *Handbook of Agricultural Economics, Vol. 1A*, Elsevier.

IAEEA (2000) "TIMSS 1999: International Mathematics Report," International Study Center, Boston College.

IAEEA (2003) "PIRLS 2001 International Mathematics Report," International Study Center, Boston College.

Imbens, Guido and Joshua Angrist (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2): 467-475.

Jacoby, Hanan (2002) "Is There an Intrahousehold Flypaper Effect? Evidence from a School Feeding Program," *Economic Journal*, 112(476): 196-221.

Jamison, Dean, Barbara Searle, Klaus Galda, and Stephen 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): 556-567.

Jimenez, Emmanuel and Yasuyuki Sawada (1999) "Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program," *World Bank Economic Review*, 13(3): 415-441.

Jimenez, Emmanuel and Marianne Lockheed (1995) “Public and private secondary education in developing countries,” Discussion Paper #309, World Bank, Washington, DC.

Kim, Jooseop, Harold Alderman, and Peter Orazem (1999) “Can Private School Subsidies Increase Enrollment for the Poor? The Quetta Urban Fellowship Program,” *World Bank Economic Review*, 13(3).

Kingdon, Geeta (1996) “The Quality and Efficiency of Public and Private Schools: A Case Study of Urban India,” *Oxford Bulletin of Economics and Statistics*, 58(1): 55-80.

Kremer, Michael (2003) “Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons,” *American Economic Review Papers and Proceedings*, 93(2): 102-115.

Kremer, Michael (1995) “Research on Schooling: What we know and what we don’t – A comment on Hanushek,” *World Bank Research Observer*, 10(2): 247-54.

Kremer, Michael and Andrei Sarychev (2000) “Why Do Governments Operate Schools?” mimeo, Harvard University.

Kremer, Michael, Sylvie Moulin, and Robert Namunyu (2002) “Unbalanced Decentralization,” mimeo, Harvard University.

Kremer, Michael and Edward Miguel (2003) “The Illusion of Sustainability,” mimeo, Harvard University.

Kremer, Michael, Edward Miguel, and Rebecca Thornton (2004) “Incentives to Learn,” mimeo, Harvard University.

Kremer, Michael, Karthik Muraldhiran, Nazmul Chaudhury, Jeffrey Hammer, and F. Hasley Rogers (forthcoming) “Teacher Absence in India: A Snapshot,” *Journal of the European Economic Association*.

Krueger, Alan B. and Mikael Lindahl (2001) “Education for Growth: Why and for Whom?” NBER working paper #7591.

Lakdawalla, Darius (2001) "The Declining Relative Quality of Teachers" NBER working paper #8263.

Lavy, Victor (1996) “School Supply Constraints and Children’s Educational Outcomes in Rural Ghana,” *Journal of Development Economics*, 51(2): 291-314.

Lavy, Victor (2002) “Evaluating the Effect of Teachers’ Performance Incentives on Pupils’ Achievements,” *Journal of Political Economy*, 110(6): 1286-1318.

Levine, R. and D. Renalt (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82(4): 942-963.

Linden, Leigh and Kartini Shastry (2005) "Grain Inflation: Misrepresenting Student Attendance in Response to a School Meals Program," mimeo, Harvard University.

Lockheed, Marlaine, and Adriaan Verspoor (1991) *Improving Primary Education in Developing Countries*, Oxford University Press, New York.

Long, Sharon K. (1991) "Do the School Nutrition Programs Supplement Household Food Expenditures?" *The Journal of Human Resources*, 26: 654-678.

Lucas, Robert (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22: 3-42.

Mankiw, N. Gregory, David Romer and David Weil (1992) "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107(2): 407-437.

Manski, Charles and Irwin Garfinkel (editors) (1992) *Evaluating Welfare and Training Programs*, Harvard University Press.

Meyer, Bruce (1995) "Natural and Quasi-Experiments in Economics," *Journal of Business and Economics Statistics*, 3(2):151-168.

Miguel, Edward (2003) "Tribe or Nation? Nation-building and Public Goods in Kenya versus Tanzania," mimeo, University of California, Berkeley.

Miguel, Edward and Mary Kay Gugerty (2004) "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, forthcoming.

Miguel, Edward and Michael Kremer (2003) "Social Networks and Learning About Health in Kenya," mimeo, Harvard University.

Miguel, Edward and Michael Kremer (2004) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1): 159-217.

Morley, Samuel and David Coady (2003) *From Social Assistance to Social Development: Education Subsidies in Developing Countries*, Institute for International Economics, Washington, DC.

Pagan, Adrian, and Aman Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press, Cambridge, UK.

Pitt, Mark M., Mark R. Rosenzweig, and Donna M. Gibbons (1993) "The Determinants and Consequences of the Placement of Government Programs in Indonesia," *World Bank Economic Review*, 7(3): 319-348.

Planning Institute of Jamaica (1992) "Jamaica Survey of Living Conditions: Analytical Review, 1990," Kingston.

Poverty Action Lab (2005) *Fighting Poverty: What Works?* Bulletin 1.

Powell, Christine, Sally Grantham-McGregor, and M. Elston (1983) "An Evaluation of Giving the Jamaican Government School Meal to a Class of Children," *Human Nutrition: Clinical Nutrition*, 37C: 381-388.

Pritchett, Lant (1996) "Where Has All the Education Gone?" World Bank Policy Research Working Paper #1581.

Pritchett, Lant and Deon Filmer (1999) "What education production functions *really* show: a positive theory of education expenditures," *Economics of Education Review*, 18: 223-239.

PROBE Team (1999) *Public Report on Basic Education in India*, Oxford University Press, New Delhi.

Psacharopoulos, George (1985) "Returns to Education: A Further International Update and Implications," *Journal of Human Resources*, 20(4):583-604.

Psacharopoulos, George (1994) "Returns to Investment in Education: A Global Update," *World Development*, 22:1325-1344.

Reinikka, Ritva and Jakob Svensson (2004) "Local Capture: Evidence from a Central Government Transfer Program in Uganda" *Quarterly Journal of Economics* 119(2): 679-705.

Reinikka, Ritva and Jakob Svensson (2003) "The Power of Information: Evidence from a Campaign to Reduce Capture," mimeo, World Bank.

Rosenzweig, Mark and Kenneth Wolpin (2000) "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38: 827-874.

Rugh, Andrea (2000) *Starting Now: Strategies for Helping Girls Complete Primary*, Academy for Educational Development, SAGE Project, Washington, DC.

Schultz, Paul (1997) "Demand for Children in Low Income Countries," in M. Rosenzweig and O. Stark, editors, *Handbook of Population and Family Economics*, North Holland.

Schultz, T. Paul (2004) "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program," *Journal of Development Economics*, 74(1): 199-250.

Schultz, T. Paul (2002) "Why Governments Should Invest More to Educate Girls," *World Development*, 30: 207-225.

Sen, Amartya (1999) *Development as Freedom*, Knopf, New York.

Sen, Amartya (2002) "The Pratichi Report," Pratichi India Trust.

Strauss, John, and Duncan Thomas (1995) "Human Resources," in J. Behrman and T.N. Srinivasan, editors, *Handbook of Development Economics: Volume 3,* North Holland.

Svejnar, Jan (1999) "Labor Markets in the Transitional Central and East European Economics," in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics: Vol. 3B,* Elsevier.

Tan, Jee-Peng, Julia Lane, and Paul Coustere (1997) "Putting Inputs to Work in Elementary Schools: What Can Be Done in the Philippines?" *Economic Development and Cultural Change*, 45(4): 857-879.

Tanzania Media Monitoring, Weekly Newsletter from World Bank Country Office Dar Es Salaam, Sept. 30-Oct. 6, 2002.

UNDP (1990) *Human Development Report*, United Nations Development Programme, New York.

UNDP (2003) *Human Development Report*, United Nations Development Programme, New York.

UNESCO (2003) "UNESCO Institute of Statistics Electronic Database" http://www.uis.unesco.org/ev.php?URL_ID=5187&URL_DO=DO_TOPIC&URL_SECTION=201

UNESCO (2002) *Education for All: Is the World On Track?* UNESCO Publishing, Paris.

UNESCO (1997) *Statistical Yearbook* UNESCO Publishing, Paris.

UNICEF (1999) *The State of the World's Children*, Paris.

Vermeersch, Christel and Michael Kremer (2004) "School Meals, Educational Attainment, and School Competition: Evidence from a Randomized Evaluation," World Bank Policy Research Working Paper, WPS3523.

World Bank (1996) “India: Primary Education Achievement and Challenges,” Report #15756-IN, Washington, DC.

World Bank (1997) *Primary Education in India*, Washington, DC.

World Bank (1997b) *Vietnam: Education Financing*, Washington, DC.

World Bank (2001a) *World Development Report 2000/2001: Attacking Poverty*, Washington, DC.

World Bank (2001b) *Engendering Development: World Bank Policy Research Report*, World Bank and Oxford University Press.

World Bank (2002) *World Development Indicators 2002*, Washington, DC.

World Bank (2003) *World Development Indicators 2003*, Washington, DC.

World Bank (2004) *World Development Report 2004: Making Services Work for Poor People*, Washington, DC.

World Food Programme (2002) “Global School Feeding Report 2002,” World Food Programme School Feeding Support Unit, Rome.

Table 1. Primary School Gross Enrollment Rates (percent of students of primary school age)

Area	*1960*	*1970*	*1980*	*1990*	*2000*
World	80	87	97	102	104
Country group					
Low-income	65	77	94	102	102
Middle-income	83	103	101	103	110
High-income	109	100	101	102	102
Region					
Sub-Saharan Africa	40	51	80	74	77
Middle East/North Africa	59	79	89	96	97
Latin America	91	107	105	106	127
South Asia	41	71	77	90	98
East Asia	87	90	111	120	111
East Europe/Former Soviet Union (FSU)	103	104	100	98	100
Organization for Economic Cooperation and Development (OECD)	109	100	102	103	102

Note: 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.
Sources: Barro and Lee data set; UNESCO (2002); World Bank (2003)

Table 2. Primary School Enrollment, Repetition, and Grade 4 Survival Rates

(percents)

Area	***Gross enrollment 2000***	***Net enrollment 2000***	***Repetition 2000***	***On-time enrollment 2000***	***Grade 4 survival 1999***
Country group					
Low-income	102	85	4	55	80
Middle-income	110	88	10	61	88
High-income	102	95	2[a]	73[b]	98[b]
Region					
Sub-Saharan Africa	77	56	13	30	76
Middle East/North Africa	97	84	8	64	96
Latin America	127	97	12	74	86
South Asia	98	83	5	-	55
East Asia	111	93	2	56	97
East Europe/FSU	100	88	1	67[a]	97[b]
OECD	102	97	2[a]	91[a]	99[b]

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: UNESCO (2003)

Table 3. Secondary School Gross Enrollment Rates
(percent of students of secondary school age)

Area	*1960*	*1970*	*1980*	*1990*	*2000*
World	29	36	49	55	67
Country group					
Low-income	14	21	34	41	54
Middle-income	21	33	51	59	77
High-income	63	74	87	92	101
Region					
Sub-Saharan Africa	5	6	15	23	27
Middle East/North Africa	13	25	42	56	66
Latin America	14	28	42	49	86
South Asia	10	23	27	39	47
East Asia	20	24	44	48	67
East Europe/FSU	55	64	93	90	88
OECD	65	77	87	95	107

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: Barro and Lee data set; UNESCO (2003); World Bank (2003)

Table 4. Average Years of School of Adults, Age 15+

Area	*1960*	*1970*	*1980*	*1990*	*2000*
Country group					
Low-income	1.6[a]	2.2[a]	3.7	4.6	5.2
Middle-income	2.8	3.5	4.2	5.1	5.9
High-income	7.4	7.9	9.2	9.5	10.1
Region					
Sub-Saharan Africa	1.7	2.0	2.3	3.0	3.4
Middle East/North Africa	1.4	2.2	2.9	4.1	5.4
Latin America	3.2	3.7	4.4	5.3	6.0
South Asia	1.5	2.0	3.0	3.8	4.6
East Asia	2.5[b]	3.4[b]	4.6	5.6	6.2
East Europe/FSU	6.5[b]	7.6[b]	8.5[b]	9.0[b]	9.7[b]
OECD	7.3	7.8	9.1	9.5	10.1

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: Barro and Lee (2000)

Table 5. Literacy Rate among Adults, Age 15+

Area	*1960*	*1970*	*1980*	*1990*	*2000*
Country group					
Low-income	32[a]	44	54	63	70
Middle-income	62	68	75	80	85
High-income	95[a]	96[a]	97[a]	98[a]	98[a]
Region					
Sub-Saharan Africa	24[b]	41	54	67	77
Middle East/North Africa	33[b]	54	66	76	83
Latin America	67	84	90	93	95
South Asia	26	43	52	61	69
East Asia	54[b]	83	91	95	97
East Europe/FSU	93[b]	99	100	100	100
OECD	95	98	99[b]	100[b]	100[b]

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: UNESCO (2003)

Table 6. Gender Disparities in Gross Primary and Secondary Enrollment Rates, 2000

Area	*Primary*		*Secondary*	
	Boys	***Girls***	***Boys***	***Girls***
Country group				
Low-income	107	98	60	47
Middle-income	112	108	77	78
High-income	102	101	100	102
Region				
Sub-Saharan Africa	83	71	29	24
Middle East/North Africa	101	92	71	61
Latin America	129	125	83	89
South Asia	107	90	53	39
East Asia	112	111	73	60
East Europe/FSU	100	99	88	89
OECD	102	102	106	108

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: World Bank (2003)

Table 7. Mean Mathematics and Reading Achievement, TIMSS and PIRLS Studies

	Mathematics (TIMSS)		***Reading (PIRLS)***
	1999		***2001***
	Grade 7	*Grade 8*	*Grade 4*
Country			
France	-	-	525
Japan	-	579	-
U.K. (England)	-	-	553
U.S.	-	502	542
Argentina	-	-	420
Belize	-	-	327
Chile	-	392	-
Colombia	-	-	422
Indonesia	-	403	-
Iran	-	422	414
Jordan	-	428	-
Korea (South)	-	587	-
Kuwait	-	-	396
Malaysia	-	519	-
Morocco	337	-	350
Philippines	345	-	-
South Africa	-	275	-
Thailand	-	467	-
Tunisia	-	448	-
Turkey	-	429	449

Source: IAEEA (2000, 2003)

Table 8. Mathematics and Reading Achievement of 15 Year Olds, PISA Study

	Mathematics	***Reading***	
	Mean score	Mean score	Percent with very low skills
Country			
France	517	505	4.2
Japan	557	522	2.7
United Kingdom	529	523	3.6
United States	493	504	6.4
Argentina[a]	388	418	22.6
Brazil	334	396	23.3
Chile[a]	384	410	19.9
Indonesia[a]	367	371	31.1
Mexico	387	422	16.1
Peru[a]	292	327	54.1
South Korea	547	525	0.9
Thailand[a]	432	431	10.4

Notes: Data are for the year 2000.
a. Data are for the year 2002.
Source: OECD and UNESCO (2003)

Table 9. Government Expenditures on Education (percentage terms)

Area	*Expenditure as percent of GDP*		*Expend. per student as percent of GDP per capita*		*Expenditures per tertiary student as a ratio of expenditures per student at lower levels*	
Country group	***Primary***	***Secondary***	***Primary***	***Secondary***	***Primary***	***Secondary***
Low-income	1.0	1.1	7.0	16.7	33.6	13.6
Middle-income	1.8	1.4	13.3	15.5	5.0[a]	4.4[a]
High-income	1.4	1.9	18.8	21.5	1.8[a]	1.3
Region						
Sub-Saharan Africa	1.9[a]	1.2[a]	10.6	25.8	198.5[a]	81.1[a]
Middle East/North Africa	1.8[a]	1.4[a]	15.0	19.5	5.4	5.3
Latin America	1.6	1.6	12.2	14.3	4.3	4.0
South Asia	1.0	1.2	7.4	22.0	5.6[b]	3.3[b]
East Asia	0.9	0.8	6.6	11.8	12.5	6.5
East Europe/FSU	0.2[b]	2.3[b]	21.4[c]	19.1	2.0[b]	1.5
OECD	1.2[b]	2.1[b]	18.6	22.8	1.8[a]	1.5[a]

Notes: Expenditures as a percent of GDP are for 2000.

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.

c. Data are based on less than 10 percent of the total population of the country group or region.

Source: UNESCO (2003)

Table 10. Government Expenditures on Education (in dollars)

Area	*Expenditure per student*			
	US dollars		*PPP dollars*	
	Primary	***Secondary***	***Primary***	***Secondary***
Country group				
Low-income	48	87	202	366
Middle-income	555[a]	660[a]	833[a]	1013
High-income	3263[c]	4279[c]	3059[c]	3915[c]
Region				
Sub-Saharan Africa	68[b]	171[b]	338[b]	638[b]
Middle East/North Africa	157[b]	316[a]	429[b]	809[a]
Latin America	364[b]	504[a]	588[b]	877[a]
South Asia	34	66	167	322
East Asia	66	101	214	347
East Europe/FSU	564[b]	555[b]	1401[b]	1250[b]
OECD	4310[b]	5655[a]	3760[b]	4933[a]

Notes: Expenditures per student are for 1996.
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.
c. Data are based on less than 10 percent of the total population of the country group or region.
Source: World Bank (2003)

Table 11. Pupil Teacher Ratios and Percentage of Teachers with Training

Area	*Pupil teacher ratio*		*Percent trained teachers*		*Teacher salary as percent of per capita GDP*
	Primary	*Secondary*	*Primary*	*Secondary*	
Country group					
Low-income	32	25	90	69[b]	
Middle-income	25	20	90	83[a]	
High-income	16	14	-	-	
Region					
Sub-Saharan Africa	43	24	69	78[b]	6.7
Middle East/North Africa	23	18	96	85[a]	
Latin America	26	19	87	77	1.4
South Asia	42	33	62[b]	-	
East Asia	22	19	96	71[b]	
East Europe/FSU	17	12	93[b]	-	
OECD	16	14	-	-	1.3

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: UNESCO (2003)

Table 12: Absence Rates Among Teachers in Developing Countries

	Primary schools
Ecuador	14
India (average over 14 states)	25
Indonesia	19
Papua New Guinea	15
Peru	11
Zambia	17
Uganda	27

Notes: The absence rate is the percentage of staff who are supposed to be present but are not on the day of an unannounced visit. It includes staff whose absence is "excused."
*Source*s: Chaudhury and others (forthcoming), Habyarimana and others (2004), and NRI and World Bank (2003).

Table 13: Summary of 96 Studies on the Estimated Effects of Resources on Education in Developing Countries

Input	Number of studies	Statistically significant		Statistically insignificant
		Positive	Negative	
Teacher-pupil ratio	30	8	8	14
Teacher's education	63	35	2	26
Teacher's experience	46	16	2	28
Teacher's salary	13	4	2	7
Expenditure per pupil	12	6	0	6
Facilities	34	22	3	9

Source: Hanushek (1995).