



Determinants of Primary Education Outcomes in Developing Countries

Background Paper for the Evaluation of the World Bank's Support to Primary Education

Maurice Boissiere

Director-General, Operations Evaluation: Gregory K. Ingram
Director, Operations Evaluation Department: Ajay Chhibber
Group Manager, Sector, Thematic, and Global Evaluation: Alain Barbu
Task Manager: H. Dean Nielsen

This paper is available upon request from OED.

2004
The World Bank
Washington, D.C.



**ENHANCING DEVELOPMENT EFFECTIVENESS THROUGH EXCELLENCE
AND INDEPENDENCE IN EVALUATION**

The Operations Evaluation Department (OED) is an independent unit within the World Bank; it reports directly to the Bank's Board of Executive Directors. OED assesses what works, and what does not; how a borrower plans to run and maintain a project; and the lasting contribution of the Bank to a country's overall development. The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank's work, and to provide accountability in the achievement of its objectives. It also improves Bank work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

OED Working Papers are an informal series to disseminate the findings of work in progress to encourage the exchange of ideas about development effectiveness through evaluation.

The findings, interpretations, and conclusions expressed here are those of the author(s) and do not necessarily reflect the views of the Board of Executive Directors of the World Bank or the governments they represent.

The World Bank cannot guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply on the part of the World Bank any judgment of the legal status of any territory or the endorsement or acceptance of such boundaries.

Contact:
Operations Evaluation Department
Knowledge Programs and Evaluation Capacity Development (OEDKE)
e-mail: eline@worldbank.org
Telephone: 202-458-4497
Facsimile: 202-522-3125
<http://www.worldbank.org/oed>

Acronyms and Abbreviations

CONFEMEN	Conference of Ministers of Education having French as a common language
EFA	Education for All
FTI	Fast Track Initiative within EFA programs
GER	Gross Enrollment Rate
Laboratorio	Latin American Program for the Assessment of Quality in Education
MDGs	Millennium Development Goals
MLA	Monitoring Learning Achievement (international group from all regions sponsored by UNESCO)
NAEP	National Assessment of Education Progress
NER	Net Enrollment Rate
OLS	Ordinary Least Square multiple regression analysis
PASEC	Program for the Analysis of Education Systems of CONFEMEM Countries (Francophone West Africa)
PISA	Program for International Student Assessment
PTR	Pupil-Teacher Ratio
Progresa	Mexican anti-poverty program, including education subsidies
SACMEQ	Southern African Consortium for Measurement of Education Quality
SES	Socio-economic status
TIMSS	Third International Mathematics and Science Study
UPE	Universal Primary Education
UPC	Universal Primary Completion
UNESCO	United Nations Education, Scientific and Cultural Organization

Contents

1. Introduction.....	1
2. Conceptual Framework	2
3. Education Production Functions	3
<i>Conceptual and Methodology Issues</i>	<i>3</i>
<i>Productivity and Efficiency Issues</i>	<i>7</i>
<i>Distribution Effects and Equity Issues</i>	<i>10</i>
4. Alternative Approaches to Determinants of Outcomes.....	11
<i>Randomized Evaluations.....</i>	<i>12</i>
<i>Natural Experiments.....</i>	<i>12</i>
<i>Comparative Benchmarking</i>	<i>13</i>
<i>Qualitative Methods.....</i>	<i>14</i>
5. Specific Inputs and Related Issues from EPF and Alternative Approaches.....	15
<i>“Hardware” Issues</i>	<i>16</i>
<i>“Software” Issues</i>	<i>18</i>
<i>Teachers</i>	<i>21</i>
<i>Management and Institutional Structure</i>	<i>23</i>
<i>Contextual and Background Factors.....</i>	<i>26</i>
6. Can Any Firm Conclusions Be Drawn?	27
Annex A: Working Definitions of Inputs and Outcomes	31
Annex B: Econometric Issues Related to Education Production Function	33
Bibliography	36

1. Introduction

1.1 This paper is a review of the literature on the determinants of primary education outcomes in developing countries. In today's world, simply getting children into schools is not enough; governments must also ensure that children complete the primary cycle and attain the basic knowledge and skills needed for personal well-being and national development. This review examines recent research into the determinants of the key outcomes — completion, numeracy, and literacy. Such a seemingly straightforward question — “What determines the outcomes of education?” — has generated volumes of controversial research over the years, and the end is still not in sight. Education policymakers and administrators have looked to this literature for guidance, and have found few clear-cut results to guide them. In the end, education is a complex enterprise, and decision-makers still have to fall back upon their experience and practical judgement. But these studies of outcomes and their determinants, when viewed in the context of each educator's and policymaker's own experiences, can help them select better choices from among the available alternatives.

1.2 Many studies have examined how total resources devoted to education or resources per student affect education outcomes. Other studies have sought to define the dimensions of quality education. For example, the World Bank's Primary Education Policy Paper (1990), based in large part upon a comprehensive review of research up to that time,¹ identified five principal contributors to primary education effectiveness: (1) curriculum, (2) learning materials, (3) instructional time, (4) classroom teaching, and (5) students' learning capacity. This review assesses how the research, especially since 1990, has addressed the importance of these as well as others factors (such as school facilities, teacher training, and management), and offers some insights into the circumstances under which the various factors make a difference.

1.3 The review also describes different research methodologies used in identifying determinants of education outcomes, since flawed methods can lead to invalid conclusions. In some cases, the methodologies are advanced statistical methods, but this is not a review of statistical and econometric methods per se. Instead, the emphasis is on providing policymakers and practitioners working in and for developing countries with fresh insights into how learning outcomes can be improved.

1.4 Although this review is concerned with developing countries, it sometimes draws on research in developed countries, since many of the research techniques used in the least-developed countries were first pioneered in developed countries, and then applied with appropriate modifications to new settings. Also, since this review focuses on the outcomes of primary education it excludes research on higher education, except for that related to training primary school teachers. Some of the

1. See Lockheed and Verspoor (1991) *Improving Primary Education in Developing Countries*. The book has a wealth of details to back up the Policy Paper and emphasizes that learning should be at the center of primary education programs, while also being aware of costs and finance from the perspective of their impact on learning. However, they note that little research had been done up to that time comparing children's learning over time. This paper reviews some of the progress that has been made since then.

research on secondary education is included since many studies include both primary and secondary as part of their research designs and results. However, secondary education has distinctive features, where teachers need to be trained more deeply in their subject areas and students are at a higher stage of cognitive development.

2. Conceptual Framework

2.1 Studies of education outcomes often are framed in terms of the supply-side factors, but demand-side factors are also important in determining education outcomes. Participation in school is regarded here as an input, and completing primary school with the acquisition of basic knowledge and skills is regarded as a desirable outcome. These both depend also upon various demand factors at the household level and within the broader social environment. While this conceptual framework adopts a more economics-oriented perspective, it is recognized that education, in addition to being a professional practice, is also an interdisciplinary social science. In particular, the insights and models of educational psychology and the sociology of education, which have studied the determinants of education outcomes before the more recent interests of economics, have an essential role to play. Examining both supply and demand factors in the determination of education outcomes provides a more complete framework for policymaking and assessment.

2.2 In analyzing the determinants of education outcomes, education economists have studied specific aspects of supply and demand for education, but rarely in a general equilibrium framework. This has been done sometimes in education planning models, which are outside the scope of this review, since these are typically mathematical constructs that assume the relationships that empirical research tries to document with evidence. Even in those studies that are mainly empirical, with very little explicit theoretical modeling to underpin them, there are usually implicit assumptions about supply and demand. The goal behind these empirical studies is to identify those factors that can be verified as significant determinants of outcomes and the policy implications thereof.

2.3 Without adopting any one of these theoretical models of supply and demand for primary education, this review outlines the factors that have been proposed as important for the supply or demand side of education, noting that some factors could actually serve both sides in a larger model. Typically, such models are formulated with households on the demand side and schools as the production units on the supply side (while noting that households also play a role on the supply side). Households demand more education because there is private economic rate of return to acquiring human capital, as well as social and cultural benefits. But more education has a cost to households, especially poor and rural ones, which face serious income, asset, and credit constraints. Cultural impediments to female education and formal sector employment (in the adult years) are also important demand factors. Thus, in poor countries households may not enroll their children, even when they have access to school, so it is important to understand the underlying reasons for household demand for education.

2.4 School quality and learning outcomes can play a role in both supply and demand of education, as with most goods and services. If parents in poor rural households perceive the quality of their children's schooling to be poor (for example, the building is unsafe or teachers do not show up) they may be reluctant to send their children (White 2004). Moreover, if they find that children who attend learn no more than those who do not, as has been demonstrated in some studies in Africa (Glewwe 2002, and Galabawa et al. 2000), then families may decide that benefits do not justify the costs. Thus, families may have access to schools (at a reasonable distance and cost) but choose not to enroll or send their children on a regular basis. Demand factors often include judgments on the part of families about the returns to schooling in terms of marketable knowledge and skills (literacy and numeracy) compared to school costs in terms of both direct costs (fees, supplies, and uniforms) and indirect ones (loss of household labor). A favorable calculation would increase the demand for education, even by poor and rural households (Glewwe 2002).

3. Education Production Functions

Conceptual and Methodology Issues

3.1 Schools can be treated analytically as production units on the supply side, but, with few exceptions, they are not profit-maximizing firms, most of them being public or private non-profit. This leads to issues of how to treat the behavior of schools in a theoretical economic model, and to further interesting observations and hypotheses related to school organization, management, and governance that are important to the delivery of quality education services from schools.

3.2 Thinking of schools as producers of education services leads naturally to the notion of the education production function as used in microeconomic theory. Education economists realize that the theory of the production function for economic firms would need appropriate modification when used for schools, but the belief is that the basic idea of employing capital, labor, and other inputs to produce specific outputs can be useful. Pritchett and Filmer (1999) note that the appropriate modification needed in the education context is that schools are treated as organizations that should try to maximize output, for example, learning in a variety of knowledge areas, subject to their budget constraints. The mathematical relationship between inputs and education outputs² is referred to as the education production function by economists, and has led to a large literature on the productivity of schools.

3.3 The output of the education production function is usually specified as some level of achievement as measured by test scores. However, the value-added specification can also be used in which the difference between test scores (hence the

2. The project planning and evaluation literature makes a distinction between outputs and outcomes. In economic theory, the output would be the result from the production function, i.e., $y = f(x)$. Then the outcome would be the value or utility of that output, namely $u(y)$, as valued by some metric, which can be monetary or non-monetary. In this paper, the distinction is not rigidly maintained.

term “value added” as used in economic production functions) is sometimes used. The value-added specification is especially useful when it can be tied to particular education interventions to see what impact they have over time, in some cases over the course of a year. The value-added approach is now being adopted in some states of the United States (North Carolina and Tennessee) as an almost “real time” assessment of individual student progress. In this assessment model, students are tested at the beginning and end of each year, using standards-based tests derived from the official curriculum standards. The actual progress of each student can be tracked year to year and related to many factors, including the individual teacher.

3.4 Before reviewing the recent research (mainly since 1990), it is worthwhile to review briefly the development of the concept of education production functions. Many researchers trace the origins of concern about the determinants of education outcomes and the production function approach to the Coleman Report (1966) done in the United States to investigate equal opportunity issues during the 1960s. Research on these topics in developing countries subsequently drew on methods used in the United States and now the developing country research has also broken new ground in terms of methodology (randomized experiments) and issues covered (effects of determinants for countries at different levels of development).

3.5 The Coleman Report stirred up considerable controversy by coming to the surprising conclusion that variations in school resources did not explain much of the variation in students’ achievement. The importance of schools and teachers for the achievement of students seemed much less critical than that of the students’ socioeconomic status (SES) as indicated by a number of family background characteristics, such as parental education, profession, and income. The controversy surrounding these conclusions inspired a large body of research in both developed and developing countries. In one of the first of these studies for developing countries, Heyneman (1979) analyzed a large sample survey of Ugandan students and found that SES was not as important in Uganda as it was in the United States.

3.6 The Coleman Report was criticized for a number of methodological reasons, and more studies were conducted on the question of the variations in effectiveness of teachers, student-teacher ratios, and other dimensions of schools. Hanushek’s more recent review (1986) of production function studies in the United States shows average spending to have risen over time while test scores remained flat, a problem he attributes to the weak affect of school inputs (see more in paragraphs 3.19 and 3.20 about school productivity). His subsequent review in developing countries (1995) reached essentially the same conclusion. He found the traditional approach to improving student outcomes — increasing inputs — an ineffective policy option, given that no systematic relationship can be found between inputs in the aggregate and test scores. Even when reviewing the studies of particular inputs, like teacher quality, he found equivocal results. Other researchers, such as Kremer (commenting on Hanushek’s article, 1995), while agreeing with the overall conclusion about aggregated inputs, maintained that particular input resources, such as more text books or use of educational radio, had been clearly demonstrated to affect student test scores.

3.7 In an even more recent article, Glewwe (2002) reviewed some of the issues surrounding econometric estimation of education production functions. He used a linear form of the production function (shown below, but see Annex B for a derivation of the linear form from a non-linear one) to provide a convenient framework to clarify the issues surrounding multiple regression analyses of cognitive achievement tests:

$$H = c + s * S + a_1 * A_1 + a_2 * A_2 + \dots + a_n * A_n + q_1 * Q_1 + q_2 * Q_2 \dots + q_m * Q_m + u$$

H is human capital using a measure of knowledge, such as achievement test scores

S is schooling (usually years of schooling)

A_i 's represent a series of individual student ability and learning capacities, such as IQ

Q_i 's represent school quality factors, such as class size, teacher qualifications, etc.

The lowercase c is the constant term, and s, a_1 to a_n , q_1 to q_m represent the coefficients multiplying the variables S, A_i 's and Q_i 's.

The variable u represents the random disturbance term due in part to the unpredictability of human behavior. In addition, it may include omitted variables and measurement errors in the variables.

Policymakers would be interested in the size and statistical significance of the coefficients in this education production function since that would give some idea of the impact of the various factors thought to be determinants of education outcomes. It is especially useful if this is combined with information about the unit cost of the inputs in order to compare relative cost-benefit ratios of the inputs.

3.8 Glewwe concludes, after a review of some prominent education production function (EPF) studies conducted in the 1990s, that there are no simple solutions to the methodological problems that beset most estimation of education production functions. He cites the possibility of biased estimates of key parameters that could result from four serious econometric problems: omitted variables such as student ability, endogenous school quality resulting from choices parents might make, measurement errors in the variables of interest such as inaccurate education data, and sample selectivity problems such as some students not being selected into the sample (see Annex B for technical discussion of these issues). It should be noted in this context that unbiased means that the estimated parameter would be correct on average with a certain statistical distribution about that average, with a larger sample size giving a tighter distribution around the average.

3.9 After reviewing some of the better conventional production function results in the literature, and in light of the above difficulties, Glewwe concludes that production function studies of student achievement should be taken as suggestive, and not definitive. He also seems to be in agreement with Hanushek's (1995) view that it is doubtful that the above difficulties could be easily resolved. However, Glewwe mentions on a positive note that if a number of good conventional studies agree on the significance of an input, there might be good reason to believe in a causal connection. His main caveat is that there are too few good conventional EPF studies,

and a much better job should be done in conducting conventional EPF studies. Due in part to the possibility of biases in the estimation of education production functions, Glewwe and Kremer (see his 1995 comment on Hanushek) make a case for using experimental designs (random control groups) for program implementation and evaluation, especially for interventions such as textbooks and materials and their impact on education outcomes (see the section below on alternative approaches).

3.10 Glewwe and Kremer also propose that future production function studies should improve their results by increasing the sample size. The four studies reviewed in detail by Glewwe, which he ranked among the better ones in terms of design, had relatively small sample sizes. The Brazil study sample size was 250 (for the best specification), Ghana was 163, India was 902 (but across only 30 schools), and Jamaica was 355. Increasing the sample size would give more precision to tests of statistical significance and estimates of the coefficients. Glewwe recommends that these decisions be considered during the planning stage of the survey. This is important for cost-benefit studies when confidence in the size and significance of coefficients used to calculate benefits is critical. However, large sample sizes will still not remove biases caused, for example, by omitting important explanatory variables.

3.11 Not all researchers are as pessimistic as Glewwe and Hanushek about the difficulties of estimating education production functions. These researchers maintain that through extra care and planning most of these problems can be overcome. A recent example is the Third International Mathematics and Science Study (TIMSS) (Wossman 2003), which involved a very large sample (more than 260,000 students from 39 countries). The parameter estimates for this production function show that international differences in student performance cannot be attributed to resource differences, in line with the Hanushek view of education productivity from U.S. studies. However, institutional features (such as centralized exam systems or school autonomy features), and competition from private schools are factors having significant explanatory power, as are family background variables (as in the Coleman Report). Wossman argues that omitted variables on IQ or other innate ability measures would not bias these results since across countries the average abilities are about the same. Concerning the endogeneity of resources to student performance, he indicates that, while this could bias estimates, it is not likely, since mobility across the 39 countries is not large enough to be a factor.

3.12 Hanushek (2003) has also used the TIMSS data set in estimating an EPF and reaches conclusions similar to his earlier ones about inefficiency in schools. He concludes that his EPF analysis using the TIMSS data set fails to confirm the original Heyneman hypothesis that school resources are relatively more important compared to family background in poor countries. Likewise, his comparison of poorer and richer countries in the TIMSS data set does not confirm his earlier finding of diminishing returns in the education production function (see Annex B for a discussion of this). However, until recently the TIMSS data set only has included only one country from sub-Saharan Africa, namely South Africa, which is not nearly the poorest in the region. So the poorest parts of the developing world, mainly most of sub-Saharan Africa and South Asia, have not been in this data set. Also, transition countries like Romania and Russia are included, which historically have a different pattern of GNP per capita and

educational achievement. The 2003 round of TIMSS starts to address this issue since Botswana and Ghana are participating.

3.13 Another assessment program of a similar scale is OECD's Program for International Student Assessment (PISA), an internationally standardized assessment of achievement in reading (2000), mathematics (2003), and scientific literacy (2006) among 15 year olds in 43 countries, including about 10 developing countries. However, since TIMSS and PISA include relatively few low-income countries, their findings have limited applicability in such settings. Fortunately, there is a new generation of assessment programs coming up which are more oriented to less-developed countries, such as (1) Southern Africa Consortium (southern and east Africa) for the Measurement of Education Quality or SACMEQ, (2) Program for the Analysis of Education Systems of the CONFEMEN Countries (Francophone west Africa) or PASEC, (3) Latin American Program for the Assessment of Quality in Education or Laboratorio, and (4) a group of 40 African, Arab, Asian, European, and Latin American countries in Monitoring Learning Achievement or MLA. These new assessments generally include achievement tests in the core subjects of language, mathematics, and sometimes science, and their results are used in multiple regression analyses such as those employed by TIMSS. In all of these programs the goals include the generation of information that can be used by decision-makers to improve quality and to enhance the research and evaluation capacity of the national education systems.

3.14 UNESCO is involved with all of these projects and published a study (Fiske 2000), which summarized the multiple regression results that tried to identify factors that contribute to achievement. Although the analyses to date are not complete, it is encouraging to see such data collection and analytical efforts from developing countries that are planned to continue on a systematic basis. The detailed results form a somewhat complex pattern and will be dealt with in the later section on the impact of specific inputs and factors. However, Fiske cites a number of studies that reach the general conclusion that increasing specific education inputs generally matters more for low-income countries than they do for the high-income ones, mainly because the level of these inputs is much lower in poor countries. This is consistent with the generic production function interpretation that involves diminishing returns to more inputs.

Productivity and Efficiency Issues

3.15 Hanushek was one of the first economists to emphasize the issue of inefficiency and declining productivity of education, first in the U.S. context (Hanushek 1986). This is based on the empirical results that show more resources per student (measured by expenditures per student) do not result in commensurate gains in achievement (as measured on a variety of test scores). In the United States, achievement gains appear to be non-existent, and maybe even declining while spending per student keeps rising. Good data from the poorest developing countries on both achievement and per student cost over time are rare, but there are some studies that also indicate widespread educational inefficiencies in developing countries (Hanushek 1995).

3.16 Another recent study (Gundlach and Wossman 2001) also indicates declining educational productivity in OECD and East Asian countries, some of them the high-performing ones like Singapore that have been leading in education. Again, this was inferred on the basis of changes in test scores and increases in spending per student that implied learning achievement per dollar spent has been declining. An important question is to what extent this may apply to poorest countries of Africa and Asia, which are at much lower levels of both achievement and resources per student. This is especially pertinent as these poor countries strive to meet the goals of primary education for all by 2015.

3.17 In the U.S. context, Card and Krueger have done studies that seem to show some significant exceptions to the model of increasing inefficiencies. In an interesting symposium on primary and secondary education sponsored by Journal of Economic Perspectives (1996), Card and Krueger used data from two southern states (North and South Carolina) of the United States to show that increasing resources did result in increased quality of education during the 1930s. Their methodology was more along the lines of a natural experiment as described above for the Indonesia study, which to some extent used the Card and Krueger work as a model. They used data on school spending and subsequent labor market outcomes to argue that the spending must have resulted in quality improvement in education.

3.18 In the same symposium, Hanushek presented an updated version of the argument about resources and quality of education using much better data. The National Assessment of Education Progress (NAEP), which has been testing students in primary and secondary schools in the United States nationwide since the early 1970s, provided a time series of test scores in reading, mathematics, and other subjects that is widely regarded as a valid picture of achievement in the U.S. education system. The overall trend in achievement is flat, sometimes slightly declining, depending upon the subject matter of the tests. While Hanushek noted that the expenditure data could be better categorized according to areas of spending, it is overall a reliable time series of per student resources allocated to primary and secondary education in the United States. The conclusion about declining productivity of U.S. primary and secondary education seems to be supported by empirical evidence, however difficult or complex it is to explain.

3.19 These national results and the North and South Carolina cases present interesting contrasts in conclusions as well as methods. Hanushek points out that one way of reconciling these opposite results is that resources do matter when they are added to an already low resource base, such as the poorer states of the south during the depression era of the 1930s. He does not say so, but by extension, it seems plausible that in poor developing countries additional resources may make a significant impact as well, especially if they are directed into the right areas. Observations like these lead to some interesting work about allocating resources to the right area as was done by Pritchett and Filmer (1999).

3.20 Pritchett and Filmer present compelling evidence that there could be significant efficiency and productivity gains by reallocating the share of expenditures to areas of higher marginal productivity, such as learning materials (textbooks and other types of instructional materials). Based upon education production function research done for northeast Brazil (Harbison and Hanushek 1992) and for an urban area of India (Kingdon 1996), they estimated the marginal product per dollar of each input. Table 1 (extracted and modified slightly) summarizes the conclusions of Pritchett and Filmer concerning the relative size of the effects of various inputs.

Table 1: Increases in Test Scores per Dollar Spent on Input (relative to teacher salary)

<i>Input</i>	<i>Northeast Brazil (1980s)</i>	<i>India (1990s)</i>
Teacher Salary	1.0	1.0
School Facility Measure	7.7	1.7
Instructional Material	19.4	14.0

Source: Pritchett and Filmer (1999) using data from Harbison/Hanushek (Brazil) and Kingdon (India).

3.21 They found, that increases in test scores per dollar spent on learning materials was about 19 times greater than that of increases in teacher salary in the Brazil case. This would mean that the system was operating very far from conditions of optimal expenditures within a production function framework, which requires that the marginal product per dollar of each input be equalized. This would imply that if more resources were provided, they should not go equally to all areas, but should be provided for learning materials in a much greater proportion. Something like this could have happened in the cases like that of North and South Carolina. A similar situation exists in many of the poorest countries that are struggling to pay teacher salaries, while starving the budget for complementary inputs into the education production function.

3.22 The reasons for the overall inefficiency in high-income and low-income countries may not be the same. The allocative inefficiencies identified by Pritchett and Filmer may be a bigger part of the story in low-income and middle-income countries where textbooks and quality of facilities is at a lower level than in the high-income countries. The review by Fiske (2000) cites the argument by Husen (1990) that above a certain threshold of resources, which has been reached by high-income countries, family background contributes more to learning outcomes than school resources. Perhaps this is due to diminishing returns in the production function for school resources.

3.23 Wossman makes an argument something like this when he uses Baumol's "cost disease" model to explain declining education productivity in the relatively high-income and high-performing Asian and OECD countries. This model is based upon the much slower productivity growth in labor-intensive service sectors relative to other sectors of the economy (Baumol 1967). Another part of the story in high-income countries could be that the universal coverage of education brings in more students at risk of low learning outcomes, such as non-native speakers of the language of instruction or special education students with higher unit costs of schooling. Much better controlled studies are needed to explore this, but the evidence points to different root causes for the low- and high-income countries. The implication seems

to be that, with the right policies, low-income countries can expect higher returns to spending on school resources than high-income countries.

Distribution Effects and Equity Issues

3.24 Before closing this section on production function studies, it should be noted that only a few of these studies explicitly treat national-level distribution effects in terms of outcomes among the poor. Most of them were concerned with average effects across all groups and not with the variance by poverty status. The Brazil study by Harbison and Hanushek is an exception, emphasizing its focus on the poor in its title “Educational Performance of the Poor: Lessons from Rural Northeast Brazil.” The motivation for this study was to assemble the data needed to evaluate the EDURURAL Project that was supported by a Bank loan during the mid-1980s. The purpose of the project was to expand access to primary education, reduce repetition and dropout rates, and increase achievement among the rural poor in this region, which is known for its poverty relative to the rest of the country. The sampling design for the education production function database was to select a sample of primary schools from the northeast Brazil states and counties. Some schools were involved in the project; others that were not served as a comparison group. Achievement tests were given in Portuguese language and mathematics, primarily in grades 2 and 4. The evaluation also assembled a large background database on access to primary schools and promotion, repetition, and dropout of students.

3.25 The production function part of the Brazil evaluation showed the factors that matter for the poor to be broadly the same as for other production function studies. The quality of school facilities and textbooks availability significantly increased learning. Pupil-teacher ratios were not a significant factor, as in many other such studies. Although the study did not include a variable that measured innate ability of the students, by using a value-added specification in which test scores for two different years were used, the production function coefficients were not biased by the omission of an ability variable. In effect, the ability and motivation variables could be subtracted out. Although the study showed that some teachers were significantly more effective at producing learning results, this teaching capacity in the classroom was not related to any variables usually thought important for selecting good teachers, such as level of education and training, experience, or salary level.

3.26 Another study that examined the performance of the poor versus the better off households is that in Ghana done by Glewwe (1999) and the more recent World Bank/OED impact evaluation by H. White (2004). Glewwe et al. used the data from a nationwide household survey in Ghana, that was supplemented by achievement and ability test data for a sub-sample of individuals. The OED evaluation used a second household survey with achievement tests, thus updating the data from the original survey. In each case, the Ravens Progressive Matrices were used to control for ability. The production function results are similar to Brazil, with the same basic variables influencing outcome among the poor as among the rich, that is, there is not a different production function for the poor. But the allocation of resources to the poor in Ghana is less, and this results in the poor achieving less. As a result of decentralization

reforms in Ghana, including more reliance on community financing, inequities have arisen between school resources in poor communities and better-off communities.

3.27 Both of these studies show that what matters most for access by the poor to primary education is government policy and commitment. This is bolstered by other studies in Indonesia and India. The Indonesia case study (Filmer and Lieberman 2002) shows how government policy made a large difference in promoting access and equity to primary education via large-scale school construction programs. The main deficiency in the government policy was inadequate attention to education quality. Likewise, the OED Education Evaluation for India (Abadzi 2002) traced the history of government education policy and Bank involvement, showing how increased government commitment in the late 1980s and during the 1990s significantly increased access to primary education for the poor. The OED report also noted some of the remaining challenges in documenting achievement, especially for the lower castes and for girls. A special case study of the Indian District Primary Education Program was prepared by one of the Indian officials (Pandey 2000) involved, providing a rich history and summary of how many difficulties had to be overcome by committed government policies. These case studies, together with the results of production function studies, indicate that government policy in developing countries can have a larger impact upon primary education access and outcomes than would be the case in developed countries. Primary education outcomes for the poor in terms of completion and learning achievement also depend critically upon government policy for ensuring that adequate resources are provided for those critical inputs identified by production function studies.

4. Alternative Approaches to Determinants of Outcomes

4.1 The production function approach usually refers to an explicit mathematical equation between outcomes and inputs and an econometric strategy for estimating those relationships. Most often the significance and the strength of these relationships are investigated by multiple regression analysis, such as ordinary least squares or more sophisticated econometric methods to deal with problems of biased estimates. There are other approaches that are similar to the production function in spirit, insofar as the concern is with outcomes and determinants, but other statistical methods are used and the causal factors of concern may be interpreted more broadly than the notion of inputs into the production function. For example, government policies and programs may be under evaluation with respect to their impact on some desired outcome. Another distinction is the statistical methodology may not involve regression analysis of data used to estimate an explicit mathematical relationship. Two such prominent areas identified in Glewwe's review of cognitive achievement are randomized trials and natural experiments used for impact evaluation of policy or program intervention. These can also shed valuable light on how inputs, such as resources, school and student characteristics, etc. can influence education outcomes.

Randomized Evaluations

4.2 Randomized evaluations are not actually new, having been used in agriculture, experimental psychology, and medical research for many years. Their use in evaluating social programs is relatively new, having been introduced in studies of welfare and labor market programs in the United States during the 1970s (see Grossman 1994; and Newman, et al. 1994, for reviews), and it is becoming more common in the evaluation of education outcomes resulting from policy or program interventions. The basic idea is to establish a control group and a treatment group, with random assignment of individuals to each group. Education researchers can adapt this evaluation technique of using randomized control groups to a variety of experimental education interventions. Such randomized trial was used in the early 1980s to evaluate the outcomes of educational radio in Nicaragua (see Jamison, et al.), which was shown to have a significant positive impact. A more recent example of a randomized trial for developing countries is the case of a textbook program in Kenya.

4.3 One of the most promising advantages of randomized evaluations over education production function studies, which typically consider supply-side factors, is that they are well suited to studying demand-side issues. The study of demand-side factors is very important for understanding strategies to achieve the education MDGs of universal primary education for all and gender equity in education at all levels of education by the year 2015. One way many analysts consider to increase school participation of the poor is to reduce the cost of school, perhaps by subsidizing attendance. The Progresas Program in Mexico, which used a randomized phase-in of program implementation, allowed for the design of a randomized study of the program outcomes (see T. Paul Schultz 2003). Schultz was able to compare the Progresas program and non-Progresas program localities, which were randomly chosen. Before the program intervention there was no statistically difference between the program group and the control group. The results of the experiment showed a program effect amounting to an average enrollment increase of 3.8 percent for all students in grades 1 through 8, the increase being largest among girls who completed grade 6 (a 14.8 percent increase for them).

4.4 As can be expected, randomized evaluations are not a panacea for sorting out program evaluation problems. They also require careful planning and can be labor intensive and costly, as in the case of a good education production function study. Care must be taken so that there is no spill-over effect from the program group and the control group in a local area. Sometimes there is resistance among the potential beneficiaries, excluding some people from a program merely for the purpose of experimental evaluation. While this can raise ethical questions, such issues of informed consent that arise in ethical protocols in psychology or medical experiments, some researchers who espouse random evaluations feel that these issues can be sorted out.

Natural Experiments

4.5 A good example of a natural experiment type study is one by E. Duflo (2001, AER) for Indonesia, which looks at the impact of government policy in making schooling more widely available. Natural experiments come about when governments

have undertaken some large program initiatives in a way that allows for some controlled comparisons. Duflo examined the outcomes of a large school construction program (and related teacher training) undertaken by the Indonesian government during the 1970s. The particular outcomes examined were changes in primary school completion rates, improvements of the average level of education and the impact on wages (taken as a proxy for increased human capital). This particular study was able to examine the inter-censal survey data years later (1995) to do a statistical analysis of the impact of the school construction program undertaken during the 1970s.

4.6 While the statistical methodology is complex, the results show a positive outcome of an increase of 0.12 to 0.19 years of education for each school constructed per 1,000 children. There was an increase of 1.5% to 2.7% in wages due to the program, with an estimated economic return of 8.8% to 10.6%. Duflo notes that rapid economic growth Indonesia experienced during the previous 20 years contributed to this success, and that such massive investment in school provision pays off when a government expects strong economic growth.

Comparative Benchmarking

4.7 Another approach to the determinants of education outcomes has been taken by Bruns and Mingat (2003). Although their approach appears similar in some ways to a production function approach, is different in fundamental ways. The outcome they examine is not achievement on test scores for a sample of students, but rather the primary school completion rates for a sample of 47 countries. Their empirical approach is to do exploratory correlations and regressions of the inputs and factors that may affect the attainment of universal primary education (UPC) in order to identify patterns among countries that perform the best and worst in this respect. While the study might be criticized for not having an explicit theoretical model to guide it, the authors themselves point out that is not their purpose. They are searching for benchmarks that could be of practical value in formulating a simulation model for estimating the costs of achieving UPC, first country-by-country based upon country demographics and enrollments, and then aggregating the country-level costs to obtain a total cost estimate. This can then be used to see how much international assistance might be needed to supplement national resources in the drive for attaining EFA by 2015.

4.8 Based upon their database of 47 countries, they develop a set of indicative benchmarks for key parameters of the primary education system that is associated with the best performance at making progress toward UPC. Table 2 illustrates the indicative benchmarks they propose.

Table 2: Benchmarks for Primary Education Efficiency and Quality

<i>Variable</i>	<i>Sample Range</i>	<i>Adjusted Sample</i>	<i>Highest-Completion Countries</i>	<i>2015 Benchmarks</i>
Service Delivery Variables				
Average Annual Teacher Salary (multiple of GDP per capita)	0.6 - 9.6	4.0	3.3	3.5
Pupil-Teacher Ratio	13:1 - 79.1:1	44:1	39:1	40:1
Non-Salary Spending (% of Total)	0.1 - 45%	24.4%	26.0%	33%
Average repetition rate (%)	0 - 36%	15.8%	9.5%	10% or lower
Finance Variables				
Government revenues as %GDP	88.0 - 55.7%	19.7%	20.7%	14/16/18%
Education Recurrent Spending (% Government Revenue)	3.2 - 32.6%	17.3%	18.2%	20%
Primary Recurrent Spending (% Education Spending)	26.0 - 66.3%	48.6%	47.6%	50%
Private Enrollment (%Total)	0 - 77.0%	9.4%	7.3%	10%

Source: Bruns et al. (2003).

4.9 A number of caveats are in order about this approach. First, they do not indicate in their data set whether or not a country follows a policy of social promotion within the primary cycle. This would be important variable in differentiating country performance with respect to the primary completion rate (PCR) outcome. Second, there is the difficulty of finding indicators of learning achievement upon completion of primary school, such as national primary leaving exams or international studies such as TIMSS if countries participate. This is important since the authors are interested in EFA of adequate quality and want to use some of the benchmarks as indicators of quality. Third, some of the financial benchmarks, such as allocating not less than 50 percent of the education budget to primary education, must be looked at in light of national education and economic development strategies. For example, the authors remark that India is well below the benchmark for spending on primary education, with 32 percent versus the benchmark of 42 percent prorated for a five-year cycle. However, the Indian pattern of expenditure may not be an ineffective or inefficient one if looked at in the light of its overall economic and education strategy. But such criticism should keep in mind the author's advice to use these indicative benchmarks in a flexible way depending upon each country's context.

Qualitative Methods

4.10 Qualitative methods of evaluation, although they have been used for a long time, should not be forgotten, since they can also shed light on many issues of policy and practice. Economists tend to focus on statistical methods, and even more on those that have some grounding in economic theory. But statistical methods, whether education production functions or randomized evaluations, wind up treating education

much like a “black box” and do not shed much light on some of the critical qualitative details, such as the teaching-learning process in the classroom (Grossman 1994). Yet the qualitative and the quantitative methods need not be seen as mutually exclusive but rather as complements to each other.

4.11 A good example of this is the use of classroom videotaping of contrasting teacher-pupil interactions as part of the TIMSS study. Using standardized protocols to videotape mathematics classroom lessons in Germany, Japan, and the United States, investigators got a much richer idea of what could lie behind the different average results uncovered by the statistical analysis (Stigler and Hiebert 1997). Another example of a more comprehensive qualitative approach is the study by Heneveld and Craig (1996) concerning the “school effectiveness” and “school improvement” literature as it relates to Africa. Their approach is to focus on relevant details of the school and the classroom as the appropriate level where inputs are integrated effectively (or not) into learning activities. The findings of this approach will be discussed further in the sections on curriculum and school management.

5. Specific Inputs and Related Issues from EPF and Alternative Approaches

5.1 The previous two sections give a brief overview of the methodologies involved in education production function studies and alternative approaches. This section is organized around specific issues concerning the determinants of education outcomes so that a more in-depth treatment can be given to each issue from the perspective of different research methodologies and policy implications. Other studies will be cited as each issue is reviewed to look at the merits of the case for various positions on each issue.

5.2 One useful way of categorizing inputs is that used by Harbison and Hanushek (similar to approaches by other authors): (1) “hardware” such as school buildings, classrooms and furniture, sanitation, etc., (2) “software” such as curriculum, pedagogy, textbooks, writing materials, etc., and (3) teachers. A fourth category of management and institutional structure and a fifth category of context and background variables (student academic ability, family and community background, etc.) could be added as well. The management category looks at how allocation decisions are made regarding the first three categories of inputs. The context and background category is often viewed as unalterable background conditions the school must take as given (Fiske 2002 and Velez 1993), though there is evidence that this may not always be the case for some of the “readiness to learn” background conditions. The discussion in this section will follow this categorization as far as possible.

5.3 Before examining each of the specific inputs, this is a good place to assemble a cross-tabulation of results from some review studies on the impact of specific inputs on learning outcomes. This will be useful for the discussion that follows concerning the role of specific inputs.

Table 3: Confirmation Percentages from Various Review Studies

<i>Specific Inputs</i>	<i>Fuller('94)</i>	<i>Harbison/ Hanushek('92)</i>	<i>Velez ('93)</i>
School Facilities		64.7(34)	32.9(70)
Textbooks	73.1(26)		76.5(17)
School Libraries	88.9(18)		
Class Instruction Time	81.8(17)		
Homework Frequency	81.8(11)		
Pupil-Teacher Ratio	34.6(26)	26.7(30)	9.5(43)
Teacher Education	50(18)	55.6(63)	45.6(68)
Teacher Experience	56.5(23)	34.8(46)	40.3(62)
Teacher Salary	36.4(11)	30.8(13)	

Note: First number indicates percentage of cases in which the variable is significantly positive. Second number in parentheses is the number of cases in that review study. Blank cells show indicate data was not available.

5.4 Table 3 used test scores, usually reading or mathematics or some combination, from production function studies. This table is extracted from individual tables in the article by Pritchett (1999) and Hanushek (1995) and consolidated to present the results in a single cross-tabulation. The main purpose is to show the percentage of cases in which a variable was confirmed as being significantly positive (usually at the 5% level) in a multiple regression analysis. However, it should be kept in mind that the underlying quality of each study has not been fully investigated (especially from the point of view of the problems raised by Glewwe 2002) by the authors of the review studies. Thus, there is more or less an acceptance of the published studies at face value. Kremer (1995) also raises the issue of probabilities of making acceptance errors (accepting a false hypothesis) and rejection errors (rejecting a true hypothesis) in statistical hypothesis testing. The fact that some variables have much higher confirmation percentages should be of interest. He also cites the meta-analysis of Hedges, Laine, and Greenwald (1994) on these same studies, in which they conclude there is a positive relation between spending on education and education outcomes, contrary to the Hanushek conclusions. Thus, caveats and debates about statistical uncertainties must be kept in mind when reviewing studies about the impact of individual inputs and education spending in the aggregate.

“Hardware” Issues

5.5 School buildings and classrooms in relation to achievement are shown in Table 2 for two review studies. The Harbison/Hanushek review shows a higher percentage of positive impact than does the Velez review study. However, the Harbison/Hanushek measure of school facilities includes libraries in addition to quality of buildings and classrooms, which may account for the difference with the Velez study. This difference may also be due partly to the fact that Velez reviews many Latin American studies where school facilities are probably better than other lower-income settings. Harbison and Hanushek also note that in comparison to data from the United States, the confirmation rate of school facilities is much higher for developing countries (64.7

percent versus 16 percent). They think that the quality of facilities may be more important in the more disadvantaged settings of developing countries.

5.6 Decisions about the quality of school facilities must also face up to the issue of construction costs. This is especially important in the drive to achieve EFA by 2015. A special study of construction costs in World Bank projects (Theunynck 2002) examined school costs and procurement issues and found that the average cost of a classroom is \$108 per square meter in Asia, \$119 in Africa, and \$187 in Latin America. Other factors come into this, such as the area allocated per student, which is lower when floor mats are used, as in Bangladesh or Mauritania. Theunynck recommends more flexible procurement methods based upon experience with social funds and community-driven projects. Case studies of India (Abadzi 2002 and Pandey 2000) show that community involvement in school design and procurement of school buildings has been a positive contribution to expanding school facilities at reasonable standards of quality. It appears that to go beyond a basic level of building quality does not yield much extra benefit for learning achievement. Yet to allow facilities to deteriorate or use very substandard temporary buildings and classrooms can also hold back learning achievement (Glewwe 2002 and Fiske 2002). The key issue is identifying what school building specifications contribute to learning at an affordable cost.

5.7 The comprehensive case studies of Brazil by Harbison and Hanushek (1992) and of Ghana by White (2004) offer specific evidence that a minimum basic quality of school facilities matters significantly for achievement outcomes. For example, in Ghana, schools would often lose days of instruction due to leaking roofs. Attention to maintenance and repairs dramatically improved the situation to the extent that it shows up as statistically significant in the multiple regression for test scores. The more qualitative literature review by Heneveld and Craig on school quality in Africa also found that a basic level of school facilities contributes to school quality in terms of student learning. A basic standard of school facilities would include enough classrooms to accommodate about 40 students per classroom, sufficient desks in preference to using floor mats, chalk boards, and maybe a storage cupboard for books and materials.

5.8 In addition to classrooms, adequate sanitation in terms of water and latrines is an important aspect of school facilities for increasing the willingness of parents to enroll their girls. In most cases, parents prefer separate toilets for boys and girls. Also, the average distance girls must walk or otherwise travel is a factor, and this becomes more important as the girls enter the upper grades of primary. Of course, reducing the average travel distance implies more schools or classrooms need to be built, but this can also lead to significant reductions in drop-out rates, especially for girls.

5.9 Access to school and enrollment rates in poor regions is critically affected by school and classroom construction, sometimes by adding rooms to existing small schools. White (2004) in the Ghana case studies documents how, despite significant improvements in school construction, in part from Bank support, there are still too many poor communities left behind in terms of quantity and quality of availability of building and places in classrooms. The recent literature on school construction issues (Theunynck 2002 and the White OED study on Ghana 2004) highlights how donors have gone from supporting “bricks and mortar,” then withdrew from that approach to

focus on “software.” They are now coming back to supporting a better balance of “hardware” and “software” inputs, to which the next section turns.

“Software” Issues

5.10 Textbook availability stands out in Table 3 as the input variable with the consistently high confirmation percentage for being positive and significant. This could reflect the scarcity of textbooks relative to other inputs in many developing countries. That would give them a higher marginal product in the production function. However, more needs to be known about the underlying quality of these studies, and about the relative chances of type one and type two errors, i.e., the power of the tests of significance. But as Kremer (1995) pointed out, the probability of getting so many cases of a variable like textbooks being significant, just on the basis of sampling probabilities would be quite low.

5.11 The importance of providing sufficient textbooks, especially when they have been scarce, is documented in the OED Ghana study by White (2004). This study has available test scores in 1988 and again in 2003 in order to assess learning improvements over the intervening 15 years. A rigorous multiple regression analysis showed large gains in reading and mathematics, and improved textbook provision was a significant factor in this. Before the improved provision of textbooks, primary schools in Ghana, which had been among the best in Africa, deteriorated to the point where primary graduates scored no better on simple reading tests than those who had not been to school. Heneveld and Craig (1995) developed a composite picture of the deterioration of primary schooling in Africa in their study “Schools Count.” They point out that textbooks are often not available, and, even when they are, they are not used because they do not get to the classroom for a variety of reasons or teachers are not trained in how to use new textbooks. The result is evident in the poor academic performance of students.

5.12 Curriculum factors are also important, but only a few of the production function or randomized evaluation studies approach them in much detail. For example, curriculum planners in all countries debate the proportion of total instruction time that should be devoted to reading, the total days and hours of instruction to be included in the school year, and complete array of subjects to be covered and the topics within each. Curriculum design also includes decisions about teaching methods and teacher preparation – for example on best ways to teach reading (a controversial topic in the United States and many other countries). It is difficult to do good production function studies on details such as this, since data on teaching methods for a particular subject are not easily available or difficult to generate, even in the most advanced countries.

5.13 A survey of official instruction time (“time on task”) in 110 countries for the 1980s showed the following figures: 870 (low income), 862 (lower middle income), 896 (upper middle income), and 914 (high income). Thus a school year can cover a widely varying number of hours, depending on the official school calendar and patterns of allocating time, a fact not taken into account in many production function

studies that only use years of schooling as inputs. One study of time allocation across countries (Benavot and Kamens 1989) revealed the following patterns (Table 4):

Table 4: Allocation of Curriculum Time to Major Content Areas by Country Income

<i>Content Area</i>	<i>Low Income</i>	<i>Lower Middle</i>	<i>Upper Middle</i>	<i>High Income</i>
Language	37	34	36	34
Mathematics	18	17	18	19
Science	7	9	8	6
Social Studies	8	10	9	9
Moral Educ.	5	6	4	5
Music/Art	9	8	11	13
Physical Educ.	7	6	7	9
Hygiene	1	2	2	1
Vocational	6	7	3	1
Other	3	3	2	1

Source: Benavot and Kamens (1989) as in Bank Primary Education Policy Paper (1990)

What is striking about these official figures is that there is not that much variation from low-income to high-income countries. The core subjects of language and mathematics take up about 50-55% of the instructional time and science and social studies about another 20%, leaving approximately 25-30% for the other subject areas.

5.14 Of particular interest should be the difference between the hours of instruction officially mandated and what actually happens in the classroom. Of course, it is the actual instructional time on task that matters for educational outcomes. The studies reviewed by Fuller (1994) show that classroom instruction time and homework frequency have the highest confirmation rates (see Table 3). Benavot (2003) has done a review of the literature on factors affecting actual instructional time on task versus what is specified in the official curriculum documents. He concludes from his review of the literature in developed countries, especially for schools in impoverished environments, that there is evidence for believing that increasing instructional time on task would improve learning achievement. His review of studies for developing countries shows that there is a large difference between official and the actual instructional time in the classroom. Although there is much variation by context, the overall reduction in time on task is fairly large, perhaps as much as 30-50% is his rough guess. The reasons for this can vary, but teacher absenteeism is one major factor. Illness, especially debilitating diseases like HIV/AIDS can also be a major cause. In some other studies there is also a high rate of tardiness among teachers. One study suggested that centralized education bureaucracy inherited from colonial times might be a factor in teacher absenteeism since headmasters have little leverage or sanctions to use against staff.

5.15 Benavot also points out how instructional time on task is related to other factors often identified as determinants of education outcomes. Class size is related to time on task insofar as many educators believe that smaller classes allow for more

time and attention to each student. Adequately maintained school buildings can help avoid loss of time on task. A striking example is that of leaking roofs during heavy rains that were shown in the Ghana study to result in lower achievement (Glewwe 2002, and White 2004). Availability of sufficient textbooks can reduce time on task since the teacher then has to resort to writing much of the material on the chalk board for students to see and/or copy.

5.16 Other kinds of education evaluation methods, often more qualitative in nature, are needed to address these questions, because they are germane to the issues underlying the determinants of educational outcomes. Reading and mathematics form the core of basic knowledge skills, being the gateway to developing reasoning and problem-solving skills in other specific subject areas. The Bank has increasingly sponsored some research into these areas, especially most recently in the area of reading skills. As mentioned above, it is very difficult to find production function studies that can incorporate sufficient details of the curriculum, such as what methods are used to teach reading. Nor is it necessary to rely upon a production function approach here, as experimental and quasi-experimental methods are feasible and appropriate. Even then, it would be hard to generalize across countries because strategies for teaching reading may vary by language. For example, the debate in the United States over teaching reading via phonics versus whole-language approaches may not be relevant to countries with a much more phonetic language, like Kiswahili. Yet it would be beneficial to know about techniques for improving the productivity of teaching reading. Many studies show that there is a benefit in terms of higher wages and productivity for improved reading, even if someone has completed only a primary level of education, but has learned to read well (see Glewwe 2002 and Knight and Sabot 1990).

5.17 In some developing countries the colonial languages are the main ones of instruction. In other countries with diverse ethnic groups, the languages of the largest or dominant groups also might be used. Thus, there has been a debate over teaching reading and mathematics in a foreign language versus the child's mother language used at home. There is evidence to suggest that young children do better at mastering concepts of reading and notions of numbers, shapes, and arithmetic in their own local languages. The OED report on Ghana (2003) finds some supporting evidence that this is the case. The UNESCO report by Fiske (2000) showed in the PASEC study that children do better when they speak the language of instruction at home.

5.18 The language of instruction matters not only for curriculum but also for assessment, especially in terms of public examinations. Kelleghan and Greaney (1992) point out that it is not unusual for children to be instructed and examined in a language that is not their mother tongue. This is a factor in poor performance on examinations in countries like Madagascar, where there is extensive use of French in the education system, or in Mauritius, where exams are in English but Creole is the lingua franca. The language of instruction and assessment can also be a sensitive political issue in many

countries. Benavot cites the examples of Albanian minorities in Macedonia, and the Russian minority in Latvia, but there are many other such examples.³

5.19 Alternative methods of delivering the curriculum when school buildings are scarce include double-shifting and multi-grade teaching methods. Double-shifting involves using the same classrooms for a morning and afternoon shift to accommodate more students. Sometimes the same teacher covers both shifts, but there is evidence to suggest that it is better to have different teachers for each shift (Verspoor 2003). It may be that teacher fatigue and morale could be factors here, since the teacher is not likely to be compensated that much extra, if at all, for the greater work load. Multi-grade teaching would combine different grade levels (say grade 1 to 3) into a single class under a single teacher. This can be cost effective in rural situations where the school population is relatively small. However, it also requires special training for the teacher to learn effective classroom management techniques, some of which can involve using older students to help younger ones. But the evidence on this is still mixed, as seen in the PASEC studies from West Africa (Fiske 2000). Jarousse and Mingat (1999) found that multi-grade teaching in Togo had positive effects, while the PASEC study found that there was a negative effect (except in Senegal). It is not clear how much this is related to lack of appropriate training for teachers in multi-grade pedagogy.

Teachers

5.20 The issue of class size or pupil-teacher ratios (PTRs) is related to the quantitative aspect of how many teachers to hire. For a given student population size, smaller class sizes means more teachers need to be hired. However, it should be noted that class size and PTRs are not necessarily the same measures, except perhaps in special circumstances. Reducing class size is the most frequent suggestion made for improving the quality of education, but it is a costly strategy. Yet the research results in this area are among the most controversial, despite many studies and reviews of studies from time to time. Hanushek and others have reviewed such studies a number of times over the past 10 to 15 years, and concluded that the evidence does not support widespread across-the-board decreases in class size (see Table 3).

5.21 Many educators relate the class size argument to that of “time-on-task” issues, claiming smaller class size minimizes disruption and allows teachers to give more individual attention to students, thereby increasing the effective time of instruction. But some research studies of actual classroom teaching practices show that teachers often do not change their method of teaching in response to a smaller class size. Instead, they still lecture and go about assigning homework in much the same way. This may explain why the gains from class size reduction is more apparent in the early grades, where teachers tend to use small groups, hands-on projects, and personal relationships with their students. On the other hand, other researchers cite the examples of Asian schools

3. Not only the language of instruction and public examinations, but also the content of subjects like history can be very sensitive politically in some countries. This should be kept in mind when analyzing the role of primary education in promoting social cohesiveness and national identity, an important topic, but outside the scope of this review.

where class sizes are larger, but achievement is better. In some cases, in Japan, teachers will trade-off increased class size for more preparation time (Ehrenberg et al. 2001).

5.22 The World Development Report 2004 had as its annual theme “Delivering Services to the Poor,” which included the topic of improving education services for the poor, as well as other services, such as health. In that context, the WDR 2004 also reviewed the class size debate worldwide and concluded that the uncertainty of research results over such a seemingly simple issue illustrates how truly complex the research question is, with the results varying across time, context, and content. Still, the overall conclusion was that a policy of promoting relatively small class sizes (below 40 student per teacher) is not cost-effective in developing countries, compared to providing more textbooks, increasing the total hours of class instruction over the year, or restructuring overcrowded curricula, so that more time is spent in class on the core subjects of reading and mathematics. Of course, excessively large classes (above 60 student per teacher) are also unacceptable, since they are detrimental to learning (Heneveld and Craig 1995).

5.23 In addition to teacher distribution, other more qualitative dimensions of teacher provision have been investigated for their relative importance. Studies done in the 1980s indicate that teachers’ education and certification have a positive association with children’s score on achievement tests (Heyneman and Loxley 1983). Behrman and Birdsall (1983) used teacher qualifications as an indicator of school quality and showed that it affected the rate of return in the labor market in Brazil. Other studies have looked at teacher qualifications in terms of levels of education, such as university degree versus lesser education levels. Hanushek (1994) in his review of studies reported that 35 of the 63 studies give a significant positive effect for teacher’s education, 26 no significant effect, and 2 significantly negative. This contrasts with the situation in the United States where teacher education usually does not show any impact, leading Hanushek to suggest that these results reflect differences in stages of development.

5.24 Hanushek’s review finds no compelling support for the belief that higher salaries would lead to better quality teachers. On the other hand, there are many individual reports of cases where excessively low teacher salaries can have a negative effect. Excessively low salary can be defined by comparison to similar occupations that are likely to compete for teachers, and the cost of living. If teachers find it difficult to maintain their living standard, as has been documented in some cases, the results can be absenteeism and low morale on their part while they pursue second jobs, leading to declining student performance (Filmer and Lieberman 2002, for an example in Indonesia).

5.25 However, many studies find that teacher training is important. Some of these studies find the pre-service education and training matter, while others seem to indicate that in-service training could be more effective. The evidence is somewhat complicated and mixed, no doubt varying by the quality of research design and data, but the overall suggestion is that better trained teachers are more effective in terms of cognitive achievement. However, it does not really appear necessary to have university educated teachers in primary schools in low-income countries. Nor is it

necessary to have a very long duration teacher training college (more than 2 years) course after secondary. For the early grades of primary, it is often sufficient and more cost-effective to give a one-year pre-service coupled with well designed in-service follow-up and support (Verspoor 2003).

5.26 Three key issues for teacher effectiveness are identified in the Bank 1990 policy paper based upon an extensive literature review: (1) knowledge of subject matter, (2) pedagogical skills, and (3) teacher motivation, of which salary is only one part. Harbison and Hanushek (1992) and Kingdon (1996) found that tests measuring teachers' knowledge of mathematics are significant determinants of the achievement of students. Education theory and philosophy suggest that teachers who are skilled in active-child-centered methods of teaching produce better learning results, especially when it comes to the capacity of students to apply knowledge as opposed to just memorizing facts and names of concepts. Abadzi (2002) and Pandey (2000) cite the use of "joyful learning" methods as an example of the child-centered approach and there is some limited evidence to believe that it is working. However, there are not many rigorous statistical studies in developing countries to confirm that active learning approaches have had large positive effects, although most educators seem to subscribe to some variant of this approach. Finally, apart from salary, low teacher morale is due to poor working conditions and lack of administrative and community support (Heneveld and Craig 1995), much of which could be rectified without significant financial expenditures.

Management and Institutional Structure

5.27 In addition to school resource issues, management and institutional features of the school system have been investigated for their potential impact. Introduction of competition via private schools and vouchers has been a topic of interest in a number of studies. Angrist et al. (2002) conducted a randomized evaluation of a voucher program in Colombia in which vouchers for private schools were allocated by lottery. Lottery winners were 15-20 percent more likely to attend private school and scored 0.2 standard deviations (equivalent to a full grade level) higher on standardized tests. The program effects were substantially higher for girls, lottery winners were more likely to complete high school and scored higher high school completion exams/college entrance exams. The cost of the program was similar to that of providing a public school place, so the conclusion is that this program is clearly more cost-effective.

5.28 Wossman also investigated this issue by using the TIMSS production function study to show a positive impact on achievement by the degree of choice available as measured by share of students enrolled in privately managed schools. In both mathematics and science, the average student score was significantly higher when there was a larger share of students in privately managed schools. Privately managed schools could encompass those like the Dutch that received public funding or those like the Japanese that rely less on public funding. The test scores for private versus public were not available for each country, but Wossman maintains that what is of interest is the impact on the whole system

5.29 There are a number of critical reviews of vouchers and privatization that contrast with the positive results above. Ladd (2002) argues that based upon the limited U.S. evidence to date, vouchers are not likely to lead to significant gains in learning achievement or in the productivity of the education system. McEwan and Carnoy (2000) reach similar conclusions about the voucher programs in Chile. However, Ladd admits that vouchers could permit low-income families access to a type of schooling they might not have available under non-voucher regimes, and for that reason targeted voucher programs could contribute to social mobility.

5.30 Since voucher use typically allows for enrollment in private schools, there are inevitably questions about the use of public money for private education. It is often difficult to disentangle issues of political philosophy from those of educational effectiveness and economic efficiency. As Ladd (2002) points out, even in studies that try to compare the relative effectiveness of public and private schools, the statistical problem of self-selection is often not addressed and therefore confounds attempts to draw policy conclusions.

5.31 The presence or absence of a central examination⁴ system is an institutional feature that Wossman was able to investigate using the TIMSS database. Some of the countries in the study have such examination systems, and the results of his production function study show that this has a significant positive impact on achievement scores. Wossman makes the interesting point that the magnitude of this impact in the TIMSS database is about three-quarters of that derived by Krueger in his analysis of the impact of Project STAR in Tennessee. However, he points out that the cost of implementing a central exam system is much less than that of implementing class size reductions on the scale of Project STAR. This could be a topic to investigate in some developing countries where central exams play a prominent role after the primary cycle as a selection mechanism for continuation to secondary education. Also worth examination are national assessments using samples of schools, which are used for assessing system performance and not student placement (thus called low-stakes exams). Such examinations could have the same positive affect on performance without the negative affects of a high-stakes exam (Greaney and Kellaghan 1996).

5.32 Decentralization and school autonomy are other features of the institutional structure of education systems that have received much attention in recent years. Wossman found that central control of curriculum and textbooks was positively correlated with mathematics and science results, while autonomy at the school level in formulating schools budgets and hiring teachers were also positively correlated with test scores. His interpretation of these results is that the distribution of

4. Here Wossman is referring to “high-stakes” examinations as it is termed by some (Greaney and Kellaghan 1996). This refers to examinations used for selection to secondary or university education or certification that can affect job prospects. The reform of such exams is the subject of controversy in many countries that use them extensively, both developing and developed. By way of contrast, the TIMSS exams, PASEC, etc. referred to earlier in this paper are assessment exams or “low stakes” in terms of consequences for individual students. Each approach serves a fundamentally different purpose, which should be kept in mind when trying to evaluate the determinants of education outcomes.

responsibilities is optimal when central administration has control of areas where school-level opportunistic behavior should be limited and schools have autonomy in those areas where school-level knowledge is important.

5.33 This conclusion fits with the belief that school-based management (or site-based management) provides an organizational approach that gets the distribution of management functions right. The Bank has been promoting decentralization in education for some years (see the policy study “Priorities and Strategies for Education” 1995). As part of a review of using school grants to promote greater school-based management in the context of Bank projects, Roberts-Schweitzer, et al. (2002) conclude that school grants could be a promising approach if carefully done. Roberts-Schweitzer et al. cite a review of SBM studies by Littlewood and Menzies (1998), which found that only 11 of 83 studies reviewed could identify actual improvements in teaching and learning outcomes, although other beneficial results were identified, such as increased interest and financing from the local community. The Roberts-Schweitzer et al. review indicates that the most positive effects on students are reported when the local community can directly influence school management, but it does not specify what these positive effects are. Part of the difficulty in such studies is isolating the influences of so many factors, and it appears a randomized evaluation approach could be applied in such situations.

5.34 As pointed out in the Roberts-Schweitzer et al. review, these SBM initiatives must be done carefully. It is not easy to figure out exactly what functions to delegate down to the lowest level, and this may vary by country and culture. Fiske and Ladd (2000) did a study of New Zealand’s education reforms, which involved decentralization to the point of public schools becoming substantially self-governing under their own school boards with the Ministry of Education setting overall policy and providing formula funding for which schools would be accountable. Moreover, parents were given choice as to where to enroll their children and schools could compete in substantial ways for students. A similar decentralization down to the school level was tried in Chicago and has been explored in other U.S. school districts. Although the jury is still out on the impact on learning achievement, it appears that arrangements need to be made for dysfunctional or failing schools. Sometimes decentralization can lead to “capture” of service provision by local elites or politically active groups to the detriment of the participation of the wider local community (World Development Report 2004); in other cases it could lead to increases in regional disparities in resources and outcomes as was found to some extent in a careful OED evaluation study in Ghana (White 2004). So while the potential benefits seem large, the caveats about careful implementation of decentralization schemes, school choice, and school-based management seem warranted.

5.35 In addition to reforms involving decentralization and vouchers, there is the beginning of a focus on systemic change of education systems. The importance of these larger systemic issues was raised in the Bank’s *Priorities and Strategies for Education* (1995). The ECA Education Sector Strategy Paper (2001) also provides an example of this where the larger governance and management issues are confronted with a view as to how they affect incentives and behavior. The ECA strategy paper draws upon the analysis of the “new institutional economics” (Williamson 2000),

which goes beyond the traditional microeconomics of getting the marginal conditions right and examines the larger governance and institutional context within which optimization of resource use takes place. Although the ECA strategy was done in the context of transition economies and their education systems, which need realign their systems to incorporate “new rules of the game,” a case could be made that almost all developing countries (and even developed ones) are facing such pressures to realign their systems.

5.36 Based upon the studies reviewed, it would be appropriate to consider the use of private schools, vouchers, and decentralization even as strategies benefiting the poor. However, the success depends greatly upon the political context of a country’s education system and the institutional history of education in a given country. In the context of low-income countries struggling to achieve EFA and improve quality, there are not many rigorous studies to draw upon at the primary education level, so care must be used in extrapolating results of studies from high- and middle-income countries.

Contextual and Background Factors

5.37 There are many contextual factors outside the school that education policy analysts examine, the most typical being individual student characteristics, family background, and community characteristics. The most frequently mentioned student characteristic is innate intelligence or IQ, and studies that have some measure of this usually find it to be a significant factor in learning achievement (Boissiere et al. 1985, Glewwe 2002, Kingdon 1996, White 2004). There is now a large literature in the developed countries about broader notions of intelligence. This debate about different types of intelligence among educators and psychologists still includes some version of innate components corresponding to ability to learn reading and mathematics (which is typically at least 50% of a primary school curriculum). However, many studies take this as a given, outside the influence of school interventions and government policy, which may not fully be the case, as shown by health and nutrition studies. Moreover, in the Ghana studies of Glewwe and White, the Raven’s Progressive Matrices Test (used as a measure of intelligence in the studies cited above) is shown to be influenced by home and school to some extent. Nonetheless, schooling still has the predominant influence on learning achievement in these studies.

5.38 Health and nutrition status can be both an input factor as well as an outcome of schooling. First, as an input factor, it has been documented that illness of various sorts, for example, malaria in tropical countries, can cause absenteeism as well as reduced energy levels in class. Lack of nutrition at home can lead to poor performance, even if attendance is regular. In addition, there are the various physical and mental disabilities that occur in all societies. It has been shown that reduced learning capacity can result from poor health and nutrition due to poverty. Flynn (1987) has documented the rise of IQ scores in the high-income countries over time, probably due in large part to improved health and nutrition. Some demand-side programs, like Progreso in Mexico or Bolsa Escola in Brazil, have addressed such issues within their incentives for school attendance. In addition, there are school feeding and health programs (such as deworming) and various approaches to special education.

5.39 With respect to health and nutrition outcomes of schooling, Glewwe (2002) reviews a number of studies that present strong evidence that mother's education can influence the health and nutrition status of children in the household (Behrman 1990). Thus intergenerational models can foresee that health and nutrition status of children in the not so distant future can be improved by ensuring that girls everywhere, in rural and urban areas, high- and low-income households, receive at least a complete primary education of adequate quality. This in turn would affect the learning capacity of the next generation.

5.40 SES variables, such as education of the parents and income level of the family (SES variables) have been mentioned before in the context of the Coleman report (1966) done in the United States, where it was concluded that SES accounts for more of the variation in achievement than does school inputs. However, in their Brazil study, Harbison and Hanushek concluded that on balance their results do not support such an overwhelming importance of SES variables in the educational performance of poor pupils. Similar results are found in East Africa (Armitage and Sabot 1987) and also in Ghana (White 2004). It is not that SES does not matter at all, but that the contribution of schooling is much larger for developing countries than is the case for developed ones.

6. Can Any Firm Conclusions Be Drawn?

6.1 Are there any firm conclusions that can be drawn from the continued effort of research to find which inputs or factors, or combinations of thereof, are most effective in determining primary education outcomes? Compared to the state of research at the time of the last Bank Primary Education Policy Paper (1990), what new things have we learned about the determinants of primary education outcomes? Overall, one could conclude that the recommendations of the 1990 policy paper have stood the test of time in terms of the research results now available since 1990. The priorities for investment identified there still seem valid. However, firm conclusions specified in great detail across many countries are not really feasible, since the research literature shows how much policy interventions depend upon context and history. Still, a few general patterns can be discerned that offer some promising pathways to examine in the light of local conditions.

6.2 One of the general patterns discerned in the literature is related to the initial controversy generated by the Coleman Report. In contrast to the Coleman Report conclusion that schools are less important than SES in determining education outcomes, many researchers conclude from the developing country evidence that schools matter much more in the setting of poor countries. A large part of the reason behind this appears to be the principle of diminishing returns — in the setting of poor countries providing more education resources has a larger impact than in a rich country where school resources are already at a relatively high level. This was suggested well before 1990, but the research literature since 1990 has given more weight to that conclusion.

6.3 A second important pattern is that the outcomes of primary education in poor countries are far below those of the rich countries. There is now much more known about the basic cognitive skills of literacy and numeracy in rich and poor countries based upon the growing body of evidence from international and regional education assessments. The gap between the rich and the poor countries on these measures is large, and the distribution of this most basic form of human capital within the poor countries also seems to be more inequitable (though more research is needed here).

6.4 The third general lesson that emerges from the evidence is that government policy and implementation capacity is important, especially for determining the provision of schools and equity of access. Although private schools spring up under many circumstances and make an important contribution, equitable access to quality primary education for many specific groups — girls, the poor, distant rural communities — depends crucially upon good government policy and implementation. These three patterns taken together suggest that it is quite feasible to make substantial progress in the outcomes of primary education in developing countries.

6.5 When it comes to more specific factors related to education outcomes, the most striking area of agreement found in studies using a variety of research methodologies is that textbooks and learning materials show the highest incidence of impact for improving primary school outcomes in many developing countries. This may reflect the fact the developing countries, due to budget constraints, have been under-funding such resources for a long time, leading to their relative scarcity. Hanushek (1995) and Kremer (1995), despite their acknowledged differences of interpretation on other issues, agree on the textbook issue. Not only production function studies, but random evaluations also support the influence of textbooks and learning materials. The OED education sector evaluations in Ghana and India support the importance of interventions for learning materials..

6.6 However, the Glewwe and Kremer randomized study in Kenya provides some caveats. The learning materials must be appropriately designed so as not to be too difficult for the typical rural primary school, and teachers must be trained in conjunction with the introduction of learning materials. Thus, local conditions must be factored into the design of learning materials.

6.7 The interesting feature of improving learning materials is that this can be done at a unit cost that is reasonable compared to GDP per capita and cost per student and the improvement in efficiency of education can be substantial (Pritchett and Filmer 1999). Providing learning materials could receive priority over reducing class size once class sizes have reached a level of about 40 students. Yet there are too many cases, especially in Sub-Saharan Africa, where class sizes are well above 50 students, a situation detrimental to good learning outcomes. In situations where rapid expansion of access is needed to achieve EFA/FTI goals, learning materials and related teacher training should not be sacrificed to reach numerical enrollment targets. In many sectors there is a trade-off between rapid expansion and quality of outputs, but the unprecedented pace of politically-driven basic education expansion in some countries makes this trade-off a critical issue. As the experience of Indonesia has shown, going

back to try to retrofit quality after rapid expansion with low quality is difficult (see Filmer and Lieberman 2002), and likely to be more expensive in the end.

6.8 In addition to the evidence on learning materials, there is also evidence supporting a better balance between “hardware” and “software” investments. The contribution of adequate school facilities has also been shown in many studies to be an important determinant (Harbison and Hanushek and OED evaluations in Ghana and India). In addition, adequate school facilities have been shown to be possible at relatively low-cost relative if local materials and procedures are used, and expensive buildings are avoided. Much has been learned about the more detailed qualitative and quantitative characteristics of adequate school buildings and furnishings, and also about methods of procurement and implementation of large-scale construction and renovation programs (Bruns et al. and Theunyck). These lessons are especially important for EFA and FTI.

6.9 Teacher quality and salaries are inter-related issues that could provide improvements if done appropriately. Getting teacher salaries right requires examining the labor market as a whole in a given country and setting a teacher salary scale that makes sense in the overall labor market context. This could help avoid situations where primary teacher salaries are too low with the result in absenteeism and second jobs that compromise teacher effort in the classroom. If teacher salaries are excessively large relative to the overall labor market and other relevant occupations, the cost of expansion becomes prohibitive. A policy of bringing them in line over time, difficult as that might be, would also result in significant reductions in cost per student without impeding teacher effectiveness.

6.10 Improving performance incentives is another area where there is broad agreement in the literature that positive gains in education outcomes could be achieved. In part this reflects the disillusionment with the results of simply providing more resources. How to structure effective performance incentives is not altogether clear, though some promising interventions are discussed. There is some evidence that more decentralization and school autonomy (school-based management) would more effective than the currently inefficient bureaucratic structures.

6.11 Also, a number of studies indicate that demand-side programs that subsidize or reduce the costs of attending primary promote greater enrollment and, ultimately, completion. This appears to have the most impact in the poorest countries and in the poorest areas of middle-income countries. Demand-side programs also appear effective in promoting gender equity in education access.

6.12 With respect to research methodology, many lessons have been learned that should be incorporated into future research to get a better picture of more effective interventions. Despite the caveats of the best researchers, other equally respected researchers appear to believe that well designed education production function studies should continue. Some prominent researchers suggest that randomized evaluations and the search for natural experiments should receive more encouragement and resources, especially for those issues, like demand-side financing or vouchers, where they are more suitable than production function approaches or shed a different kind of light on

the same issue. Finally, the literature review shows that good qualitative research, especially at the school and management level, can provide very practical insights.

6.13 There is no reason why research methods cannot be combined in studying education in a given country, or even the same region a country. EPF methods and random evaluations can both be used as cross checks. However, it is worth remembering that these statistical methods will always have some inherent uncertainties that result from drawing probability samples and trying to make inferences about the population. Despite Glewwe's critique of the four major econometric problems faced by EPFs (omitted variables, endogenous variables, measurement errors, and sample selectivity), random evaluations also have their statistical difficulties and pitfalls (such as contamination of control groups). Finally, Pritchett's caveat about collecting relevant and reliable education cost data along with EPF and random evaluations should be kept in mind. Otherwise, good cost-benefit analysis would not be possible.

6.14 Countries need to build up their local research capacity as they experiment and innovate, as is being done under the EFA initiatives supported by the Bank, UNESCO, and others. Although international comparisons have their uses, each country needs to analyze its own experiences with a wide spectrum of research techniques ranging from the quantitative to the qualitative ends of the spectrum. Each method has its advantages, and their results should be integrated to form a complete picture of a country's situation with respect to quality primary education. While such a research and evaluation enterprise at the country level may seem expensive, the cost is relatively small compared to the resources spent on education. Moreover, the cost of ignorance about the outcomes of education interventions is likely to be more expensive.

Annex A: Working Definitions of Inputs and Outcomes

Inputs

Access. Access to primary education refers to households having a school with places available to its children within a reasonable distance and cost. It is a behavioral matter, subject to social science analysis, whether or not households choose to enroll their children in the available schools.

Gross enrollment rate. The usual definition takes the ratio of primary enrollment of all ages to the primary enrollment of those in the official primary age group, typically 6 to 11 years old. As a result of under-age and over-age enrollments and high repetition rates, this ratio could be greater than 100%. While this ratio has a number of shortcomings, it does give some indication of the capacity of the school system.

Net enrollment rate. This is the ratio of the number of students enrolled of official primary school age to the total population of official primary-age children. This should be 100% or less. This ratio is considered to be a better indicator of primary school access than the gross enrollment ratio.

Primary cohort progression rate. This is the ratio of those graduates of the last year of the primary cycle who entered primary year 1, 5, or 6 years ago depending upon the length of the cycle. This is different from Primary Completion Rate defined below in that it only looks at those who actually entered primary, which clearly may not include all age-eligible children. Still, it gives some idea of the efficiency of the system for those children who do enter.

Outcomes

Primary completion rate. This is the ratio of primary graduates in a given year to the total number of children of official graduation age, which can vary by country according to definitions of the primary school cycle and official age of entry. This definition may also vary from study to study under review. The EFA/FTI initiatives are now formulated in terms of primary completion rates, whereas originally they were formulated in terms of net enrollment rates.

Where primary completion rates are available, they would be preferred over net or gross enrollment rates, since these are not really outcomes, but rather process indicators. Assessments must also be done concerning the curriculum standards, resources available, and promotion policies (whether or not there is generous social promotion). The idea is to ascertain if completion rates in a given study or data set can imply that adequate learning has been associated with primary completion, and hence the quality of primary schooling. Improving net enrollment rates together with primary completion rates, with improving balance by gender and urban/rural areas, can give a reasonably good indicator of an improving school system.

Learning achievement. Actual learning achievement in the core curriculum (reading, writing, mathematics, and science) as measured in reliable tests that have some appropriate international comparisons would be the ideal outcome measure to combine with net enrollment rates or primary completion rates in making judgments about a school system. This would give an almost complete picture of the quality of primary education. Some countries have participated in TIMSS and TIMSS-R, and others have a time series of their national exams at the end of the primary cycle. Others have done special testing studies to examine learning achievement.

Average years of education of the adult population and labor force. This could also be used as an indicator over time in conjunction with enrollment (gross and net) and primary completion rates. If a primary education system is making rapid progress toward UPE, this should be reflected in the census data about the education of the population, or preferably the labor force if available either directly or by estimation. Times series for this indicator is increasingly available for many countries over the past 20 to 30 years, including the younger cohorts. Of course, secondary and tertiary enrollment would also increase this indicator, but the enrollment and completion rates for secondary and tertiary can give some idea of their separate contribution.

Earnings and productivity. Many micro-economic studies use labor market earnings as a proxy for productivity, since productivity is difficult to measure directly for individual workers. This literature review will accept that approach, noting along the way some of the debates economists have about that. However, this review will not devote much time to studies that analyze the macro-economic impact of primary education on productivity, except to note in the appropriate places some of the controversy surrounding these macro-economic studies. In general, the micro-economic studies of earnings and education have better data sets to support them than the macro-economic studies of the impact of education on economic growth and productivity.

Annex B: Econometric Issues Related to Education Production Function

1. In Glewwe's (2002) review article referred to in the main text (para. 16), he notes that a linear production function, which is often used in empirical work, can be derived from a non-linear production function, such as a Cobb-Douglas type, where $H = k * S^x * A^y * Q^z$ and H is human capital, S is schooling, A is ability, and Q is school quality. A constant factor is represented by k, and x, y, and z represent exponents of the variables S, A, and Q. If the exponents are constrained to have values between 0 and 1, such a non-linear function would give diminishing returns to more inputs, which is not the case with a linear function. In the education context, this would mean that there could be a high marginal return to more textbooks when they are scarce relative to other inputs. However, once there are adequate textbooks per student in a given subject area, the marginal returns would decrease.

2. If $x+y+z=1$, there would be constant returns to scale for all inputs if increased in the same proportion, as opposed to varying them individually. If $x+y+z<1$, there would be decreasing return to scale, and increasing returns to scale when $x+y+z>1$. Also, there are other more complex mathematical forms of production functions, but they appear not to have been used in education studies, probably because the statistical problems involved are not tractable. A Cobb-Douglas type form is perhaps sufficient to capture the phenomena of non-linearity and diminishing returns. Another approach is the use of dummy variables for some specifications to capture non-linearity, especially for variables like pupil-teacher ratios or levels of teacher education and training.

3. Taking logarithms of both sides of this Cobb-Douglas equation just above would give a linear equation in all of the logs of the variables, referred to as log-linear form. In this case, the coefficients estimated are actually the exponents x, y, and z, which can be used to calculate the impact effect of increasing the inputs. However, many studies use test scores and other variables directly, instead of using the logarithms of these variables. In this case, the interpretation is that a linear function is a reasonable approximation to a non-linear function over the range of the data variables being used. The issue of linearity versus non-linearity is not just an arcane mathematical debate. It relates to the practical policy issue of when and where diminishing returns could be expected to set in and therefore what would be the relative cost-effectiveness and cost-benefits of various interventions.

4. Thus a linear form of the production function is used to provide a convenient framework to clarify the issues surrounding multiple regression analyses of cognitive achievement tests:

$$H = c + s * S + a_1 * A_1 + a_2 * A_2 + \dots + a_n * A_n + q_1 * Q_1 + q_2 * Q_2 \dots + q_m * Q_m + u$$

H is human capital using a measure of knowledge, such as achievement test scores,
S is schooling (usually years of schooling),

A_i 's represent a series of individual student ability and learning capacities, such as IQ,
 Q_i 's represent school quality factors, such as class size, teacher qualifications, etc.

The lower case c is the constant term, and s , a_1 to a_n , q_1 to q_m represent the coefficients multiplying the variables S , A_i 's, and Q_i 's.

The variable u represents the random disturbance term due in part to the unpredictability of human behavior. In addition, it may include omitted variables and measurement errors in the variables.

5. Taking the problem of omitted variables first, it is difficult to collect data on all aspects of quality of a school or of the learning ability of children. Many studies do not have IQ-type scores, and, even if they do, there is still debate about how much learning ability is captured by IQ scores. However, econometric theory shows that omitting an explanatory variable in the simple ordinary least squares (OLS) regression would produce biased estimates of the coefficients of the other variables. The basic reason for this is that omitting an explanatory variable can cause a correlation of the explanatory variables with the random disturbance term. This violates one of the key assumptions needed for OLS multiple regression to give unbiased coefficient estimates, namely that the explanatory variables should not be correlated with the random disturbance term. Hence, it is important to try to obtain measures of ability as has been done in some education production function studies. When ability measures have been included, it has been shown that ability is a positive and significant determinant of learning achievement.

6. Glewwe also cites other econometric problems that could lead to correlation of an explanatory variable with the random disturbance term, and thus give biased estimates of coefficients. For example, endogenous school quality means that the quality of school could be influenced or chosen by the household, that is, it is a variable of choice in a larger behavioral model with other equations. For example, studies in a number of African countries show that parents often send their children to live with relatives so that the children can attend better schools. This is somewhat similar to parents in the United States buying a house in a school district with a reputation for quality schools. Such a situation could lead to the explanatory variables representing quality being correlated with the random disturbance term. Glewwe presented a simplified model of household choice to illustrate how the education production function can be derived in such a setting and how quality of schooling can be a variable of choice. Glewwe also points out that this means that all of the coefficient estimates could be biased, not just the coefficient for the variable correlated with the disturbance term.

7. The problem of measurement error in the explanatory variables could result from random errors, such as respondents in interviews recalling information incorrectly, or nonrandom errors, such as poor wording in the sample survey questionnaire leading to erroneous responses. This could lead to overestimation or underestimation of the parameters of interest, depending upon whether or not the measurement error is random or non-random.

8. Sample selectivity problems can result from how the sample is chosen, which may not be truly random due to the behavior of the people determining whether or not they are in the sample. For example, not all children attend schools in low-income

countries and the behavior of households could affect whether or not a child is in a sample of pupils. Again, this could lead to a correlation between the explanatory variables used in the multiple regression and the random disturbance term, potentially resulting in biased coefficient estimates.

9. As noted in the text, Glewwe concludes that these are difficult problems to overcome. However, he notes that careful planning in the early stages of the research studies can go a long way in addressing these issues. Each of the four good conventional studies (for Brazil, Ghana, Jamaica, and India) reviewed in detail by Glewwe addressed some of these issues. He notes that, if good conventional studies agree on the significance of an input, there is likely to be a causal relationship. The problem is that there are too few conventional studies, a problem that can be rectified as a result of research in the future.

Bibliography

- Abadzi, Helen. 2002. *India: Education Sector Development in the 1990s, OED Country Assistance Evaluation*, Operations Evaluation Department, World Bank, Washington, D.C.
- Angrist, Joshua, Eric Betting, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92(5):1535-58.
- Armitage, Jane, and Richard H. Sabot. 1987. "Socio-Economic Background and the Returns to Schooling in Two Low-Income Countries." *Economica* 54(213):103-08.
- Bruns, Barbara, Alain Mingat, and Ramahatra Rakotomalala. 2003. "Achieving Universal Primary Education by 2015: A Chance for Every Child." World Bank Policy Study.
- Baumol, W.J. 1967. "Macroeconomics of Unbalanced Growth: the Anatomy of Urban Crisis." *American Economic Review* 57:415-426.
- Behrman, Jere. 1990. "The Action of Human Resources and Poverty on One Another: What Have We Yet to Learn." Living Standards Measurement Study Working Paper No. 74. World Bank.
- Boissiere, Maurice, J.B.Knight, and R.H. Sabot. 1985. "Earnings, Schooling, Ability and Cognitive Skills." *American Economic Review* 75(5):1016-30.
- Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91:4, pp 795-813.
- Ehrenberg, Ronald, Dominic Brewer, Adam Gamoran, and J. Douglas Willms. 2001. "Does Class Size Matter?" *Scientific American* 285(5).
- Filmer, Deon. 2003. "Determinants of Health and Education Outcomes," Background Paper for World Development Report 2003. World Bank.
- Fiske, Edward. 2000. "Assessing Learning Achievement." Report of UNESCO for the International Consultative Forum on Education for All.
- Fuller, Bruce, and Prema Clark. 1994. "Raising School Effects While Ignoring Culture? Local Conditions and the Influence of Classroom Tools and Pedagogy." *Review of Education Research* 64(1):119-57.
- Flynn, J.R. 1987. "Massive IQ Gains in 14 Nations: What IQ Tests Really Measure." *Psychological Bulletin* 101, pp 171-191.
- Galabawa, J.C., F.E.M.K.Senkoro, and A.F.Lwaitama. 2000. *The Quality of Education in Tanzania*. Dar-es-Salaam: Institute of Kiswahili Research.
- Glewwe, Paul. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes." *Journal of Economic Literature* 40:436-482.
- Glewwe, Paul; Michael Kremer and Sylvie Moulin. 2001. "Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya," Development Research Group, World Bank.
- Greaney, Vincent, and Thomas Kellaghan. (1996). "Monitoring the Learning Outcomes of Education Systems." World Bank Directions in Development Series.

- Grossman, Jean Baldwin. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *World Bank Research Observer* 9(2):159-180.
- Gundlach, Eric, and Ludger Wossman, 2001. "The Fading Productivity of Schooling in East Asia." *Journal of Asian Economics* 12(3):401-417.
- Gundlach, Eric, and Ludger Wossman, and Jens Gmelin. 2001. "The Decline of Schooling Productivity in OECD Countries." *Economic Journal* 111(471):C135-C147.
- Hanushek, Eric, and Javier Luque. 2003. "Efficiency and Equity in Schools Around the World." *Economics of Education Review* 22: 481-502.
- Hanushek, Eric. 1995. "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer* 10:2, pp.227-46.
- _____. 1994. "Making Schools Work: Improving Performance and Controlling Costs." Brookings Institution. Washington, D.C.
- _____. 1986. "The Economics of Schooling." *Journal of Economic Literature*.
- Harbison, Ralph, and Eric Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. Oxford U. Press for the World Bank.
- Heyneman, Stephen. 1979. "Why Impoverished Children Do Well in Ugandan Schools." *Comparative Education* 15(2), pp. 175-85.
- Heyneman, Stephen, and William Loxely. 1983. "The Effect of Primary of Primary School Quality on Academic Achievement Across Twenty-Nine High and Low Income Countries." *American Journal of Sociology* May 1983, 88, pp.2262-94.
- Kingdon, Geeta. 1996. "The Quality and Efficiency of Private and Public Education: A Case Study of Urban India." *Oxford Bulletin of Economic Statistics* 58(1),pp.57-82.
- Knight, John B., and Richard Sabot. 1990. "Education, Productivity and Inequality" Oxford University Press for the World Bank.
- Kremer, Michael. 2003. "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." *American Economic Review Papers and Proceedings* 93:2, pp. 102-106.
- Kremer, Michael. 1995. "Research on Schooling: What We Know and What We Don't: A Comment on Hanushek," *World Bank Research Observer* 10:2, pp.247-54.
- Leithwood, Kenneth, and Tereza Menzies. 1998. "Forms and Effects of School-Based Management: A Review." *Educational Policy* 12(3), pp 325-346.
- Lockheed, Marlaine and Adriaan Verspoor. 1991. "Improving Primary Education in Developing Countries." Oxford. Oxford University Press.
- McEwan, Patrick J., and Martin Carnoy. 2000. "The Effectiveness and Efficiency of Private Schools in Chile's Voucher System." *Educational Evaluation and Policy Analysis* 22(3):213-39.
- Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *World Bank Research Observer* 9(2):181-201.
- Pandey, R.S. 2000. "Going to Scale with Education Reform: India's District Primary Education Program, 1995-99," *Country Studies in Education Reform and Management Publication Series*, World Bank.

- Pritchett, Lant, and Deon Filmer. 1999. "What Education Production Functions Really Show: A Positive Theory of Education Expenditure." *Economics of Education Review* 18(2):223-239.
- Roberts-Schweitzer, Eluned, Andrei Markov, and Alexander Tretyakov. 2002. *Achieving Education for All Goals: School Grants*. Staff Working Paper. World Bank.
- Rosenzweig, Mark. 2000. "Natural 'Natural' Experiments in Economics." *Journal of Economic Literature* 38:4, pp. 827-74.
- Schultz, T. Paul. 2003. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics*.
- Stigler, James W., and James Hiebert. 1997. "Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study." *Phi Delta Kappan*. Vol.78.
- Theunynck, Serge. 2002. "School Construction in Developing Countries: What Do We Know?" Working Paper in Education for All Case Studies Series. World Bank.
- Velez, Eduardo, Ernesto Shiefelbein, and Jorge Valenzuela. 1993. "Factors Affecting Achievement in Primary Education: A Review of the Literature for Latin America and the Caribbean." Human Resources Division, Technical Department, Latin American and Caribbean Region. World Bank.
- Verspoor, Adriaan. 2003. "The Challenge of Learning: Improving the Quality of Basic Education." *Newsletter of the Association for the Development of Education in Africa* 15(4): 4 -7.
- Williamson, Oliver E. 2000. "The New Institutional Economics: Taking Stock, Looking Ahead." *Journal of Economic Literature*. 38(3):595-613
- White, Howard. 2004. "Books, Buildings, and Learning Outcomes: An Impact Evaluation of World Bank Support To Basic Education in Ghana." OED World Bank.
- World Bank. 2004. "World Development Report: Making Services Work for Poor People."
- _____. 2001. "Hidden Challenges to Education Systems in Transition Economies". Human Development Sector Unit. Europe and Central Asia.
- _____. 1999. *Education Sector Strategy Paper*.
- _____. 1995. *Priorities and Strategies for Education*. A World Bank Review.
- _____. 1990. *Primary Education*. A World Bank Policy Paper.
- Wossman, Ludger. "Schooling Resources, Educational Institutions and Student Performance: The International Evidence." *Oxford Bulletin of Economics and Statistics*, May 2003, 65(2), pp.117-170.