CHAPTER 10

Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications

P. Glewwe*, K. Muralidharan†
*University of Minnesota, St. Paul, Minnesota, United States
†University of California San Diego, La Jolla, California, United States; NBER; JPAL

Contents

Abstract

Improvements in empirical research standards for credible identification of the causal impact of education policies on education outcomes have led to a significant increase in the body of evidence available on improving education outcomes in developing countries. This chapter aims to synthesize this evidence, interpret their results, and discuss the reasons why some interventions appear to be effective and others do not, with the ultimate goal of drawing implications for both research and policy. Interpreting the evidence for generalizable lessons is challenging because of variation across contexts, duration and quality of studies, and the details of specific interventions studied. Nevertheless, some broad patterns do emerge. Demand-side interventions that increase the immediate returns to (or reduce household costs of) school enrollment, or that increase students' returns to effort, are broadly effective

Handbook of the Economics of Education, Volume 5
ISSN 1574-0692, http://dx.doi.org/10.1016/B978-0-444-63459-7.00010-5

at increasing time in school and learning outcomes, but vary considerably in cost-effectiveness. Many expensive "standard" school inputs are often not very effective at improving outcomes, though some specific inputs (which are often less expensive) are. Interventions that focus on improved pedagogy (especially supplemental instruction to children lagging behind grade level competencies) are particularly effective, and so are interventions that improve school governance and teacher accountability. Our broad policy message is that the evidence points to several promising ways in which the efficiency of education spending in developing countries can be improved by pivoting public expenditure from less cost-effective to more cost-effective ways of achieving the same objectives. We conclude by documenting areas where more research is needed, and offer suggestions on the public goods and standards needed to make it easier for decentralized and uncoordinated research studies to be compared across contexts.

Keywords

Education in developing countries, School enrollment and attendance, Learning outcomes, Demand for education, School inputs, Pedagogy, School governance, Policy evaluation, Cost effectiveness

1. INTRODUCTION

Governments in developing countries, international aid agencies, and almost all economists agree that an educated populace is necessary — though not necessarily sufficient — for long-run economic growth and, more generally, a high standard of living. The governments in these countries spend approximately one trillion dollars each year on education, and households spend hundreds of billions more (a precise amount is difficult to calculate) on their children's education. This spending, along with several other factors, has led to large increases in school enrollment at all levels in the past 25 years. Indeed, most children in developing countries now complete primary school and obtain at least some schooling at the secondary level.

Yet enrolling in school does not guarantee that children acquire the human capital that their schooling should provide. Indeed, there is ample evidence that in many developing countries children learn much less than the goals set out in the official curriculum. Also, in some of these countries a large fraction of children still are not finishing primary school, and there are still millions of children who never attend school at all. Moreover, high rates of enrollment often mask low rates of actual school attendance in many low-income settings.

Given this state of affairs, economists and other researchers who study education in developing countries have focused on two main questions. The first is: What education policies increase student enrollment, attendance, and completed years of schooling? The second is: What education policies increase student learning? Further, given the tight fiscal constraints under which developing country governments operate, a key question for researchers has been to identify the most cost-effective ways of achieving positive education impacts, and to identify the policies that are less cost effective.

Much research has been conducted in the last 25 years that attempts to answer these two questions, and this research has accelerated in the past 5–10 years. In particular, during the decade that has passed since the publication of the previous Handbook chapter on the economics of education in developing countries (Glewwe and Kremer, 2006) there has been a sharp increase in the quantity and quality of empirical research in developing countries. This growth in evidence has been facilitated by the increasing use of randomized experiments to evaluate the impacts of education interventions, as well as by the increasing availability of administrative and survey data sets on education in developing countries.

The growing body of this literature is reflected in the fact that there have been several other reviews of this literature in just the last 3 years, including Glewwe et al. (2013), Kremer et al. (2013), Krishnaratne et al. (2013), McEwan (2015), and Murnane and Ganimian (2014).[1] Further, the variation across these reviews in their interpretation of the research literature and of the corresponding policy implications has also prompted a recent survey of surveys by Evans and Popova (2015) to understand the source of this variation. The main conclusion of Evans and Popova (2015) is that the differences across these syntheses can partly be explained by the variation in the studies included in the reviews (due in part to differences in evidence-quality thresholds used for including studies), and partly by variation in how interventions are classified. They also recommend that future reviews combine a formal meta-analysis with a more narrative explanation of the specific interventions and contexts in order to better interpret the results and the variation across studies.

Given this proliferation of recent reviews of research on education in developing countries, this chapter aims to provide a framework that can be used to synthesize what can be learned from the evidence about the nature of education and human capital production in developing countries, and the ways in which governments can deploy their limited resources more effectively to achieve their goal of increasing the human capital of their citizens. While mechanical meta-analyses focus on classifying interventions and summarizing the estimates, we emphasize that many challenges arise when synthesizing the evidence to reach a definitive view of what works, and what does not. These challenges include variations in context, variations in duration of the evaluations and the outcomes studied, and perhaps most importantly variations in the details of the interventions themselves.

We therefore aim not only to summarize what has been found, but also to discuss and interpret these results and to discuss the reasons why some interventions appear to be effective and others do not, with the ultimate goal of drawing implications for both research and policy. Thus, in addition to summarizing and synthesizing the high-quality studies on education in developing countries, an important goal of this chapter is to provide a framework for understanding what the results *mean* in addition to summarizing what they *are*. In doing so, we also hope to provide guidance and suggestions for future

[1] This list does not even include other reviews that have reviewed education research in specific geographical settings, such as Muralidharan (2013), which focuses on India, and Conn (2014), which focuses on Africa.

research. Some of the key themes we seek to emphasize (especially relative to the chapter written a decade ago by Glewwe and Kremer, 2006) include the following, which are discussed in detail in Section 5 of the chapter.

First, the design of the details of any intervention matter enormously and should be more theoretically informed. Second, while the standard education production function approach seeks to estimate population average parameters, heterogeneity across students is likely to be a first order issue. Thus, the optimal policy is likely to be different at different points in the student learning distribution. Third, it is particularly important to try to understand *why* well-intentioned interventions may not have much of an impact on learning outcomes, and to aim such inquiry toward identifying binding constraints in the education system of interest. Fourth, interactions across components of an education system are likely to be of first order importance, which poses a challenge for most traditional research methods, which typically are not well suited to studying interactions. Finally, while several individual high-quality studies have been produced in the past decade, more public goods and standards for measurement and reporting need to be created to make it easier for highly decentralized (and often opportunistically conducted) research studies to be compared across contexts. This will contribute to a more systematic understanding of the most cost-effective ways to improve education outcomes in developing countries.

The plan of this review chapter is as follows. The next section presents recent trends in education outcomes in developing countries and places the performance of developing countries in perspective by accounting for the extent to which they over- or under-perform relative to their per capita income. Section 3 provides a theoretical framework for interpreting the evidence that will be reviewed later in the paper, discusses the empirical challenges in generating credible evidence on the impacts of policies to improve education, and outlines the methodology for determining how studies were selected to be included in this review. Section 4 presents our own review of the evidence to date, organized in a way that is consistent with the theoretical framework. Section 5 aims to interpret the body of evidence reviewed in Section 4, and discusses key themes that emerge for both research and policy. Section 6 concludes with a brief set of recommendations for education researchers, as well as for funders and users of such research.

2. TRENDS IN EDUCATION OUTCOMES, 1990 TO 2014

Primary and secondary school enrollment rates have increased in all regions of the developing world in the 50 years from 1960 to 2010, as seen in Table 1.[2] In 1960, primary

[2] Data are from the World Bank Edstats data based. Data are missing for some countries, especially in 1960. These are usually very small countries, but there are a few cases of missing data for very large countries. In particular, data are missing for China and the United States in 1960 and for Brazil in 2010. Limiting the data to countries with data in both 1960 and 2010 does not have a large effect on the figures in Table 1.

Table 1 Primary and enrollment rates, 1960 and 2010, by region

	Primary enrollment and completion				Secondary enrollment		
	Enrollment rate		Completion rate	Income adjusted residual (2010)	Enrollment rate		Income adjusted residual (2010)
Region	1960	2010	2010		1960	2010	
OECD countries	110	102	99	−8.0	52	102	−2.6
East Asia and Pacific	73	122	99	10.9	12	82	4.8
Eastern Europe and Central Asia	98	100	98	7.4	53	93	13.7
Latin America and Caribbean	88	108	98	4.2	14	85	3.1
Middle East and North Africa	54	108	98	1.7	10	85	−0.3
South Asia	56	111	92	8.0	18	61	2.0
Sub-Saharan Africa	41	101	68	−10.4	3	44	−12.1

school gross enrollment rates (GER) in the OECD countries[3] and in the countries of Eastern Europe and Central Asia (these are the countries and allies of the former Soviet Union) were above or very close to 100%, and were again close to 100% in 2010. Among developing country regions, the primary school GER in Latin America and East Asia were somewhat lower in 1960 (88% and 73%, respectively), but were over 100% in 2010. Farther behind in 1960 were South Asia (56%), the Middle East and North Africa (54%) and Sub-Saharan Africa (41%). Yet all three of these regions had achieved a 100% rate or higher by 2010.

While the primary GER data for 2010 suggest that all children in all regions are now completing primary school, this is not the case. GER can be misleading because they are calculated as the number of children enrolled in a particular level of school divided by the number of children in the population in the age range associated with that level. If some students repeat grades, or start school at a late age, they may still be in that level of schooling even though they are older than the associated age range. Repetition will, in general, increase GER and so a rate of 100% or even higher is consistent with a sizeable proportion of the population never finishing, and even never enrolling, in that level of schooling.

To assess more fully current progress in primary education it is most useful to examine what proportion of children *complete* primary school. This is shown in the third column of Table 1, for the year 2010 only since very few countries have completion data for 1960. By 2010 close to 100% primary school completion had been achieved in all other regions except South Asia and Sub-Saharan Africa. Yet South Asia is on course to reach full primary completion within a few years, having attained a 92% completion rate by 2010.

[3] The OECD countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

Even though Sub-Saharan Africa's primary school completion rate was only 68% in 2010, this is a large increase from 46% in 2000 (not shown in Table 1), and suggests that a rate close to 90% would be possible by 2020.

Secondary school GER data, by region, are shown for 1960 and 2010 in the fifth and sixth columns of Table 1. The rates in 1960 were only around 52–53% for the OECD countries and the Eastern European and Central Asian countries, but they had reached 100% by 2010. The rates in all five developing country regions were much lower, all below 20%, in 1960, with a very low rate of 3% for Sub-Saharan Africa. Yet by 2010 all regions of the developing world except for South Asia and Sub-Saharan Africa had rates of 80% or higher. In South Asia, the rate increased from 18% in 1960 to 61% in 2010, which is higher than OECD rate in 1960. In Sub-Saharan Africa, the rate increased from a very low rate of only 3% in 1960, reaching 44% in 2010, which is close to the OECD average in 1960.

These patterns raise the question of whether the variation in education attainment between OECD and developing countries are primarily driven by differences in per capita income. In other words, are developing countries lagging behind developed countries in terms of their enrollment rates given their income levels? This question is of little interest for primary school enrollment rates since all regions have attained average GER of 100% or higher. Indeed, Fig. 1 shows that the relationship between the primary GER and (log) per capita income in 2010 is essentially flat. Turning to the primary school completion rate, we find a strong positive correlation with the log of per capita income

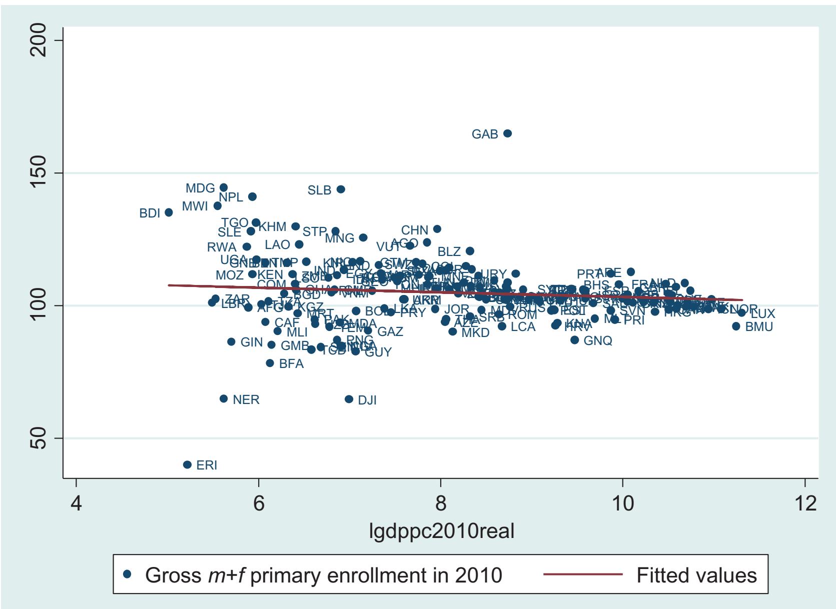

Figure 1 Primary gross enrollment rate by log real GDP/capita, 2010.

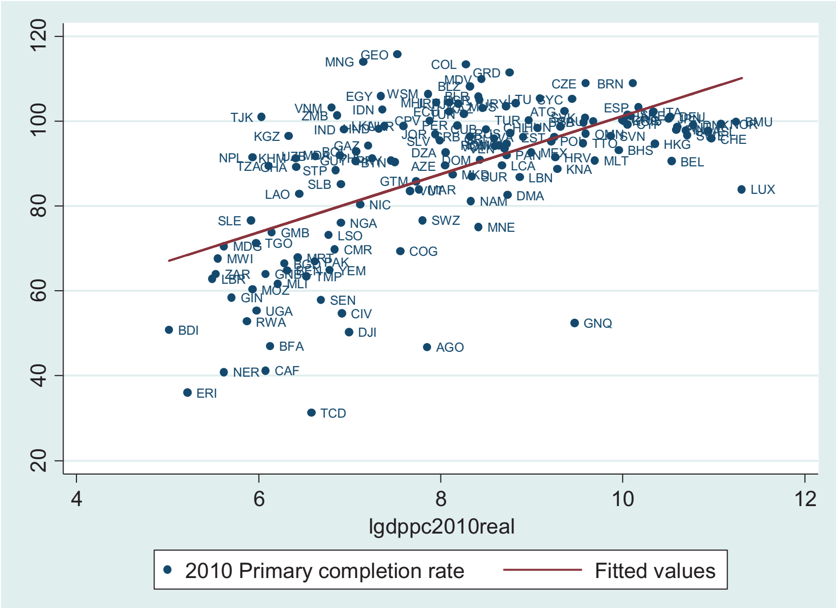

Figure 2 Primary completion rate by log real GDP/capita, 2010.

(Fig. 2). If developing countries are lagging behind the OECD countries even after conditioning on income, they would tend to lie below the line while the OECD countries would tend to be above the line. But this is not the case, as seen in the fourth column of Table 1. Indeed, the average residual for OECD countries is −8.0, which means that those countries tend to lie below the line, and so developing countries lie above it, on average. In other words, developing countries on average outperform on primary school completion rates conditional on per capita income. However, there is variation among developing countries, with East Asian, Eastern European and Central Asian, and South Asian countries generally well above the line while Sub-Saharan African countries are, on average, well below it (−10.4), even after accounting for their income levels.

Fig. 3 and the last column of Table 1 examine secondary school GER conditional on income. That relationship, from an analogous regression, is clearly positive. The last column in Table 1 investigates the performance of different groups of countries conditional on their income levels. On average, OECD countries have a slightly negative residual (−2.6), which implies that they perform slightly worse than average given their income level, so again developing countries, on average, perform somewhat better than average conditional on their income. Yet again there is variation among these countries, with Eastern Europe and Central Asian countries performing well above average, and Sub-Saharan African countries again performing well below average even after conditioning on their income levels.

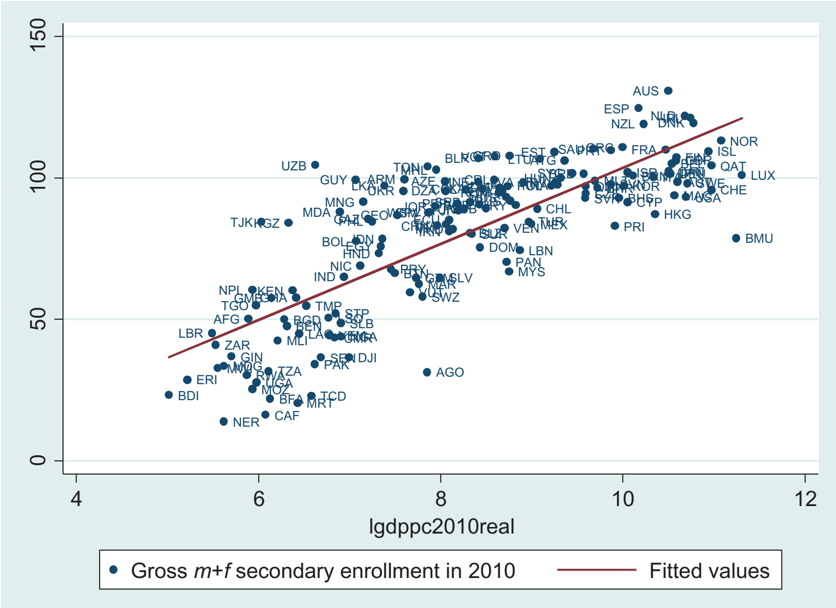

Figure 3 Secondary gross enrollment rate by log real GDP/capita, 2010.

Another way to assess the performance of developing countries in a way that controls for income levels is to examine school enrollment rates of developed countries when their income levels were similar to those of developing countries today. This is done in Table 2, examining three prominent developed countries: the United States, the United Kingdom, and France. Today's low-income developing countries (as defined by the World Bank), have a (purchasing power adjusted) per capita income in 2010 of $1307. France had a somewhat higher (unadjusted) per capita income ($1987) almost 200 years ago, in 1830. At that time, France's primary enrollment rate was only 39%, and its secondary enrollment rate was only 1%, well below the respective rates for low-income countries in 2010 (104% and 39%).

In Table 1, Sub-Saharan African countries seemed to perform poorly even after conditioning on their income levels, but Table 2 demonstrates that their performance is much better than those of today's developed countries when they had income levels similar to the average income level today in Sub-Saharan Africa. More specifically, in 2010 Sub-Saharan African countries had a (purchasing power adjusted) per capita income of $2148. The United States had a slightly higher (unadjusted) per capita income ($2582) almost 200 years ago, in 1830. At that time, its primary enrollment rate was only 55%, and its secondary enrollment rate was less than 1%, well below the respective average rates in Sub-Saharan Africa in 2010 (100% and 36%).

Table 2 Primary and secondary school gross enrollment rates: a historical perspective

Country or country group	Year	GDP/capita (2010 US$)	Enrollment rate Primary	Enrollment rate Secondary
Low-income developing countries	2010	1307	104	39
Middle income developing countries	2010	6747	109	69
South Asian countries	2010	3124	110	55
Sub-Saharan African countries	2010	2148	100	36
United States	1830	2582	55	<1
	1850	3085	68	<1
	1900	6827	88	37
United Kingdom	1890	6691	72	1
France	1830	1987	39	1
	1870	3130	74	1
	1930	7563	80	3

Sources: All 2010 figures are from World Development Indicators 2012. Note that the GDP/capita numbers are adjusted for purchasing power parity. Historical income data for the United States, United Kingdom, and France are from the Maddison-Project, http://www.ggdc.net/maddison/maddison-project/home.htm. Historical enrollment data from those countries were compiled by Peter Lindert, http://lindert.econ.ucdavis.edu/data-and-estimates/lindert-data-for-cup-book.

South Asian countries also have relatively low levels of primary and secondary school enrollment, but their enrollment rates are much higher than those of the United States and France when they were at the average income level of South Asian countries in 2010 ($3124). In particular, the United States in 1850 and France in 1870 had very similar income levels, but their primary school enrollment rates were much lower (68% in the United States and 74% in France) than the average for South Asia in 2010 (110) and their secondary school enrollment rates were only about 1%, compared to 55% in South Asian countries today.

A final interesting historical comparison is with the United Kingdom, which had a much higher income in the 1800s relative to France and the United States. In 1890, the per capita income in the United Kingdom (in 2010 dollars) was $6691, which is similar to that of middle income countries in 2010 ($6747). However, in 1890, the primary and secondary school enrollment rates in the United Kingdom were only 72% and 1%, respectively. These are much lower than the 2010 rates for middle income countries, which were 109% and 69%, respectively.

These findings are similar to those of Deaton (2013) for health outcomes; he showed that populations in today's developing countries are much healthier than those in the developed countries a century earlier. These findings for health are almost certainly due to major advances in medicine over that period of time (such as the introduction of vaccinations and antibiotics), but it is not clear that analogous advances in education "technology" explain, even in part, the higher levels of education in today's developing countries compared to developed countries at the same income level in previous decades and centuries. Some possible explanations for these results include: (1) international campaigns for universal primary education,[4] which may have influenced global social norms in a way that persuaded governments in developing countries to prioritize primary education and also led to increased international donor funds to support this goal; (2) greater rates of democratization in developing countries now than in developed countries at a comparable level of per capita income, which may have created domestic political pressure for prioritizing education; and (3) greater returns to education in a world with more advanced technologies, many of which are complements to human capital.

We now examine international comparisons of learning outcomes, and the corresponding correlation with GDP per capita. There are two international learning assessment initiatives that administer comparable mathematics, reading and science tests to school-age children across a large number of developed and developing countries, the PISA and the TIMSS assessments. Both normalize the distributions of test scores so that the mean equals 500 and the standard deviation equals 100.

As others have shown (eg, Hanushek and Woessmann, 2015), students in many, and perhaps most, developing countries learn much less than students of the same age, or in the same grade, learn in OECD countries. However, OECD countries have much higher incomes than developing countries, and parents in OECD countries have much levels of education than do parents in developing countries. Thus, it should not be surprising that students in developed countries learn more each year, since their parents are likely to be both more educated and more able to afford additional school inputs, and the school system can afford more and higher quality inputs. So we conduct a similar exercise as in the case of school completion and plot cross-country test scores as a function of per capita income (Figs. 4–7).

To start, Fig. 4 examines the math scores of 15-year-old students in the 2012 PISA assessment. As expected, there is a general trend that countries with higher (log) incomes have higher test scores. The line in this figure is the least squares line for the all the points in the figure from a regression of test scores on a constant term and the log of per capita income. The OECD countries are "hollow" circles, while all other countries are solid ("filled") circles. However, there is no tendency for OECD countries to be "better than average" given their incomes — they are just as likely to be below the line as they are to be above it. This implies that developing countries as a whole are also found above the line as often as they are below it. Fig. 5 repeats this analysis for the 2012 PISA reading scores, and the results are similar. Similar patterns are seen in the TIMSS data, which are for younger children (in grade 4, so about 10 years old), as seen in Figs. 6 and 7.

The figures presented above provide useful perspective for understanding the challenges of improving human capital accumulation in developing countries. On one hand,

[4] Examples include the "Education for All" campaign and the Millennium Development Goals, the latter of which prioritized both universal primary education and the elimination of gender gaps in education.

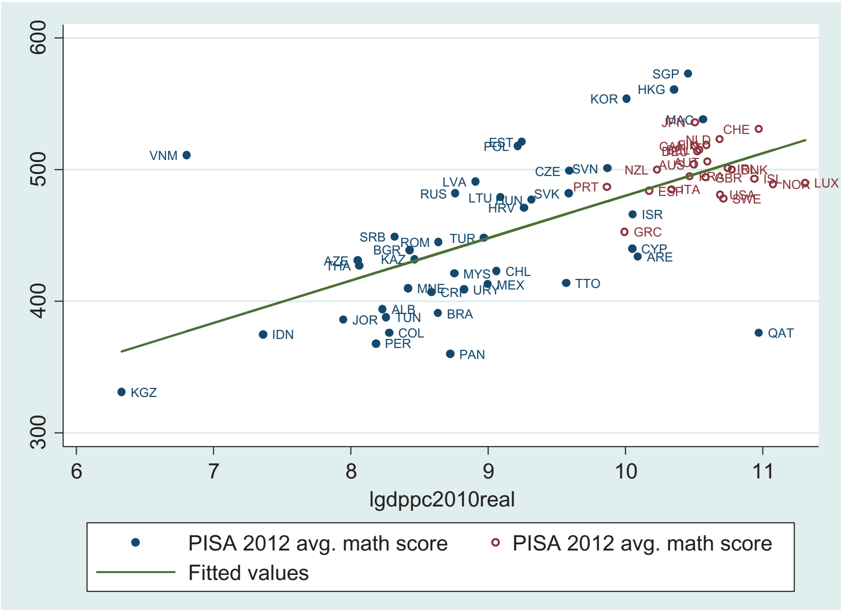

Figure 4 Mean age 15 math scores in 2012 (PISA), by 2010 log real GDP/capita.

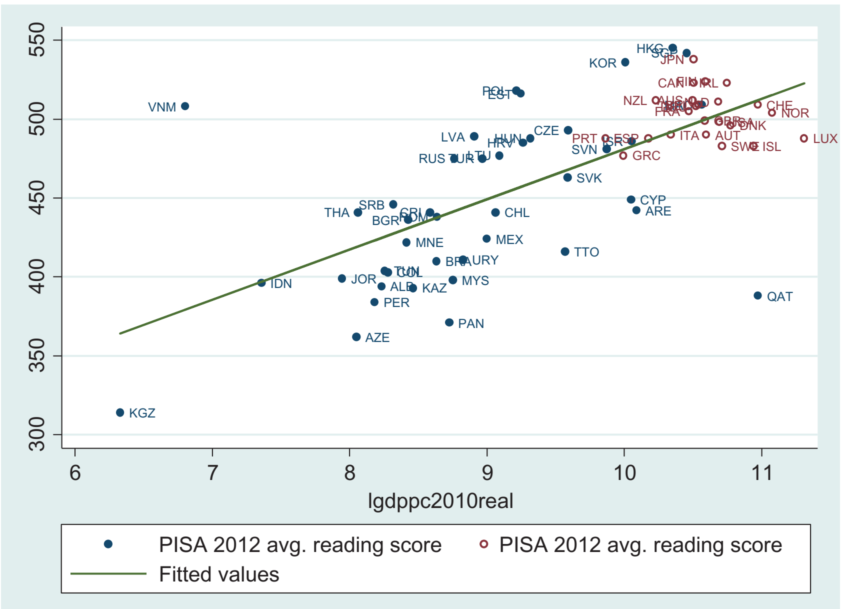

Figure 5 Mean age 15 reading scores in 2012 (PISA), by 2010 log real GDP/capita.

Figure 6 Mean grade 4 math scores in 2011 (TIMSS), by 2010 log real GDP/capita.

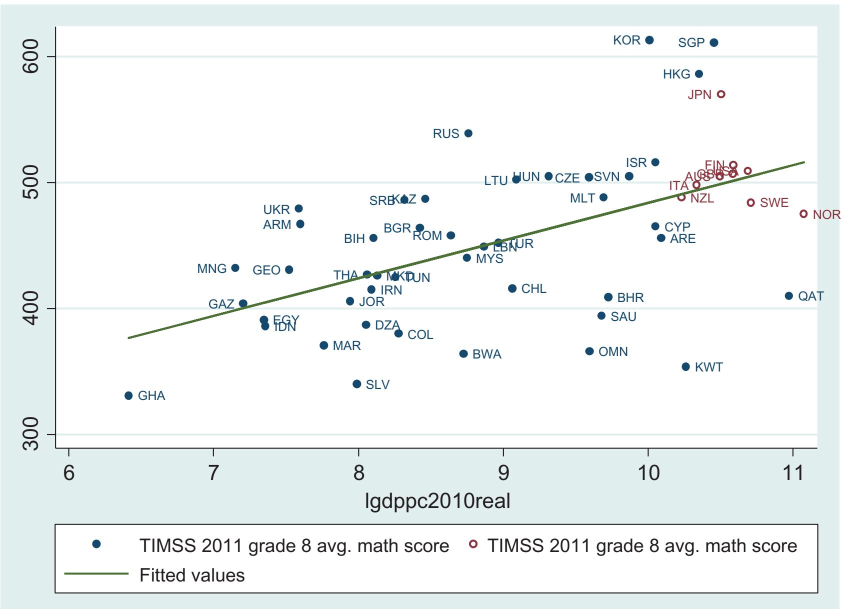

Figure 7 Mean grade 8 math scores in 2011 (TIMSS), by 2010 log real GDP/capita.

the data suggest that developing countries are doing a poor job of translating enrollment into learning outcomes. On the other hand, it is also possible to interpret the results as suggesting that developing countries "overperform" on enrollment conditional on per capita income, and perform "as expected" on learning outcomes conditional on per capita income.[5] Figs. 4–7 suggest that human capital accumulation will likely automatically improve as countries get richer both because parents of future cohorts will be more educated and because countries will have more resources to spend on education.

However, the key policy question of interest should be whether it is possible for developing countries to outperform the levels that would be predicted based on their income levels. One way of doing this is to identify education systems that are strong positive outliers on performance (conditional on per capita income) and to study the features of such education systems in a detailed case study to identify policies that could be translated to other settings for improved performance. For instance, Figs. 4 and 5 suggest that Vietnam is a strong positive outlier in this regard, and may be a role model for other developing country education systems. A second approach is to conduct micro-level (as opposed to system-level) research on the returns to different forms of education spending and reallocate resources from less to more efficient ways of spending limited resources. The advantage of this approach is that it allows for much better causal identification of the impacts (or nonimpacts) of specific categories of education policies, and policy recommendations based on such research is likely to be better grounded in credible evidence.[6] This is the approach that we will focus on for the rest of this chapter.

3. CONCEPTUAL ISSUES

In this section, we discuss two key conceptual issues. First, we present a simple theoretical framework to help organize our thinking to better understand how various policy and

[5] Note however, that the sample of students from developing countries who are tested in the TIMSS and PISA is much more nonrepresentative than the corresponding sample in OECD countries. Since these tests are conducted in school, they are only representative of students who are in school (and likely to be attending school regularly). Thus, it is likely that the true levels of learning in developing countries are considerably lower than those measured by TIMSS and PISA, and that these measured scores represent an upper bound on true levels of learning. If we had representative data on learning outcomes in all countries, it is possible that developing countries would be found to be "underperforming" relative to per capita income.

[6] For instance, it is possible that countries like Vietnam and Cuba (see Carnoy et al., 2007) outperform on education relative to per capita income because they were communist societies that (a) prioritized equality in education over outstanding individual achievements, (b) placed greater value on the "indoctrination" of citizens in party ideology and hence prioritized state-run education. A country case study would yield useful insights into such factors, but would not be able to credibly isolate the specific components of communist education systems that were responsible for the "positive deviance." Further, it is not clear whether replicating elements of the policies of these countries would work in the absence of a larger communist society. Nevertheless, we do not argue against the value of such system-level case studies, and consider them to be useful complements to the approach taken in this chapter.

household decisions may affect the production of skills and human capital by education systems. Second, we discuss the key empirical challenges to identifying the impacts of policies to improve education outcomes and discuss the process by which we identify the "high-quality" studies for which the evidence is summarized in Section 4.

3.1 Theoretical Framework

Education policy and research in developing countries have been motivated by finding answers to two key questions. The first one is "how to increase school enrollment and daily attendance among students," and the second one is "how to translate increases in student enrollment and attendance into improvements in skills and human capital." On school enrollment and attendance, a simple model of optimizing households in the tradition of Becker (1962) and Ben-Porath (1967) yields the result that households will invest in an additional year of education for their child only if the present discounted value of the expected increase in benefits exceeds the costs of doing so. Thus, policies that seek to improve school enrollment and attendance typically aim to increase the immediate benefits to households of sending their children to school or to reduce the costs of doing so. The magnitude of the impact of these policies will in turn depend on the distribution of the household and child specific factors, and the extent to which the policy increases the benefits of enrolling a child in school.

As summarized in the previous section, most developing countries have made substantial progress with regard to enrolling children in school; indeed, primary and secondary school enrollment rates are considerably higher than the rates in OECD countries when they were at comparable levels of per capita income. Nevertheless, while enrollment rates are high, in practice school attendance rates are quite varied and may not be as high.[7] This may be particularly true in cases where there is pressure on schools and parents to show high enrollment rates (in response to budget allocation rules and/or compulsory schooling laws). Thus, the challenge of "enrollment" is now better thought of as the challenge of "attendance" in most developing countries. However, the same theoretical framework of comparing costs and benefits of school enrollment can be applied to analyzing the household (and children's) decision of whether to attend school on a given day, and policies that seek to improve attendance similarly focus on increasing the returns to attending school on a given day (for instance through the provision of school meals) or on reducing the costs of doing so (for instance through improved school access).

[7] For instance, nationally representative survey data in India collected by the NGO Pratham as part of their annual ASER report finds that on average only 71.4% of students enrolled in primary schools were present during visits to schools, even though the enrollment rate was above 96%. Further, there is substantial variation across Indian states, with several large and economically and educationally backward states such as Bihar, Jharkhand, Uttar Pradesh, West Bengal, and Madhya Pradesh, with a cumulative population of over 500 million people having attendance rates as low as 50–60% (ASER, 2014).

The standard theoretical framework for understanding how various inputs at the household, school, and classroom level translate into learning outcomes is to specify and estimate an education production function. The production function for learning is a structural relationship that can be depicted as:

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

where A is skills learned (*achievement*), S is years of schooling, $\mathbf{Q}$ is a vector of school and teacher characteristics (*quality*), $\mathbf{C}$ is a vector of child characteristics (including "innate ability"), $\mathbf{H}$ is a vector of household characteristics, and $\mathbf{I}$ is a vector of school inputs under the control of households, such as children's daily attendance, effort in school and in doing homework, and purchases of school supplies.

The household decision making process on both school attendance (S) and the extent of household investments in education ($\mathbf{I}$) is based on optimizing household utility subject to the production function above and a set of constraints. Economists typically assume that each child is a member of a household in which the parents of that child maximize, subject to constraints, a (life-cycle) utility function. The constraints faced are the production function for learning, the impacts of years of schooling and of skills obtained on the future labor incomes of children, a life-cycle budget constraint, and perhaps some credit constraints or an agricultural production function (for which child labor is one possible input).

Assume that all elements in the vectors $\mathbf{C}$ and $\mathbf{H}$ are exogenous. Examples of such variables are credit constraints, parental tastes for schooling, parental education, and children's "ability." Some child characteristics that affect education outcomes (such as child health) could be endogenous; they can be treated as elements of $\mathbf{I}$, all of which are endogenous. Parental decisions regarding the endogenous variables, S and $\mathbf{I}$, depend on another important set of variables, the prices related to schooling, $\mathbf{P}$. These include school fees, prices for school supplies, and even wages paid for child labor. $\mathbf{P}$ does not appear in Eq. (1) because it has no direct effect on learning; its effect works through decisions made for the endogenous variables S and $\mathbf{I}$.

In the simplest scenario, assume that only one school is available and that parents cannot change the characteristics of that school. Thus all variables in $\mathbf{Q}$ and $\mathbf{P}$ are exogenous.[8] Parents choose S and $\mathbf{I}$ (subject to the above-mentioned constraints) to maximize household utility, which implies that both S and $\mathbf{I}$ can be expressed as functions of the four vectors of exogenous variables:

[8] In settings where households can choose from more than one school, it is possible that that $\mathbf{Q}$ and $\mathbf{P}$ may be endogenous even if they are fixed for any given school. Here, households maximize utility for each school in their schooling choice set, and then choose the school that leads to the highest utility. Conditional on choosing that school, they choose S and $\mathbf{I}$, as in the case where there is only one school from which to choose. In practice, rural settings are more likely to have limited choice across schools, while higher-density urban settings may feature more choice.

$$S = f(\mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{P}) \tag{2}$$

$$\mathbf{I} = g(\mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{P}) \tag{3}$$

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

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

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

Policymakers are primarily concerned with the impact of education policies on eventual academic achievement, A. For example, reducing class size is a change in one element of $\mathbf{Q}$, and changing tuition fees is a change in one component of $\mathbf{P}$. Eqs. (2) and (4) show how such changes would affect S and A. Assuming that the costs of such changes are not difficult to calculate, the benefits in terms of increases in S and A can be compared to those costs. Of course, the costs should include costs borne by households from the policy change, so changes in $\mathbf{I}$, as expressed in Eq. (3), and in household leisure must be included in the overall cost figure (though in practice this is not easy to do).

Consider a change in one element of $\mathbf{Q}$, denoted by Q_i. Eq. (1) shows how changes in Q_i affect A holding all other explanatory variable constant, and thus provides the *partial* derivative of A with respect to Q_i. This represents a *production function parameter*. In contrast, Eq. (4) provides the *total* derivative of A with respect to Q_i as it allows for changes in S and $\mathbf{I}$ in response to the change in Q_i. This represents a *policy parameter*. Parents may respond to better teaching by increasing their provision of educational inputs; alternatively, if they consider better teaching a substitute for those inputs, they may reduce those inputs. In general, the partial and total derivatives could be quite different, and researchers should be clear which relationship they are estimating.

When examining the impact of policies on academic skills, A, should policymakers look at Eq. (1), or Eq. (4)? Eq. (4) is of interest because it shows what will actually happen to A after a change in one or more element in $\mathbf{Q}$ or $\mathbf{P}$; Eq. (1) will not show this because it does not account for changes in S and $\mathbf{I}$ in response to changes in $\mathbf{Q}$ and $\mathbf{P}$. Yet the partial derivative from Eq. (1) is also of interest because understanding the actual production function helps to identify the most cost-effective policies for improving education outcomes. Knowing the pure production function impact of a policy change may also better capture overall welfare effects. Intuitively, if parents respond to an increase in Q_i by reducing purchases of inputs $\mathbf{I}$, they will be able to raise household welfare by purchasing more of other goods and services. The reduced form impact (total derivative) reflects the drop in A due to the reduction in $\mathbf{I}$, but it does not account for the increase in household welfare from the increased purchase of other goods and services. In contrast, the production function impact, by holding $\mathbf{I}$ constant, captures the full impact before the benefit is shifted away from A to something not meaured by the reduced form impact. Thus, knowledge of the relationships in both Eqs. (1) and (4) is useful for policymakers; Eq. (4) shows what will happen to education outcomes, while Eq. (1) best captures the overall welfare effect.

In practice, there are two main challenges when attempting to estimate the relationship in either Eq. (1) or Eq. (4). The first is that these equations represent the relationship between inputs and the *total stock* of human capital. Thus, estimating the production function in Eq. (1) would require the econometrician to have data on *all prior inputs* into human capital — including early childhood experiences and even in-utero conditions. This is an extremely challenging requirement and is unlikely to be feasible in almost all settings. Thus, the standard approach to estimating education production functions is to treat the lagged test score as a sufficient statistic for representing prior inputs into learning, and to use a value-added model to study the impact of changing contemporaneous inputs into education on test scores (see Todd and Wolpin, 2003, for further details, including the assumptions needed for this approach to yield consistent estimates of production function parameters of interest).

Specifically, the estimating equation that is typically used has the following form:

$$A_{i,t} = \gamma A_{i,t-1} + \boldsymbol{\beta}'\boldsymbol{X}_{i,t} + \varepsilon_{i,t} \tag{5}$$

where $A_{i,t}$ represents test scores of child i at time t, $A_{i,t-1}$ represents the (lagged) test score at time $t-1$, and $\boldsymbol{X}_{i,t}$ represents a full vector of contemporaneous home and school inputs. While the production function above is linear in $\boldsymbol{X}_{i,t}$ and is typically estimated this way, the specification does not have to be very restrictive, because $\boldsymbol{X}_{i,t}$ can include nonlinear (eg, quadratic) terms in individual inputs, and can also include interaction terms between specific sets of inputs.

However, even with the value-added specification of Eq. (5), the second challenge to consistently estimating $\boldsymbol{\beta}$ is that variation in the right-hand side variables is likely to be correlated with the error term $\varepsilon_{i,t}$. In other words, the observed school, teacher, and household characteristics are all likely to be correlated with unobserved (omitted) school, teacher, and household variables that directly determine learning outcomes — which will lead to biased estimates of $\boldsymbol{\beta}$. In practice, this problem is quite likely to occur. For instance, communities and parents that care more about education are likely to be able to successfully lobby for more school inputs, and are also likely to provide unmeasured inputs into their children's education, which would lead to an upward bias on $\boldsymbol{\beta}$ when Eq. (5) is estimated using cross-sectional data.[9] In other cases, governments may target inputs to disadvantaged areas to improve equity, in which case areas with higher values of due to the program may have lower values of $\varepsilon_{i,t}$, leading to negative correlation between and $\varepsilon_{i,t}$ and thus downwardly biased estimates of $\boldsymbol{\beta}$.

Relative to the evidence available when Glewwe and Kremer (2006) reviewed the literature, the largest advances in the empirical literature in the past decade have come from the increased prevalence of studies that pay careful attention to credibly identifying

[9] This problem is mitigated but not eliminated by including lagged test scores. While including the lagged test scores can help control for time invariant omitted variables, doing so does not eliminate concerns of contemporaneous omitted variables.

the causal impact of various education policies on learning outcomes. In particular, the sharp increase in the number of randomized experiments to study the impacts of education policies has greatly increased the volume of credible evidence on education in developing countries.[10]

The importance of accounting for omitted variable bias in the evaluation of education interventions is starkly illustrated by Glewwe et al. (2004), who compare retrospective and prospective studies of the impact of classroom flipcharts on learning outcomes. They find, using observational data, that flipcharts in classrooms appear to raise student test scores by 0.2 standard deviations of the distribution of students' test scores. However, when they conduct a randomized controlled trial (RCT) of flipcharts in classrooms, they find no impact on test scores at all, suggesting that the nonexperimental estimates were significantly biased upwards (even after controlling for other observable factors). These results underscore the value of field experiments for program evaluation in developing countries, and Glewwe et al. (2004) can be considered analogous to LaLonde (1986) in the US program evaluation literature, which showed that nonexperimental methods were not able to replicate the estimates from experimental evaluations of the impact of job training programs.

Note that the framework in Eq. (5) is quite general and can be used to examine policies that do not directly change $\boldsymbol{X}_{i,t}$ but instead change the way schools are organized such as decentralization, promoting competition by removing restrictions on private schools, or developing incentive schemes that link teacher pay to student performance. In principle, these types of policies affect schooling outcomes by changing what happens in the classroom. For example, increased competition may change the behavior of teachers, and these behaviors can be included as components of the vector $\mathbf{Q}$. Further, Eq. (5) is also general enough to allow for both heterogeneity and complementarity in the impact of specific household and school inputs on learning outcomes. However, in practice, most studies focus mainly on credibly estimating the average value of $\boldsymbol{\beta}$ for specific changes in $\boldsymbol{X}_{i,t}$ since the data requirements for credibly estimating heterogeneity and complementarities are more demanding.

3.2 Empirics

One of the main contributions of this chapter is its review of the evidence on the impacts of different types of education policies and programs on student learning and time in

[10] See the complementary chapter by Muralidharan (2016a) for further details on field experiments in education in developing countries, including a practical guide on how to conduct such experiments successfully. While there is some overlap between that chapter and this one, the focus of that chapter is methodological and aims to provide a user guide for running and interpreting field experiments in education in developing countries, whereas the focus of this chapter is more substantive and aims to synthesize the literature and what we have learned from it.

Table 3 Steps used to select papers used in conducting synthesis

Review step	Procedures used	Number of papers
1	Search EconLit and ERIC databases	13,437
	Review abstracts to eliminate duplicate papers and papers that did not estimate the impacts of school or teacher characteristics for a developing country	1017
2	Review full papers, eliminate papers based on lack of relevance or lack of quantitative analysis	320
3	Exclude papers that are not "high quality" (RCT, RDD, DD)	118
4	Number of high-quality studies that are RCT studies	80

school. However, as we highlight above, an important challenge for empirical research in this area is that of credible causal identification. We therefore limit our synthesis of the evidence to 118 high-quality studies conducted from 1990 to 2014.[11] In this section, we discuss the criteria for selecting these studies from the hundreds of relevant studies in the literature. To identify evaluations that have produced credible impacts of education programs or policies, the following four-step selection process was implemented. The number of studies corresponding to each step is shown in Table 3.

3.2.1 Step 1: Search for Possibly Relevant Papers/Reports and Read Their Abstracts

To begin, a search was conducted on a wide variety of sources, after which evaluations were systematically eliminated that did not meet a series of criteria for relevance and quality. To be included in the search, studies had to have been published in the (peer-reviewed) academic literature from 1990 to 2014, inclusive. Unpublished academic working papers written from 2010 to 2014 were also included. Academic working papers written before 2010 that had not yet been published by the end of 2014 were judged as likely to have some methodological flaws that have resulted in their not being published in peer-reviewed journals, so only academic working papers that were written from 2010 to 2014 were included.

The very first task was to conduct a search for journal articles published between 1990 and 2014 using two search engines for the economics and education literatures, respectively: EconLit and the Education Resources Information Center (ERIC). All papers that list both "education" as a keyword and any one of a list of 124 educational programs or policies as keywords (see the online Appendix for this list) were included in this initial sweep of the literature. The search was also limited to papers that include the name of at least one developing country or the term "developing country" or "developing

[11] Note that almost all papers that are cited with a 2015 publication date are either a revised or final versions of prior working papers that were available in 2014 or earlier, which (as explained below) are included in our review.

countries" in the abstract. Developing countries are defined using the International Monetary Fund's list of emerging and developing countries.

This initial search yielded 13,437 publications. For these papers, information found in the abstract (and, in some cases, by looking at the introduction or conclusion of the paper) was used to limit the studies to those that appear to be potentially relevant. In particular, this eliminated evaluations that did not focus on developing countries, or that did not estimate the impact of a program or policy on students' education outcomes.

In addition to published papers, a search was conducted of several prominent working papers series: National Bureau of Economic Research (NBER) working papers; World Bank Policy Research working papers; the Institute for the Study of Labor (IZA); the Center for Economic and Policy Research (CEPR); the CESIfo Research Network; the Rural Education Action Project (REAP) at Stanford University; and Oxford University's Young Lives Study. Papers listed as education papers on the Abdul Latif Jameel Poverty Action Lab's website were also searched. As mentioned above, working papers that appeared before 2010 were not included due to the assumption that high-quality working papers written before 2010 should have been published by 2014. As with the selection criteria for published papers, evaluations that do not focus on developing countries (for instance, several good studies from Israel were not included because Israel was not considered to be a developing country), or that do not estimate the impact of a program or policy on students' educational outcomes, were not included. The number of published papers and working papers that remained after reading their abstracts was 1017.

3.2.2 Step 2: Read Entire Paper/Report to Verify Relevance

In the second step, all of the evaluations that were not eliminated in the first step were read in full to obtain further information about each study. During this step, additional papers were eliminated for lack of relevance that was not evident from reading the abstracts. Possible reasons for lack of relevance were: (1) The evaluation did not focus on a developing country (which was not always clear in the abstracts); (2) The paper did not evaluate any type of education policy or program; and (3) The paper did not include quantitative analysis of the impact of an education policy or program on students' educational outcomes. After this step was completed, 320 papers remained that were relevant for this review of the literature.

3.2.3 Step 3: Retaining Only High-Quality Evaluations

In a third step, the evaluations that were not eliminated in the first two steps were reviewed for their quality. While regression analysis is commonly used to estimate the impact of a policy or program on an educational outcome, a very serious problem with regression analysis is that some factors that have a causal impact on the education outcome variables of interest are unlikely to be available in the data, which can lead to bias in regression estimates of the impacts of education policies and programs, and thus misleading results. This is the problem of *omitted variable bias*. Another problem with regression analysis is that regressions often included many school and teacher characteristics, but only to serve as control variables, and in many cases authors are not particularly interested in the coefficients associated with those variables, and so they should not be interpreted as estimates of the causal impacts of those variables.

Given these problems with regression estimation methods, all studies based on those methods alone are deemed not to be high-quality studies. Since matching estimators invoke similar assumptions, in particular the assumption that conditioning on (other) observed variables implies that (observed and counterfactual) education outcomes are independent of the program participation variable, studies based on matching methods were also excluded. This leaves three types of studies that are considered to be high-quality studies in this review. First, all evaluations based on a well-implemented RCT are included in the set of high-quality studies, as these studies avoid, or at least minimize, many types of estimation problems. Second, estimates based on a difference-in-differences (DD) regression (which requires longitudinal data) are deemed to be high-quality studies. Finally, evaluations based on regression discontinuity design (RDD) are also considered to be high-quality studies. The set of papers that were retained after this third step contained 118 "high-quality" studies.

3.2.4 Step 4: Identify RCT Evaluations

The fourth and final step of the review set an even higher bar for the quality of an evaluation. Well-implemented RCT studies arguably have the highest credibility, when implemented correctly. In particular, DD studies must rely on the parallel trends assumption, which can be difficult to verify, and RDD evaluations identify impacts only at the "cutoff point," strictly speaking. Of the 118 high-quality studies, about two-thirds (80) were RCTs. Note that we include high-quality non-RCT studies, but we typically give priority to the RCT evidence.

4. REVIEW OF THE EVIDENCE

This section reviews the evidence on the impacts of different types of education policies and programs on student learning and time in school. This is based on 118 high-quality studies conducted from 1990 to 2014. Following the theoretical framework above, we have organized this review as follows: First, we review interventions that are intended to increase the demand for schooling by students and their parents; second, we examine standard educational inputs that are typically provided through schools; third, we review the evidence on policies that are related to pedagogy; finally, we look at policies and programs related to the governance of schools, and to education systems more broadly.

The results of the high-quality studies discussed in this section are summarized in Tables 4–11. These tables list the number of estimates of positive and negative impacts for specific interventions, and they also break these down by whether the results are statistically significant (at the 10% level). The number of studies from which these estimates come are shown in parentheses for each type of estimate (positive or negative, and significant or insignificant) for each intervention. While our discussion does not distinguish between positive and negative point estimates that are statistically insignificant, the tables present this breakdown to help the interested reader see the general direction of impacts even in cases where they are not significant.[12] Note that, unless otherwise indicated, we refer to impacts as "positive" or "negative" only if they are statistically significant at the 10% (or lower) significance level.

4.1 Demand-Side Interventions (Interventions That Focus on Households)

As discussed in Section 3.1, household decisions on whether to send their child to school are based on comparing the (perceived) costs and benefits of doing so. However, there are several reasons why household decisions may not be socially optimal. Possible reasons include: (1) Parents may not accurately perceive the returns to education, and therefore under- or overinvest in education; (2) Households may be credit constrained, which makes them unable to afford the upfront costs (both direct and indirect) of sending their children to school even though the investment in education has, on average, a positive expected value; (3) Households may be risk-averse, which may make them not invest in an extra year of education even if the average expected return is positive due to the variance of the actual return that their child will obtain; (4) Parents may have higher discount rates than the social planner, which may lead to suboptimal investments in education relative to what is socially optimal; and (5) Parents may not account for the social spillovers from their child being educated (as is standard in any model with complementarities across the productivity of workers, such as Kremer, 1993), and so they may underinvest in their child's education.

Thus, the first class of interventions we review are those that aim to alleviate demand-side constraints to school enrollment and daily attendance. Table 4 summarizes the results for the impacts of these types of interventions on measures of time in school, such as daily attendance, current enrollment, and years of completed schooling, and Table 5 does the same for learning outcomes as measured by test scores. Below, we discuss these studies in more detail, grouping them into four general categories: information-based interventions, cash transfer programs, scholarship programs and other household interventions.

[12] This may be especially useful in cases where individual studies lack power to estimate statistically significant impacts, but where multiple studies obtain insignificant estimates that are primarily either positive or negative.

Table 4 Summary of impacts on time in school of demand-side interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Information-based interventions					
Information on returns to education (all RCTs)	0 (0)	1 (1)	1 (1)	2 (1)	2
Career counseling (RCT)	1 (1)	0 (0)	0 (0)	0 (0)	1
School counseling (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Cash transfer programs					
Conditional cash transfer					
RCTs	0 (0)	0 (0)	3 (2)	24 (13)	13
Other high-quality studies	0 (0)	0 (0)	0 (0)	16 (7)	7
Unconditional cash transfers (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Labeled cash transfer (RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Eliminating school fees (non-RCT)	0 (0)	2 (2)	1 (1)	1 (1)	2
Scholarship programs					
Merit-based scholarship (all RCTs)	0 (0)	0 (0)	1 (1)	3 (2)	2
Other household interventions					
Mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Mother literacy class (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Adult literacy program (non-RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Female sanitary products (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Bicycle program (non-RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Matching remittances funds for education (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

4.1.1 Information-Based Interventions

Several demand-side interventions are designed to provide information to students and their parents about the benefits of education and on how to take advantage of educational opportunities. Examples of such policies include providing information on the returns to education in the labor market and more general types of student counseling. The findings

Table 5 Summary of impacts on test scores of demand-side interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Information-based interventions					
Inform. on returns to schooling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Career counseling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Cash transfer programs					
Conditional cash transfer					
RCTs	0 (0)	1 (1)	1 (1)	3 (3)	5
Other high-quality studies	0 (0)	1 (1)	1 (1)	0 (0)	2
Unconditional cash transfers (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1
Labeled cash transfer (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Promise of high school financial aid (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Scholarship programs					
Merit-based scholarship (all RCTs)	0 (0)	1 (1)	0 (0)	6 (4)	4
Other household interventions					
Mother literacy classes (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

from these studies are summarized in the first three lines of Table 4, which focuses on measures of time in school, and the first two lines of Table 5, which focuses on student learning as measured by test scores.

One potential reason why parents invest relatively little in their children's education is that they, or their children, may not be aware of sizeable returns to education in the labor market. Two high-quality studies have examined the impact on students' time in school of providing this type of information to parents and/or their children. As seen in the first line of Table 4, of the four estimates from these two studies, two (one from each study) are statistically insignificant while two (both from the other study) are significantly positive.

The first study of this type was by Jensen (2010), who implemented an RCT that provided information on (estimated) returns to schooling to boys in grade 8 from poor households in the Dominican Republic. This information was provided to these students in the form of a brief statement read to the students at the end of a survey. The motivation for this intervention was the finding that these students generally assumed that the returns to additional years of schooling were very low, which may explain their low rates of secondary school completion. Jensen found that the boys who received the information were 4.1% more likely to be in school 1 year after receiving the information (significant at the 10% level). He also found, using data collected 4 years later, that the boys who received the information were 2.3% more likely to finish secondary school (not significant) and had completed, on average, 0.2 more years of schooling (significant at the 5% level). The effects were strongest for the least poor of those in poor households, and weakest for the poorest of the poor households, which suggests that the latter may face other barriers, such as credit constraints.

In contrast, a similar but more intensive intervention in China by Loyalka et al. (2013) found little effect of providing information. They focused on boys and girls in grade 7 in poor rural areas of two Chinese provinces (Hebei and Shaanxi). As in the Dominican Republic, an initial survey suggested that these students had inaccurate information about the costs and benefits (in terms of future earnings) of education (in particular, they overestimated the cost of vocational education), and that they also lacked career planning skills. The intervention consisted of providing a 45 min information session on earnings associated with different levels of schooling. The evaluation was implemented as an RCT, and the main result is that the program had no significant effect on the dropout rate. The authors speculate that students felt that the quality of their schools was low, and thus they assumed that for their schools the additional time in school would not produce high returns, but they present no evidence to support this conjecture. Loyalka et al. (2013) also examined whether this intervention increased students' test scores, and found no significant effect, as seen in the first row of Table 5.

The Loyalka et al. study also examined another information-based initiative, that of providing four, 45-min sessions on career counseling. As seen in the second row of Table 4, this program had a significantly negative impact on time in school; the authors speculate that this may reflect that students learned that upper secondary and postsecondary entrance requirements were more difficult than they had previously thought. It had no statistically significant impact on student learning as measured by test scores, as seen in the second line of Table 5.

A somewhat similar information-based initiative, also in China, was a school counseling intervention that was evaluated by Huan et al. (2014). This program involved students in grades 7 and 8 who were preparing to take upper secondary school entrance exams. The main goal of the counseling was to reduce students' anxiety. Using an RCT to estimate the impact of this program, the authors found that it reduced the dropout rate by about 2% in the first half of the school year, an estimate that is significant at the 10% level, but there was no effect at the end of the school year (see the third line of

Table 4). The paper did not examine the impact of this intervention on student learning as measured by test scores.

4.1.2 Cash Transfer Programs

One of the most widely implemented demand-side programs during the last two decades, for which there is also a large amount of evidence, is conditional cash transfer (CCT) programs, which provide monetary payments to parents if their children are enrolled in school and have a high rate of attendance (usually 80% or 85%). Note that these programs address two different market imperfections that lead to suboptimal household investments in education. First, by providing resources to the household that can be used for the education of children, CCT programs alleviate credit and resource constraints that may make it difficult for households to bear the upfront direct and indirect costs of schooling. Second, by conditioning the provision of the cash transfer on enrollment, CCT programs aim to increase the *immediate* returns to households of investing in education and thereby reduce potential differences between the individually and socially optimal decisions regarding education of children, which were mentioned above.

While these programs focus on (and indeed condition on) enrollment and attendance, as opposed to test scores directly, to the extent that they increase enrollment and attendance they should also (hopefully) increase student learning. For a more detailed description, and an earlier assessment, of these programs, see Fiszbein et al. (2009). Beyond the question of whether CCTs are effective at improving time in school and learning outcomes, key related questions of interest include: (1) Is the "conditionality" actually necessary, or can it be dropped or replaced by "framing" the transfer as being for children's education (which would reduce the monitoring cost)? (2) Can the design of the CCT be modified to improve education outcomes, for example by adjusting the timing of payments and conditioning them on learning outcomes? (3) How does the impact of the program vary by the amounts of the transfers?

Since CCT programs provide payments to parents if their children are enrolled in school and maintain a high rate of attendance, one would expect that they increase the time that children spend in school. The results are summarized in the fourth and fifth rows of Table 4. Of the 43 estimates from 20 studies, almost all (40) are significantly positive (the other three are statistically insignificant). Thus the evidence is clear that this policy is effective at increasing time school. We now discuss these studies in some detail.

The most studied CCT program is Mexico's *PROGRESA* (later renamed as *Oportunidades*) program, which was both one of the first CCT programs in the world and, crucially, the first to be randomized to allow for credible estimation of its impact. It provided monthly cash payments to the mothers of children in grades 3–9 whose daily attendance rate in the previous month was 85% or higher. The amounts varied by grade, with grade 3 students receiving only about $7 per month and grade 9 students receiving either $22 (boys) or $25 (girls) per month.

Several studies have looked at the impact of *PROGRESA* on outcomes for students at different ages and after different durations of exposure. The first published study of its impact on time in school is that of Schultz (2004). He presents estimates, separately by grade for grades 3–9, of the impact of the program on enrollment in those grades, conditional on completing the previous grade. The largest effect is in the transition from primary to secondary school; conditional on completing grade 6 (completing primary school), students in the communities randomly assigned to the program had a grade 7 enrollment rate 8.7 percentage points higher than students in the control communities.

Behrman et al. (2009) provide estimates on grades of schooling completed that include both urban and rural areas, while Schultz included only rural areas. Another difference is that Behrman et al. (2009) examine children who were 6–14 years old in 2003, while Schultz examines a slightly older cohort: children who were 6–16 in 1999. The estimated impacts on boys' and girls' grades of school completed in Behrman et al. (2009) are small and statistically insignificant after 1.5 years of exposure to the program. In contrast, they obtain much larger and statistically significant estimated impacts after 5.5 years of exposure, impacts of 0.25 grades for boys and 0.32 for girls. A subsequent paper by the same authors, Behrman et al. (2011), also examined impacts after 5.5 years but it focused on an older cohort, children 9–15 years old in 1997, and is limited to rural areas. The children in this cohort, unlike those in the younger cohort considered in the earlier paper, are old enough to have completed their schooling and so the impact on grades of schooling completed is likely to be both larger and to be the full, final impact of the program. For girls aged 15–18 years old in 2003, and thus 9–12 years old when the program started in 1997, the program increased grades completed by about 0.7, and for boys of that age the impact was 0.9 to 1.0 additional grades completed. The impacts for older boys and girls (19–21 in 2003) were lower, especially for girls, likely because these children had fewer years of exposure to the program.[13]

More recent papers have examined CCT programs in other countries. Baird et al. (2011) estimated the impact of a CCT program for girls in Malawi on time in school. For girls already in school when the program started, the number of terms enrolled over the next 2 years increased by 0.535; since each year consists of three terms, this equivalent to one sixth of a year of schooling. Also, the daily attendance rate over the 2 years rose by 8.0 percentage points. They also find that variation in the size of the cash transfers had little effect on these outcomes. In a companion study, Baird et al. (2013) examined impacts of the same program for girls not in school when the program began, many of whom returned to school due to the program. They find a much larger impact on

[13] A third paper by these authors (and two others), Behrman et al. (2012), examined the short-term effects of the *PROGRESA* program on children in urban areas. After 1 year of program exposure both boys and girls who were aged 6–20 in 2002 had 0.11 higher grades completed; these effects are highly statistically significant.

terms enrolled over the 2-year period, an increase of 2.35 that is highly significant. This is equivalent to almost 1 year of schooling over a 2-year period.

Even more recently, Mo et al. (2013) estimated the impact of a CCT program on students in grade 7 in rural China. Households were offered about US$70 if their children had maintained an 80% or higher rate of attendance over one semester (4–5 months) of school. The authors find that the program reduced the dropout rate by 8%.

An interesting study by Barrera-Osorio et al. (2011) examined the impacts of three different versions of a CCT program in Colombia on daily attendance and enrollment among secondary school students. The first version was a standard CCT program that provided about $15 per month to students' families conditional on a daily attendance rate of 80% or higher in the previous month. The second version imposed the same conditionality but reduced the monthly payments to $10 per month and put $5 per month into a bank account that was made available around the time when students are preparing to enroll in the next year of schooling, which coincides with the time of year that educational related expenses are needed. Finally, the third version also provided payments of $10 per month for regular attendance but then put about $5 per month into a fund that was made available 6 years later, at the time of graduation from secondary school, but only if the student had enrolled in tertiary education. If the student did not enroll in tertiary education he or she had to wait another year before the money was made available. The objective was to provide an incentive to enroll in tertiary education.

All three versions of this intervention led to statistically significant increase in the daily rate of attendance. The first two interventions increased daily attendance by about 3 percentage points, while the third increased it by 5.6 percentage points. The first version, with no "forced savings," had no statistically significant effect on the enrollment rate. The second and third interventions both increased the enrollment rate by about 4 percentage points, increases that were statistically significant. These results suggest that adding a "forced savings" component to a CCT program could increase enrollment at the secondary and tertiary levels by a much larger amount than a CCT without that feature, and more generally that CCT programs can be made more effective by changing the design of those programs. These results also indicate that households may have difficulties committing to long-term investments in education, and that making "commitment devices" available leads to greater educational investments.

Colombia also has a national CCT program; it provides monthly cash grants to parents of children in primary school ($7 per month) and secondary school ($14 per month), conditional on a monthly attendance rate of 80%. The program, which started in 2002, was examined by Attanasio et al. (2010). Using difference-in-differences estimation, these authors found small (1.4 in urban areas and 2.8 in rural areas) but statistically significant percentage point increases in enrollment rates for children aged 8–13 years old, as well as larger (and also statistically significant) impacts for children aged 14–17 (4.7 percentage points for urban areas, and 6.6 for rural areas). Baez and Camacho (2011) also studied the impact of this program on time on school in Colombia, though they focused on more long-term impacts. They found that the program increased the probability of finishing secondary school by 3.9 percentage points.

Two papers have examined the impact of Nicaragua's CCT program, which focused on children in primary school, on time in school. Gitter and Barham (2008) examined the short-run impact, before the control group was allowed to participate in the program in 2003. They found that the program increased the current enrollment of 7–13-year-old children by 16.6 percentage points. Barham et al. (2013) examined the longer run effect of the program on boys; girls were not included in the study since their enrollment rates at ages 9–12 were quite high. Note also that this study had no pure control group; instead one set of randomly selected communities had the program for 5 years (2000 to 2005) while the other set had the program for only 2 years (2003 to 2005). They found that the program increased the number of grades attained by one-half (0.50).

There is only one high-quality study of a CCT program in a South Asian country, that of Chaudhury and Parajuli (2010) on Pakistan. The program targeted girls in grades 6–8. Their families received about $3 per month if their daughters' monthly attendance rate was 80% or higher. The evaluation was not based on an RCT, but instead uses difference-in-differences and regression discontinuity methods. The authors find that the program increased girls' enrollment by 8.7% (a 4 percentage point increase on a base of 43%).

Brazil's CCT program, originally called *Bolsa Escola* but now known as *Bolsa Familia*, is likely the largest such program in the world. At the time of the evaluations discussed below, it was targeted at 6–15-year-old children in poor households (it was later expanded to include older teenage children). Conditional on their child having an 85% attendance rate, parents received about $8 per month per child for up to three children. Unlike Mexico's *PROGRESA* program, Brazil's *Bolsa* program was not implemented as an RCT.

Two recent studies have examined Brazil's CCT program. De Janvry et al. (2012) applied difference-in-differences methods to household survey data; they found that the program reduced the dropout rate by 9.6 percentage points. Glewwe and Kassouf (2012) also used difference-in-differences estimation, but on school level data. They find that the Brazil program increased enrollment by about 3 percentage points and reduced the dropout rate by 3 percentage points. The likely cause of the difference in these two papers' estimated reductions in the dropout rates is that de Janvry et al. examined only four states in the poor Northeast region of Brazil, while Glewwe and Kassouf used nationwide data; the impact of the program is likely to be much larger among populations with higher poverty rates.

Honduras implemented a CCT program in the early 2000s for children aged 6–12 who were in grades 1–4. Conditional on daily attendance of 85% or higher, the mothers of these children received monthly payments of about $6. The program was implemented as an RCT. Galiani and McEwan (2013) found that it increased the enrollment rate by 8.3 percentage points, an estimate that is highly statistically significant.

Two recent studies have examined CCT programs in Cambodia. Barrera-Osorio and Filmer (2013) used an RCT to examine the impact of a standard CCT program that provided poor households (as measured by self-reported household characteristics) $10, twice per year, conditional on enrollment and maintaining passing grades. They found that the program increased the probability that students reach grade 6 increased by 17 percentage points, and more generally that their grades completed increased by 0.33 grades. They also examined a variant of the CCT program which targeted students in grades 4–6 if they scored well on a test at the end of grade 3.[14] This variant also provided $10, twice per year, to scholarship recipients who stayed enrolled, attended regularly, and maintained passing grades, all up to the end of grade 6. They find that this program increased students' probability of reaching grade 6 by 12 percentage points and increased the probability of an additional grade by 18 percentage points.

Filmer and Schady (2014) examined the medium-term impacts of a different Cambodian CCT program that provided scholarships to low-income students who graduated from primary school (finished grade 6). The scholarships paid for costs of schooling for grades 7, 8, and 9, conditional on staying in school and not repeating a grade. They were provided to approximately 20% of students, those who were deemed most at risk of dropping out based on information provided by those students. This estimation was done using regression discontinuity methods, taking advantage of the fact that eligibility for the program was determined by the dropout risk score; those whose risk score was just high enough to make them eligible for the program were compared to those whose score was slightly too low to make them eligible. Five years after the scholarships were offered, they find that the program increased the average number of grades completed by a statistically significant 0.6 years.

Two recent studies have examined the impact of cash transfer programs that did *not* condition the transfers on students' educational outcomes. The first is an *unconditional* cash transfer program in Malawi that was examined by Baird et al. (2011), the results of which are summarized in the sixth line of Table 4. Among girls who were already in school when the program started, the number of terms that they were enrolled over the next 2 years increased from 4.79 to 5.02, a small but statistically significant increase. The second is a "labeled" cash transfer program in Morocco that was examined by Benhassine et al. (2015). It provided monthly cash payments of $8 to $13 to primary

[14] The authors describe this as a merit-based scholarship program, but it differed from the merit-based scholarship programs discussed below because receipt of the scholarships was based on academic performance before the program started, so the program itself provided no incentive for students to increase their academic performance. Thus we view this program as a CCT program targeted at students whose past performance was relatively high.

school students' parents, with higher amounts for higher grades. The payments were "labeled" as assistance for costs of education, but there was no formal requirement that students enroll or attend school regularly for their parents to receive the payments. The results are summarized in the seventh line of Table 4. The authors find that the program increased the enrollment rate by 7.4 percentage points (average over boys and girls); this estimated impact is statistically significant. It did not matter whether the mother or father received the payment.

A final way to, in effect, transfer cash to parents whose children attend school is simply to reduce school fees. Indeed, in the past 10–20 years many developing countries, especially in Sub-Saharan African, have eliminated school fees at the primary level. There are two high-quality studies that examined the effect of such a program on students' time in school. The first is that of Borkum (2012). He used a regression-discontinuity approach to examine the impact of eliminating school fees in South African primary and secondary schools that served poor populations. Unlike the results for CCT programs, he found no effect of the elimination of school fees on enrollment at either the primary or the secondary level.

The second is a "scholarship" program that was not merit-based in China, which was evaluated by Yi et al. (2014). The program promised financial scholarships to pay for upper secondary school (either the vocational track or the academic track) to poor students in Shaanxi and Hebei provinces who were currently in lower secondary school (grades 7 and 9). There is no condition on this financial assistance other than being admitted to upper secondary school. Thus, this is essentially a fee reduction program. For students in grade 9, the intervention increased the rate of entry into upper secondary by 7.9 percentage points, which was statistically significant, but for those in grade 7 the increase was only 3.0 percentage points and not statistically significant.

Almost all of the assessments of cash transfer programs, both conditional and unconditional, have examined the impacts of those programs on students' time in school. Yet a few of the more recent studies have also examined their impact on student learning as measured by test scores. As shown in rows 3–6 of Table 5, some — but not all — cash transfer programs have also increased student learning, Seven high-quality studies have produced seven estimates of the impact of CCT programs on student learning. Three of these studies have found positive and statistically significant impacts on learning outcomes, while the results for the other four are statistically insignificant.

The first published study of the impact of a CCT program on test scores is that of Baird et al. (2011). As explained above, this CCT program focused on 13–22-year-old girls who were in school when the program started. The authors find that, for these girls, 2 years of exposure to the program increased English test scores and math scores by 0.14 and 0.12 standard deviations of the distribution of students' test scores (hereafter denoted as σ), respectively, both of which are statistically significant. The size of the transfers had little effect on these impacts. The follow up study by Baird et al. (2013), which (as explained above) focused on girls not in school when the program started (some of whom returned to school while others did not), found that 2 years of exposure to the program increased English test scores by 0.13σ and math scores by 0.16σ, both of which are statistically significant. They did not examine whether these impacts varied by the size of the transfer.

The Baez and Camacho (2011) study also examined the impact of the Colombia CCT program on student learning. The estimates are from 2009, and since the program began in 2002 they are long-run estimates of the impact, although the years of exposure varied from 2 to 7 years, depending on the age of the child. The overall impact on math scores was statistically insignificant while that for Spanish was negative and significant at the 10% level (-0.05σ). The average impact over these two subjects (see Table 4 of that paper) was statistically insignificant.[15] The lack of a positive impact, and perhaps even a negative impact on Spanish scores, could reflect some selection bias in that weaker students who did not receive CCT payments may have dropped out of school and so were not tested (which would increase the average test scores of those who remained in school); indeed, as discussed above this program greatly increased the probability of completing high school.

Mo et al. (2013) also examined the impact of the CCT program in rural China on the test scores of students in grade 7: the point estimate was very small and statistically insignificant. This may also reflect the fact that only students who stayed in school were tested; those who dropped out, who were more often from the control schools (as explained above), very likely had relatively low test scores, so if they had been tested it is possible that the results would have shown a positive impact of the program on learning.

The CCT program in Cambodia evaluated by Barrera-Osorio and Filmer (2013) also estimated the impact of that program on test scores. After 3 years of operation the impact on test scores was not significantly different from zero. Note that the test score data used in this study were obtained by testing children in their homes, which avoids the downward bias that could occur if only children who are still in school are tested.

The long-term impact of exposure to the Nicaraguan CCT program on (former) primary school students' test scores was examined by Barham et al. (2013). The children were exposed to the program when they were 9–12 years old, but the test score data were collected (for boys only) in 2010, 10 years after the start of the program. The results show that, 10 years after the start of the program, average test scores were 0.20σ (0.23σ for home language and 0.17σ for math) higher for the boys who were exposed longer (5 years instead of 2 years) to the CCT program, and these impacts are statistically significant.

[15] This average result is the one shown in Table 5; to minimize giving larger "weight" to studies that report many estimates for different population subgroups, Tables 4–11 use aggregate estimates across tests and grades whenever possible.

Finally, the study by Filmer and Schady (2014) that examined the long-term effects of a Cambodia program that provided (conditional) scholarships to low-income students who had graduated from primary school (finished grade 6) also examined student learning. The estimated impact on test scores 5 years after the program began (by which time many of the students were in grade 11) was small and not statistically significant. As in the Barrera-Osorio and Filmer (2013) study of Cambodia, test score data were obtained by testing children in their homes, avoiding downward bias that could occur if only children still in school are tested.

Recall that two studies have examined the impacts of other types of cash transfer programs. Baird et al. (2011) examined the impact of unconditional cash transfers in Malawi, While those transfers did increase enrollment, as explained above, they had no significant impact on students' test scores among the girls who were already in school when the program began. Lastly, the Moroccan "labeled" cash transfer program examined by Benhassine et al. (2015) had no effect on student test scores even though it did increase students' time in school.

The "scholarship" program in China evaluated by Yi et al. (2014), which was essentially a fee reduction program, also examined the impact of that program on student learning. As seen in seventh line of Table 5, this program had no significant impact on students' test scores.

To summarize the results for CCT programs, for which there is now a large amount of evidence, they almost always increase time in school, and in some cases they also increase student learning as measured by test scores. The significantly positive effects on time in school were found in all nine countries studied: Brazil, Cambodia, China, Colombia, Honduras, Malawi, Mexico, Nicaragua, and Pakistan. Regarding student learning, in two of five countries (Malawi and Nicaragua) estimates done several years after the end of the program show that CCT beneficiaries had higher cognitive skills, as measured by test scores. In contrast, in three other countries (Cambodia, China, and Colombia), there was no evidence of impacts on learning. However, data shortcomings, in particular lack of data on students who dropped out of school from the control groups, may downwardly bias the estimated impact on learning in the China and Colombia studies, but this potential bias does not apply to the two Cambodia studies.

Finally, one important limitation of the experimental evidence on CCT programs is that they typically do not allow us to evaluate alternative monetary values of the transfers, which is an important input into policy decisions. One approach that addresses this limitation is to use household and school level data to estimate a structural model of households' educational decisions, and to validate the model using the experimental CCT evaluations. While doing so is rather complicated and requires assumptions that may not hold, it has the advantage of enabling prediction of the impact of variation in programs, including variants that have not yet been implemented, on students' educational outcomes. Examples of such an approach are Todd and Wolpin (2006) and Attanasio et al. (2012).

4.1.3 Scholarship Programs

A further source of inefficiency on the demand side may come in translating school enrollment and attendance into learning outcomes. Since the returns to student effort typically take many years to realize, students who are present-biased may not exert adequate effort in school.[16] Thus, policies that help make the rewards from exerting effort to acquire human capital more immediate may be a promising option for improving student learning outcomes. A common way to do this (that may also alleviate resource constraints, as discussed above) is to provide merit-based scholarships to students, which provide tangible intermediate rewards on the basis of levels or improvements in student performance.

Two high-quality studies, both from Kenya, have examined the impact of merit-based scholarships on students' time in school. As seen in the ninth line of Table 4, three of the four estimates from these two studies find significantly positive effects (the fourth estimate is statistically insignificant). The first such study was that of Kremer et al. (2009), who conducted an experimental evaluation of a scholarship program in rural Kenya that targeted girls who were in grade 6 (Kenya's primary schools teach students from grade 1 to grade 8). At the beginning of the school year, girls in grade 6 were told that those who scored in the top 15% on end-of-year exams would be given, for each of the next 2 years (grades 7 and 8), an amount of money equal to $6.40, which was enough to cover school fees, and in addition their parents would be given an amount of money equal to $12.80. The second study in Kenya, by Friedman et al. (2011), was a follow-up of this 2009 study; it examined the educational outcomes of the same girls 4–5 years after the original program started (about 2 years after the program had ended). The earlier study found that the program significantly increased participation (daily attendance, where girls who leave school have a zero attendance rate) by 3.2 percentage points. The more recent study found that the program had significant positive impacts on enrollment in secondary school (8.6 percentage point increase) and current enrollment in any school (7.9 percentage point increase), but not on grades completed.

Consider next the impact of scholarship programs on student learning. Four studies, all RCTs, have examined the impact of providing scholarships based on student learning as measured by test scores. Their results are shown in the seventh line of Table 5. The findings are almost unanimous: five of the six estimates are significantly positive, while the sixth was statistically insignificant.

[16] An extensive literature in psychology documents that "fronto-cortal" brain development (that is associated with the ability to make long-term trade-offs) is completed only by the age of 18, and that youth and children below this age are less able to evaluate the long-term trade-offs that the neoclassical economic framework assumes will be made to reach optimal decisions regarding human capital investments. As a practical application of this point, this is partly why the juvenile justice system (for offenders under 18) is separate from the adult justice system.

The two Kenya studies, which were described above, found significantly positive effects. More specifically, Kremer et al. (2009) found a 0.27σ increase in scores on a grade 6 year-end exam, and Friedman et al. (2011) found that test scores were 0.20σ higher 4–5 years after the program started (2 years after it ended).

A more recent study is that of Blimpo (2014), who examined a program in Benin that randomly assigned 100 secondary schools to a control group or to one of three different types of scholarships: scholarships based on individual-level performance with respect to a set goal, with no limit on the number of scholarships offered; scholarships based on average performance for (randomly assigned) teams of four students, again with respect to a set goal and no limit on the number of scholarships; and a "tournament" in which 84 teams of four students each (randomly assigned) from 28 schools competed for a large prize that was given only to the three top performing teams. For the first two types, the payments were $10 per person ($40 for a team of four) for a relatively low level of performance, and $40 per person ($160 for a team of four) for a high level of performance. For the third type, the prizes were much larger, at $640 for each of the top three teams. All three types of incentives had similar (and statistically significant) impacts, increasing grade 10 test scores by 0.24σ to 0.28σ.

Finally, a scholarship program in China, evaluated by Li et al. (2014), was also based on a tournament. In one version (individual incentive), groups of 10 low-performing students would compete with each other in terms of improvement over time on their test scores. The top student received about $13, the second and third about $6, and the rest each received about $3. In the other version (peer incentive), each of ten low-performing students was paired with a high-performing student; prizes for the low-performing students were the same as in the first version, but in this case the high-performing student also received a similar reward in order to encourage that student to assist the low-performing student with whom he or she had been paired.[17] The individual incentive intervention in China had no statistically significant impact. However, the authors find that combining student incentives with peer tutoring (where academically higher-achieving students were paired with lower achieving ones and both students were rewarded for improvements) increased the test scores of the weaker students by 0.27σ. Thus, in some cases it is possible that student incentives on their own may not be effective unless also accompanied by pedagogical support.

4.1.4 Other Household Interventions That Increase the Demand for Schooling

This section presents evidence on the remaining interventions shown in Tables 4 and 5. This set of interventions is diverse, but they have two common characteristics. First, they generally are intended to increase the demand for schooling, although in some cases

[17] This version could be interpreted as a hybrid intervention that involves peer tutoring as well as merit-based scholarships.

rather indirectly. Second, they are implemented directly with households rather than in schools.

One way to make schooling more effective, which should in turn increase the demand for schooling, is to provide information to mothers on how to develop their child's learning. Such an initiative was examined by Banerji et al. (2013), who also examined the impact of a mother literacy program and the combination of both programs. As seen in lines 10–12 of Table 4, neither of these two interventions, nor their combination, had an impact on children's time in school (as measured by enrollment and daily attendance). Turning to the effects on student learning, which are summarized in lines 9–11 of Table 5, the lessons on child learning led to small but statistically significant increases in the test scores (average over literacy and mathematics) of students in grades 1–4 (0.04σ), but there was no significant impact of the literacy class alone. Interestingly, a somewhat higher impact (0.06σ) on student test scores was found for women who took both classes. These impacts after 1 year of the program are relatively small, and there were no significant impacts on children's enrollment rates.

In a similar vein, Handa (2002) used a difference-in-differences approach to examine the impact of an adult literacy program in Mozambique on their children's enrollment rate. As seen in line 13 of Table 4, he found that the program had no effect. Yet a longer term program of the kind evaluated by Banerji et al. (2013) may have larger effects, and adult literacy is an outcome with positive value in itself beyond the impact on child outcomes.

The nonavailability of female sanitary products is often posited as a factor inhibiting female school participation after puberty. Oster and Thornton (2011) conduct an RCT to evaluate the impact of an intervention that provided female sanitary products to girls of secondary school age in Nepal, but found that the program had no effect on the daily attendance of secondary school girls (as seen line 14 of Table 4). This may have been because only 1% of girls actually stated that the lack of sanitary products was a binding constraint to school participation, suggesting that other constraints may have been more first order.

A more successful intervention for girls of secondary school age was a program in India that offered their families funds to purchase bicycles which their daughters could ride to attend secondary school. Muralidharan and Prakash (2013) used a difference-in-differences strategy (using triple and quadruple differences) to estimate the impact of this program on enrollment. They find that this program increased secondary school enrollment by 5.2 percentage points. For girls who lived more than three kilometers from the nearest secondary school, the impact was about 9 percentage points. These are very large impacts given that the initial enrollment rate for these girls was only 17.2%.

A final household-based intervention to increase the demand for schooling in Table 4 is a program that matched remittances sent by Salvadoran migrants in the United States to students of their choosing in El Salvador if those remittances are committed to educational purposes. Using an RCT, Ambler et al. (2014) found that when the match amount is three times the amount of remittance, the program increased enrollment in a private school by a statistically significant 10.9 percentage points. Note, however, that the overall increase in enrollment of 3.1 percentage points was not statistically significant.

To summarize this section, there is broad and consistent evidence that CCT programs successfully increase students' time in school. There is also evidence that scholarship programs may increase time in school and suggestive evidence that information-based interventions could serve this purpose. Turning to the impact of demand-side interventions on students' test scores, merit-based scholarships appear particularly effective, and in some cases CCT programs have also had positive impacts on test scores. Overall, the results suggest that interventions that increase the benefits of attending school (such as CCTs) increase time in school, while those that increase the benefits of higher effort and better academic performance (such as merit scholarships) improve learning outcomes. These results are consistent with theory and suggest that the outcomes that are explicitly rewarded by policies are more likely to be changed by those policies, and highlight the importance of theoretically informed program design.

4.2 Inputs

The majority of the costs of running a school system are typically input costs — including the costs of building schools, and the costs of operating schools (with teachers being a key input and teacher salaries typically being the largest component of education expenditure). Thus, it is no surprise that many studies have examined the impact of inputs, broadly defined, on students' educational outcomes. However, credibly studying the impact of school inputs is quite challenging given the limited number of studies with exogenous variation for these inputs. Thus, many studies that have examined inputs were deemed to not be credible enough to include in this review. As a result, many important categories of input interventions, such as school infrastructure, teachers' education levels, and teacher training, are not included in this review, and there are important evidence gaps in this area.

The impacts of inputs on students' time in school are shown in Table 6, while Table 7 provides results for the impacts of inputs on student learning, as measured by test scores. We organize our discussion of school inputs by classifying them into six broad categories: access to schools, pedagogical materials and facilities, teacher quality and quantity, provision of food, provision of medical services, and composite provision of better resources.

4.2.1 Access to Schools

Perhaps the most obvious school "input" is an actual school. More generally, access to schools affects both students' time in school and how much they learn. Access can be increased by: (1) Building new schools, which reduces the distance to the nearest school; (2) Increasing the hours per day, or the days per year, that schools are open; and

Table 6 Summary of impacts on time in school of school inputs

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Interventions that increase access to schools					
Building new schools					
RCTs	0 (0)	0 (0)	0 (0)	3 (2)	2
Other high-quality studies	0 (0)	0 (0)	0 (0)	3 (3)	3
Providing school uniforms (RCT)	1 (1)	0 (0)	0 (0)	0 (0)	1
Pedagogical materials and facilities					
Textbooks (all RCTs)	0 (0)	0 (0)	2 (2)	0 (0)	2
Provision of libraries (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Multilevel teaching materials (RCT)	1 (1)	0 (0)	0 (0)	0 (0)	1
Multilevel teaching materials and parent-teacher partnership (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Teacher quantity and quality					
Extra teacher/materials (high quality)	0 (0)	0 (0)	0 (0)	2 (1)	1
Provision of food					
School meals					
RCTs	0 (0)	0 (0)	2 (2)	1 (1)	3
Other high-quality studies	0 (0)	3 (2)	1 (1)	0 (0)	2
Take-home rations					
RCTs	0 (0)	0 (0)	0 (0)	1 (1)	1
Other high-quality studies	0 (0)	0 (0)	1 (1)	0 (0)	1
School feeding/parent-teacher partnerships (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Medical services					
Deworming medicine (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Health insurance (high quality)	0 (0)	0 (0)	1 (1)	0 (0)	1
Large-scale provision of resources					
School infrastructure investment (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Support circles (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

Table 7 Summary of impacts on test scores of school inputs

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Interventions that increase access to schools					
Building new schools					
RCTs	0 (0)	0 (0)	0 (0)	2 (1)	1
Other high-quality studies	0 (0)	0 (0)	0 (0)	1 (1)	1
Hours per school day (high quality)	0 (0)	1 (1)	0 (0)	3 (2)	2
Pedagogical materials and facilities					
Textbooks (all RCTs)	0 (0)	2 (1)	1 (1)	0 (0)	2
Flipcharts (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Provision of libraries (RCT)	1 (1)	1 (1)	0 (0)	0 (0)	1
Multilevel learning materials (RCT)	0 (0)	1 (1)	0 (0)	2 (1)	1
Multilevel teaching materials and parent-teacher partnerships (RCT)	0 (0)	0 (0)	0 (0)	3 (1)	1
Teacher quantity and quality					
Pupil-teacher ratio					
RCTs	0 (0)	1 (1)	0 (0)	0 (0)	1
Other high-quality studies	3 (2)	1 (1)	0 (0)	0 (0)	2
Provision of food					
School meals					
RCTs	0 (0)	0 (0)	0 (0)	3 (2)	2
Other high-quality studies	0 (0)	2 (1)	1 (2)	1 (1)	2
Take-home rations (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
School feeding /parent-teacher partnerships (RCT)	0 (0)	0 (0)	1 (1)	2 (1)	1
Medical services					
Deworming medicine (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1
Iron supplements (all RCTs)	0 (0)	1 (1)	2 (2)	1 (1)	2
Provision of eyeglasses (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Large-scale provision of resources					
Support circles (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Attending an elite public school (both are other high-quality studies)	0 (0)	0 (0)	1 (1)	1 (1)	2
Infrastructure/materials/ teacher training (high quality)	0 (0)	0 (0)	2 (1)	2 (1)	1

Continued

Table 7 Summary of impacts on test scores of school inputs—cont'd

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Unexpected school block grant (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Expected school block grant (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Incentivized community block grant (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1
Nonincentivized community block grant (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

(3) Removing barriers to enrollment, such as the lack of a school uniform (which are required in many developing countries).[18] This section examines all three types of interventions, starting with the building of new schools.

Five high-quality studies have examined the impact of building new schools on time in school. Each of these five studies examined a different country, so evidence is available from Afghanistan, Burkina Faso, Indonesia, Mozambique, and Pakistan. Building new schools reduces a very important *indirect* cost of attending school, the distance to the nearest school. More time spent traveling to school is time lost for other activities, including work, and greater distances may also lead to direct transportation costs and worries about safety. The first two rows of Table 6 show that there are two studies based on RCTs and three studies that use other high-quality estimation methods. All six estimates from these five studies show significantly positive impacts from building new schools on students' time in school.

The earliest of these five studies is that of Duflo (2001), who used difference-in-differences estimation to examine the impact of a massive school construction program in Indonesia in the 1970s on years of schooling in that country (the paper also examined the impact of years of schooling on wages, but that is not the topic of this chapter). She focused on boys born between 1950 and 1972, and found that an additional school built per 1000 school-age children increased years of education by 0.19 years.

The second study is that of Handa (2002), who used difference-in-differences estimation to estimate the impact of the construction of new primary schools in Mozambique on children's probability of being enrolled. He finds that the marginal probability of enrollment of boys and girls increased by 0.3 percentage points for each new primary school built within an "administrative post" area (administrative posts are relatively large

[18] Of course, tuition and other required school fees can effectively limit access as well, while cash transfer programs and scholarships can expand access. These interventions were discussed above in Section 4.1.

areas, with on average 21 primary schools). While this estimated impact is small, it is highly statistically significant (p-value < 0.01).

Alderman et al. (2003) conducted an RCT in Pakistan that provided funding to construct new (or support existing) private girls' primary schools; the support of existing schools made them affordable to poor families. In urban areas, the program increased the enrollment rate for girls by 25 percentage points, while in rural areas the enrollment rate increased by 15 percentage points.

A more recent study, based on an RCT conducted in Afghanistan, is that of Burde and Linden (2013). They examined the impact of the opening of primary schools on children of primary school age in rural villages of Afghanistan that did not have a school. In Ghor province, where this intervention took place, only 29% of families lived within 5 km of a primary school in 2007. The authors find very large impacts of that program on enrollment rates of children of primary school age in those villages. The program increased the enrollment rate by 51.5 percentage points for girls and by 34.6 percentage points for boys. The higher impact on girls likely reflects that they are often not allowed to travel to neighboring villages, some of which may have schools.

Finally, Kazianga et al. (2013) evaluated the impact of providing "girl friendly" schools in rural villages in Burkina Faso. These schools have amenities that are particularly attractive to girls, such as sources of clean water and separate latrines for boys and girls. Overall, opening these schools increased the enrollment rate of all children (average over boys and girls) by 18.5 percentage points, a very large effect that is highly statistically significant. This is the estimate included in Table 6; separate estimates indicate that the impact is 16.3 percentage points for boys and 21.9 percentage points for girls.

Together, these five studies from five different countries show that building schools in communities that do not have them can lead to very large increases in school enrollment. While the vast majority of primary school-age children in developing countries live very close to a primary school, for the small percentage who live much farther away building a school in their local community will likely have a very large effect on their probability of being enrolled.

Two of these five high-quality studies that examined the impact of building new schools on time in school also examined the impact of that intervention on student learning. As seen in the first two rows of Table 7, these two studies have produced three estimates of this impact. While the number of studies, and the number of estimates, is very small, the findings are unanimous: all three estimates are statistically significant and show that constructing new schools increases learning among children of school age.

Burde and Linden (2013) find that constructing new schools in rural Afghanistan had a large impact on test scores over a period of about 6 months, generating increases of 0.66σ for girls and 0.41σ for boys, both of which are highly statistically significant. These estimates include all children in those villages, not just those who are enrolled in school.

Similarly, Kazianga et al. (2013) find that building new "girl friendly" primary schools in rural Burkina Faso in villages that previous had no primary school increased test scores by 0.41σ. This result is an average over French and mathematics exams, and over girls and boys.

Perhaps the second most obvious "input" that affects students' educational outcomes (the first being providing an actual school) is the amount of time that children spend in school. For any given year, this time could be lengthened by increasing the number of hours per day that the school is open or by increasing the number of days per year that the school is in session. We found no high-quality studies that examined the impact of a longer school day on time in school, perhaps because it is rather obvious that longer school days by definition increase time in school, assuming that the rate of student absenteeism does not dramatically increase.

However, two high-quality studies have produced four estimates of the impact of an increase in the length of the school day on student learning, as seen in the third line of Table 7. Of these estimates, three are positive and significant and one is insignificant. The positive and significant estimates come from studies of Chile (Bellei, 2009) and Ethiopia (Orkin, 2013), both of which are based on difference-in-differences estimation. The one insignificant, estimate is also from the Orkin study. Overall, the evidence, while based on only two studies, is generally supportive of the common sense notion that longer school days increase student learning.

One high-quality study examined a third and final way to increase access to schools: provision of free school uniforms. Hidalgo et al. (2013) present results from an RCT in Ecuador that provided free school uniforms to poor primary school students in urban Ecuador. Surprisingly, they find that provision of free uniforms decreased daily attendance by 2 percentage points. Unfortunately, their data do not allow them to determine why provision of uniforms had this unexpected effect.

4.2.2 Pedagogical Materials and Facilities

Two RCT studies from Sub-Saharan Africa have examined the impact of textbooks on students' educational outcomes: Glewwe et al. (2009) and Sabarwal et al. (2014). These studies examined textbook distribution programs in Kenya and Sierra Leone, respectively. The results were surprising, and thus they merit further discussion.

The impact of these two programs on the time that students spend in school is examined in the fourth row of Table 6. Both of the two estimates (one from each of these two studies) are statistically insignificant; providing textbooks did not decrease dropping out or grade repetition in Kenya and did not increase daily attendance in Sierra Leone. However, in Kenya if one focuses on students in grade 8, who are relatively strong students (about half of students who start in grade 1 drop out before reaching grade 8), the program increased the probability that these students finished primary and enrolled in secondary school (not shown in Table 6, which shows only averages over all grades).

While textbooks may not increase students' time in school, one would expect that they would increase student learning. This is examined in the fourth row of Table 7, which summarizes the results of three estimates from the Kenya and Sierra Leone studies. Surprisingly, none of these three estimates is statistically significant. These findings are quite unexpected given that one would think that textbooks would have a strong impact, or conversely that lack of textbooks would have a strong negative effect, on student learning.

The study of Kenya by Glewwe, Kremer, and Moulin investigated the reasons behind the unexpected insignificant findings. The authors found that the textbooks provided, which were the official government textbooks, were too difficult for the average child to read in the region of Kenya (Busia and Teso districts) where the study took place. Indeed, when the sample is restricted to the top 20% of students (and, in some regressions, to the top 40% of students), as measured by their preintervention test scores, the textbooks did improve students' learning outcomes (not shown in Table 7).

For the study of Sierra Leone, by Sabarwal et al. (2014), the reason for the unexpected finding of no impact is rather obvious: few of the textbooks reached the students. In this RCT, the program was implemented by the Ministry of Education, and there was little follow up action to encourage the teachers and school administrators to distribute the textbooks to students; instead, most of the textbooks were kept in storage. The authors present evidence suggesting that school administrators stored most of the textbooks because they were unsure whether textbooks would be provided in future years. Overall, this small amount of evidence suggests that textbooks can have a positive impact when they are actually provided to students, and when the textbooks are at the appropriate level for those students, which was the case for the top students in Kenya. But if textbooks are too difficult, which was the case for most students in Kenya, or are never provided to the students, which was the case in Sierra Leone, they will have little or no effect on both time in school and student learning.

Another study from the same area of Kenya by the same authors (plus another author), Glewwe et al. (2004), used an RCT to examine the impact of the provision of "flip charts" (sets of large posters to place on an easel or hang on a wall) on students' test scores in Kenya. As seen in the fifth line of Table 7, the authors find that the provision of flip charts had no significant impact on student learning, and they note that nonexperimental estimates of the impact of this intervention find significant positive effects.

We found only one high-quality study of the impact of libraries on students' educational outcomes, a study of the provision of school libraries in India by Borkum et al. (2012). The authors used an RCT to examine the provision of both "in school" libraries and traveling libraries. As seen in the fifth line of Table 6, there was no effect of either library on students' daily attendance (which was already quite high, at around 90%). Turning to learning outcomes (sixth line of Table 7), "in school" libraries had no effect on students' language scores and the traveling libraries had an unexpected negative effect

(-0.22σ) on students' language scores. Thus in this setting school libraries did not lead to improved educational outcomes.

The last two pedagogical material interventions are from the same study, which examined primary schools in the Philippines. Tan et al. (1999) experimentally investigated the impact of the provision of "multilevel learning materials," as well as an intervention that combined those materials with "parent-teacher partnerships." The multilevel learning materials intervention by itself significantly reduced the probability of dropping out, but had no significant impact on dropping out when combined with parent-teacher partnerships (lines 6 and 7 of Table 6). The impact on test scores (in Filipino, math and English) was significantly positive for two of the three tests when only multilevel materials were provided, and for all three tests when those materials are combined with parent-teacher partnerships (lines 7 and 8 of Table 7). Of course, it is unclear what the relative contributions of two components are. Even the multilevel learning materials intervention by itself had many components (several different types of learning materials), so it is not clear which components led to increased student learning.

4.2.3 Teacher Quantity and Quality

Teachers vary in many ways, but we found no high-quality studies that have examined the impact of teacher characteristics on student learning or time in school.[19] However, one high-quality study has examined the impact of providing an extra teacher to very small primary schools in India, and three studies have examined the impact of variation in the pupil-teacher ratio. All four of these studies can be thought of as attempts to change the "quantity," as opposed to the "quality," of teachers. Indeed, a very important evidence gap is the lack of well-identified studies on the impact of teacher education and training on learning outcomes in developing countries.

Only one of these studies focused on the impact of the quantity of teachers on students' time in school. Chin (2005) used a DD approach to evaluate a program that provided extra teachers *and* additional educational materials (including blackboards) to very small primary schools in India. As seen in the eighth row of Table 6, she found that the program significantly increased students' primary school completion rates (by 1–2 percentage points), but it is not possible to determine how much of this effect is due to the extra teacher and how much is due to the additional educational materials.

Turning to the pupil-teacher ratio, three high-quality studies have produced five estimates of the impact of the pupil-teacher ratio on student learning. These results are summarized in the ninth and tenth rows of Table 7. Intuitively, one would expect that increased pupil-teacher ratios reduce learning because larger classes reduce opportunities for teachers to give individual attention to students. Indeed, three of the five estimates,

[19] We examine the impact of hiring contract teachers in Section 4.4.5 below. This is not a teacher characteristic per se, but rather pertains to the type of contract that a teacher has.

from two different studies, find a significantly negative effect. On the other hand, two of the five estimates, also from two studies, including the only RCT study, are statistically insignificant.

The two papers that produce the expected negative finding are those by Urquiola (2006) and Urquiola and Verhoogen (2009). Urquiola (2006) used regression discontinuity methods to estimate the impact of the pupil-teacher ratio on student learning in Bolivia. In particular, he used the fact that schools with pupil-teacher ratios above 30 can apply to the education authorities for another teacher, and he presents evidence that these schools often do obtain another teacher. He finds that schools that obtain another teacher, which greatly reduces the pupil-teacher ratio, have significantly higher language scores, but the effect on math scores is not statistically significant.

Urquiola and Verhoogen (2009) also use regression discontinuity methods to estimate the impact of class size on student test scores in Chile. The focus of the study is on children in grades K-8 in private schools (about half of students in Chile are enrolled in private schools). The authors found that increased class size has significantly negative impacts on both math and language test scores.

Duflo et al. (2012a, 2015) conducted an RCT in Kenya that randomly assigned some schools an extra contract teacher, and within those schools students were randomly divided into classes which were taught by the current teacher (who were civil-service teachers) and those taught by the contract teacher.[20] For the purpose of identifying the impact of the pupil-teacher ratio on student learning, the classes taught by the current (civil service) teacher can be compared to those taught by same type of teacher in the control schools, which have much larger pupil-teacher ratios. They find that although this reduction in class size led to higher test scores (about 0.09σ), this increase was not statistically significant.

In summary, while the amount of evidence is small, three of the four studies find evidence that reducing the pupil-teacher ratio improves students' time in school and their test scores. Yet the study on India could not distinguish the impact of an additional teacher from the impact of additional learning materials, and the one RCT study did not find a statistically significant impact of class size reductions on student learning. Clearly, more evidence on the size of these effects is needed. It is also important to bear in mind that reductions in class size can be very expensive, which implies that even if they do lead to better students outcomes there may be other interventions that achieve the same results at a lower cost. See Section 5 below for a brief discussion of the costs of some of the interventions discussed in this chapter.

4.2.4 Provision of Food

Child malnutrition is a common problem in many developing countries, and there is a large amount of evidence that well-nourished children have better educational outcomes

[20] The findings of this study on the impact of contract teachers are discussed below in Section 4.4.5.

(Glewwe and Miguel, 2008). The strongest evidence pertains to early childhood malnutrition, but it may also be the case that adequate nutrition during a child's time in school can improve his or her academic performance. Thus many developing countries have implemented programs that provide meals to students, and in some cases to their families as well. This section reviews the evidence on the effectiveness of such interventions based on eight high-quality studies.

The most common type of program that attempts to improve students' nutritional status is the provision of school meals. Lines 9 and 10 of Table 6 show the results from five studies in five different countries (Burkina Faso, Chile, India, the Philippines, and Uganda) that have estimated the impact of school meals on time in school. Of the seven estimates, six are statistically insignificant and one, for the program in Burkina Faso that was evaluated by Kazianga et al. (2012), finds (using an RCT) a significantly positive impact on enrollment of children aged 6–15 years old. This evidence indicates that in most cases school meals do not increase students' time in school. Even if school meals do not increase years of enrollment, one would think that it would increase daily attendance. Yet only one of the five studies, that of Alderman et al. (2012), measured the impact of school meals on daily attendance; the authors found no statistically significant, effect.

Of these five high-quality studies, four have examined the impact of school meals on student learning. As seen in lines 11 and 12 of Table 7, three of the seven estimates from these four studies are statistically insignificant and four are significantly positive. More specifically, McEwan (2013) found statistically insignificant impacts of a school feeding program on math and language test scores among grade 4 students in Chile. In neighboring Argentina, Adrogue and Orlicki (2013) found a small and statistically insignificant impact of school feeding on the mathematics test scores of students in grade 3, but a larger (0.17σ) and statistically significant impact on their language scores. Tan et al. (1999) found significantly positive impacts of a school feeding program in the Philippines on the math (0.25σ) and Filipino (0.16σ) of grade 1 students. Finally, Kazianga et al. (2012) found that school meals increase math scores by 0.10σ in Burkina Faso. These results suggest that school meal programs can — at least in some settings — increase student learning.

Two studies have examined the impact of take-home rations on students' education outcomes. The results on students' time in school are summarized in lines 11 and 12 of Table 6. Adrogue and Orlicki (2013) found no impact of such a program on student attendance in Argentina, but Kazianga et al. (2012) found a significantly positive 4.8 percentage point impact on the enrollment of children aged 6–15 in Burkina Faso. Only the latter program examined the impact of take-home rations on student learning; as seen in line 13 of Table 7, the Burkina Faso program increased students' math scores by 0.08σ, an impact that is statistically significant (p-value < 0.05).

Finally, the Tan et al. (1999) study also examined a program that combined school feeding with "parent-teacher partnerships." This combination had no effect on students' time in school, as seen in line 13 of Table 6, but it did lead to two (out of three) statistically positive impacts on student learning (line 14 in Table 7). However, it is not possible to determine the extent to which this is due to the school meals or to the parent-teacher partnerships.

4.2.5 Provision of Medical Services

A natural extension of the provision of food to improve students' educational outcome is the provision of basic medical services. This section examine four types of interventions for which there are high-quality studies: deworming programs; iron supplementation; health insurance; and the provision of eyeglasses.

Many individuals in developing countries have worm infections of various types (roundworm, hookworm, whipworm, and schistosomiasis). School-age children are particularly vulnerable to these types of infections, and those infections lead to anemia and other problems that may reduce children's attentiveness in school. Miguel and Kremer (2004) implemented an RCT to estimate the impact of providing deworming medicine to primary school students in rural Kenya. As seen in Table 6, this intervention increased students' time in school. More specifically, it reduced their absence rate by about 7–8 percentage points. However, as seen in Table 7, the authors did not find an impact of the program on students' test scores.

However, it is worth noting that a later study (not in the table) examines spillover effects of the Kenyan deworming program on younger siblings of children in treated communities and finds positive effects on the test scores of children who were less than 1 year old when the deworming program was implemented. Ten years after the program, Ozier (2014) finds that these children have large test score gains — equivalent to between 0.5 and 0.8 years of schooling. Further, a long-term follow up study of treated cohorts finds that 10 years after the program, women who were eligible for the program as school girls were 25% more likely to have attended secondary school, and men who were eligible were more likely to have completed primary school (Baird et al., 2015). The study also finds large positive labor market effects and estimates an annual financial return on investment of over 32%. Thus, it appears that the deworming program clearly improved long-term human capital even though there were no short-term effects on test scores.

The other medical services initiative for which there is a high-quality study that examined impacts on students' time in school is a health insurance program in China that was evaluated by Chen and Jin (2012). They used a DD (combined with propensity score matching) estimation procedure. As seen in Table 6, they found no impact of the program on enrollment of children aged 6–16.

Turning to student learning outcomes, there are high-quality studies for two other types of interventions. One of them has some similarities to deworming, namely the provision of iron supplementation tablets. There are two high-quality studies, Luo et al. (2012) and Sylvia et al. (2013), both of which were conducted in China. As seen in

Table 7, there are four estimates from these two studies. Three are statistically insignificant, while the fourth is significantly positive. Together they provide some initial evidence that iron supplements can increase student learning, at least under some circumstances, but there are only two studies and both were conducted in the same country.

The last medical services intervention is the provision of eyeglasses. Glewwe et al. (2014) conducted an RCT to investigate the impact of this intervention on the learning of primary school students in rural China (they did not look at enrollment since it is already very close to 100%). The authors found that the eyeglasses significantly increased average test scores by at least 0.16σ.

4.2.6 Large-Scale Provision of Resources

The last set of input studies is a set of interventions that provided either entire packages of school inputs (broadly defined) or large amounts of money that schools could use to buy the inputs of their choice (or spend in other ways). As will be seen, these interventions are often quite different from each other.

One of the earliest published RCT studies in education is that of Newman et al. (2002), who examined a program in Bolivia that provided funds to schools to make infrastructure improvements. They estimated the impact of this program on the dropout rates of boys and girls in grade 7. As seen in Table 6, the program had no statistically significant impact on those dropout rates.

The other large-scale program for which there is a high-quality study that examined impacts on students' time in school is a program in Malawi that combines a remedial program for students who have dropped out of primary school, interactive radio instruction, a new type of curriculum and pedagogy ("Escuela Nueva") and mobilization of networks of families, friends and neighbors (which are referred to as "circles of support"). This intervention was evaluated using an RCT by Pridmore and Jere (2011). They find that this program significantly reduced the dropout rates (Table 6) but had no effect on the learning outcomes (Table 7) of primary school students in that country.

There are two studies that have examined the admittedly rather general policy of being admitted to an elite public school, which could be interpreted as devoting the resources required to transform a typical high school into an elite high school. Many developing (and developed) countries have elite public schools that restrict admission to the best students in the country, and that have much more qualified teachers and many other types of resources that are not found in a typical secondary school. Places in these schools are highly valued by many parents, and the students of these schools often have very successful careers, but it is not clear that those students' successes were due to their attending those elite schools. Two studies have examined this, one for Romania and the other for Kenya.

Pop-Eleches and Urquiola (2013) examined the effect of "going to a better school" in Romania. They find that students who are able to get into a better, more academically challenging school perform better on graduation tests. This indicates that "better" schools do lead to greater student learning, but it provides no information on which of the many characteristics that make a school "better" are the ones that bring about this improved performance.

In contrast, Lucas and Mbiti (2014) find, using regression discontinuity methods, essentially no evidence that going to an elite high school in Kenya leads to increased learning. The only possible exception is that students who attend such schools appear to do better on Swahili exams, which may reflect use of that language as the common language for communication among students outside of the classroom. Thus they attribute the success of graduates of those schools to characteristics that they already possessed that would have helped them be successful even if they had not attended an elite high school.

In a similar vein, in a paper for which the main contribution is methodological, Chay et al. (2005) examine Chile's P-900 program, which provided infrastructure, materials and teacher training to schools in Chile. They use a regression discontinuity approach to estimate the impact of this program. While two of the four estimates are significantly positive, it is virtually impossible to determine which component of the program led to increased student learning.

Finally, two recent studies have examined the impact of providing block grants, but in different ways. Das et al. (2013) examine a school block grant program in India. The authors show that when schools received large unexpected grants (about $3 per student) students test scores increased by about 0.09σ. However, expected grants had little or no effect, because households decreased spending on education when they knew that the schools that their children were attending would receive the grants.

The second such study is by Olken et al. (2014). It examines a sort of cash transfer program that is targeted at entire villages, rather than households. The "incentivized" transfer provides villages with about $10,000 and the promise of addition funds if the village scores well on 12 different health and education indicators. They also examined a similar but "unconditional" program where villages received the $10,000 without any promise to provide additional funds based on health and/or education indicators. Neither program had any effect on students' test scores.

To summarize the evidence on the impact of standard "inputs" on students' educational outcomes, there is broad evidence that improving school access through building schools increases time in school and also typically improves learning outcomes (especially when the counterfactual is that children would not be going to school). However, there is more limited evidence to suggest that other categories of inputs have meaningful impacts on either time in school or test scores. It is important to caveat however that there are several categories of inputs for which we do not have much credible evidence on impacts — including teacher training and education. On a more positive note, however, there is evidence to suggest that some kinds of inputs, such as school health interventions, may be both effective as well as highly cost effective.

4.3 Pedagogy

As outlined in our theoretical framework (Section 3.1), a critical determinant of the extent to which increases in schooling inputs translate into improvements in learning outcomes is the way in which these inputs are used in practice and the way in which teaching and learning is organized. In other words, the "technology of instruction" is a key determinant of how inputs are translated into outcomes, and this section aims to review the evidence from high-quality studies in this area and to synthesize the implications of these findings for our understanding of education in developing countries.

In most developing countries, this "technology of instruction" has remained unchanged for decades, and consists of teachers using a "chalk and talk" method to teach a classroom of students. Classroom time is typically spent in a "lecturing" style starting from third grade, with students expected to follow the textbooks. This pedagogical method offers relatively little scope for differentiating instruction to account for the large heterogeneity in student ability and preparation levels within a typical classroom in a developing country. Instead, the focus is on completing the materials in the syllabus — typically specified in government-approved text books, with the goal of being able to answer questions and problems provided at the end of the chapter — which in turn will be tested in low- and high-stakes examinations. While some innovations in pedagogy are transmitted through teacher training programs, the evidence base for these innovations is typically not very strong.

An important point to note is that researchers have only recently started to use well-identified empirical strategies for identifying the impacts of pedagogical innovations on learning outcomes. Since most of the empirical innovations in improved causal identification have come from economists and statisticians, the focus of research in the economics of education has been more on questions of interest to economists, such as the impact on test scores of: programs designed to increase households' demand for education; inputs into the education production function; school resources; teacher pay levels and structure; information, choice and competition; and governance and accountability. Thus, there is a relative paucity of well-identified quantitative evidence on the critical question of how best to structure classroom instruction and pedagogy to most effectively improve learning outcomes in developing countries. Nevertheless, progress has been made on better understanding these issues in the past decade, which in turn has provided important insights into the nature of the education production function, and the binding constraints that confound attempts to improve education outcomes in developing countries. The rest of this section examines high-quality studies that have looked at the impact of pedagogy on student outcomes in developing countries; the results are summarized in Tables 8 and 9.

4.3.1 Supplemental Remedial Instruction and Teaching at the Right Level

Teaching effectively may be particularly challenging in many developing country contexts because of the higher variation, relative to developed countries, in the initial preparation of children when they enter school. This challenge has probably been magnified in the past two decades, since the focus on increasing school enrollment has brought several million first-generation learners into primary school systems in developing countries, which likely has led to a substantial increase in the entry-level variation in student abilities and preparation. Also, as Muralidharan and Zieleniak (2014) show, the variance in student learning levels within a cohort increases over time as they pass through the school system. How does a teacher effectively teach a classroom where students are so varied in their skill levels?

Remedial schooling interventions are one method that may be able to reduce the variance of achievement in the classroom and thus ensure that all students are progressing. Remedial programs offer the possibility of focusing on those students who are lagging behind and teaching at a level that is appropriate for their current level of skills. Ideally, such an intervention would increase their progress, and decrease the heterogeneity of student learning levels in a given grade. The evidence suggests that this may be the case, with several high-quality studies finding strong impacts of remedial instruction programs on learning outcomes, even when implemented by volunteers or informal teachers with little formal training and paid only a modest stipend that is several times lower than the salary of regular civil-service teachers.

The first rows of Tables 8 and 9 summarize the results from three studies that examined the impact on students' educational outcomes of interventions that focus on "teaching at the right level," which typically involves remedial/supplemental instruction and/or tutors or volunteers to provide that instruction. Only one of these studies examined the impact of teaching at the right level on students' time in school; as seen in Table 8, that study found no significant impact. All three studies considered impacts on student learning, and together they have produced six estimated impacts. As summarized in Table 9, two of these estimates are statistically insignificant while four are significantly positive. The following paragraphs discuss these studies in detail.

First, Banerjee et al. (2007) report results from an experimental evaluation of a program run by the education nonprofit organization Pratham that was specifically targeted at the lowest performing children in public schools in the Indian cities of Mumbai and Vadodara. The program provided an informal teacher hired from the community (known as a *Balsakhi* or "friend of the child") to schools, with an explicit mandate to focus on children in 3rd and 4th grade who had not achieved even basic competencies in reading and arithmetic. These children were taken out of the regular classroom for 2 hours

Table 8 Summary of impacts on time in school of pedagogy interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Teaching at the right level (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Computers, electronic games, and access to technology (high quality)	0 (0)	0 (0)	2 (1)	0 (0)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

Table 9 Summary of impacts on test scores of pedagogy interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Teaching at right level/ supplemental instruction (all RCTs)	0 (0)	1 (1)	1 (1)	4 (3)	3
Tracking/streaming (RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Computers/electronic games					
RCTs	1 (1)	0 (0)	3 (3)	10 (6)	8
Other high-quality studies	3 (1)	0 (0)	0 (0)	0 (0)	1
Reading intensive pedagogy and reading materials (RCT)	0 (0)	0 (0)	2 (1)	2 (1)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

per day, and were provided with remedial instruction targeted at their current level of learning. This is the only study that examined the impact of teaching at the right level on a time in school variable. There is only one estimate and, as seen in Table 8, the impact of this program on daily attendance was statistically insignificant. In contrast, this program improved students' test scores (average of math and English scores) by 0.14σ after 1 year of the program, and by 0.28σ after 2 years. Most of the gains were observed for students who were "pulled out" of their regular classroom (who were at the lower end of the learning distribution) and not for those who continued in the regular class (though the latter students did experience a reduction in class size for 2 hours per day). The authors therefore interpret the results as being driven by the fact that the students who were pulled out were being taught at a level corresponding to their current proficiency, as opposed to the proficiency presumed by of the textbook.

Second, Banerjee et al. (2010) report results from several interventions designed to improve community participation in education in India. Of all the interventions tried, the only one that was found to be effective at improving learning outcomes was a remedial instruction program implemented by youth volunteers hired from the village, who were provided a week of training and then conducted after school reading camps for 2 to 3 months. These increases in learning were substantial (albeit starting from a low base), even though only 13.2% of students actually attended the camps. For the average child who could not read anything at the baseline, exposure to the remedial instruction program increased the fraction who were able to read letters by 7.9 percentage points. The instrumental variable estimates suggest that for children who were not able to read, the average impact of attending a camp raised the probability of being able to read letters by 60 percentage points, which is a very large effect.

A third piece of experimental evidence is provided by Lakshminarayana et al. (2013), who study the impact of a program run by the Naandi Foundation. The program recruited community volunteers to provide remedial education to children in a randomly selected set of villages in Andhra Pradesh. After an initial outreach to households to communicate program details, the volunteers provided 2 hours of remedial instruction per day after normal school hours in the students' school. The subject matter covered in these sessions was tailored to students' class-specific needs and learning levels, and aimed to reinforce the curriculum covered in school. After 2 years of this intervention, student test scores in program villages were 0.74σ higher than those in the comparison group, suggesting a large impact of the after-school remedial instruction program. Note, however, that the large magnitudes reported in this study (it is noteworthy that the large effects reported here are "intention to treat" estimates and not scaled up "treatment on the treated" effects) also reflect high program implementation quality and monitoring over a period of 2 years.

Finally, Banerjee et al. (2015) present results from multiple RCT's conducted across several Indian states in partnership with the education nonprofit organization Pratham to evaluate the impact of different models of implementing remedial instruction in public schools in a scaled up way.[21] Overall, they find support for the hypothesis that Pratham's instructional approach, which focuses on teaching children at a level that matches their level of learning, can significantly improve learning outcomes. They find that implementing the pedagogy in dedicated learning camps that are held outside of normal school hours, where teachers use learning-appropriate remedial materials, was effective in raising test scores. However, they found no impact of other models, which attempted to incorporate this pedagogy into the regular school day. The authors interpret their findings as suggesting that the remedial pedagogy was successful, but that it was difficult to get teachers to implement new curriculums during school hours, and that successfully scaling up remedial pedagogy within an existing schooling system can be challenging because teachers are focused on completing the syllabus prescribed in the textbook.

[21] Note that this study is not included in Tables 8 and 9 because the results are still preliminary. But we include it in the discussion because the underlying RCT's have been underway for several years and we are quite familiar with the studies.

One challenge in interpreting this body of evidence is that most interventions studied include both extra instructional time as well as teaching that is targeted at the level of the student. Nevertheless, the combination of evidence on tracking (see below) and on the success of programs that provide supplemental instruction targeted at the level of the student (and the descriptive evidence on learning trajectories over time) suggests that a likely binding constraint for why substantial increases in school inputs have not translated into much improvement in learning outcomes may be ineffective pedagogy.

The explanation that is most consistent with these findings is one articulated in Chapter 4 of Banerjee and Duflo (2011) and also by Pritchett and Beatty (2012), and in the conclusion of Glewwe et al. (2009), which is that the curriculum has been designed by highly educated elites and reflects a period of time when there was no expectation of universal primary education. Indeed, as they note, the historical purpose of education systems in many developing countries may not have been to provide "human capital" to all students as much as to screen gifted students for positions of responsibility in the state and the clergy. Note that such a design of an education system may also have been efficient in times when the structure of the economy was predominantly agrarian since the returns to education also depend on occupational choice, with lower returns in occupations that mainly involve physical labor (Munshi and Rosenzweig, 2006; Pitt et al., 2014). However, the design of curricula and textbooks in developing countries have typically not been adapted to the entry into the school system of millions of first-generation learners. Since teachers continue to follow the textbook as the default mode of instruction, and define their goals in terms of completing the curriculum over the course of year, it is not surprising that they are effectively "teaching to the top" of the distribution and that a large number of children are in the class but not learning because the lessons are too advanced for them.

Thus, improvements in pedagogy that make it possible to: (a) more effectively account for the variation in initial level of student preparation; (b) break the tight link between pedagogy and the textbook; and (c) move from a "selection" paradigm to a "human capital" paradigm, are likely to lead to significant improvements in the performance of developing country education systems.

4.3.2 Tracking of Classrooms

The findings above suggest that policies that reduce variation in student learning within a classroom could improve pedagogy. Indeed, an old question in the economics of education is whether students in a given grade benefit from being "tracked" into classrooms based on their initial learning levels or ability.[22] The main argument for tracking is that the reduction in variance in the ability levels of students in the classroom may make it

[22] See Betts (2011) for an excellent theoretical discussion on the tracking debate as well as a review of the empirical evidence from developed countries.

easier for teachers to more effectively match the difficulty level of the content and material they teach to the level of their students. The main argument against tracking is the concern that students who are tracked to "lower" level classrooms may suffer further from negative peer effects and from stereotyping and loss of self-esteem, which may place them on a permanently lower trajectory of learning. A further concern is that some education systems may track students very early using data that may be noisy and not sufficiently reliable for tracking.

Unfortunately, the problem of endogenous tracking and the lack of credible identification limits the extent to which we have good causal estimates of the impact of tracking. Fortunately, however, there is one high-quality study of tracking that was conducted in the context of a developing country, that of Duflo et al. (2011). They did not examine the impact of tracking on time in school but, as seen in the second row of Table 9, they examined the impact of tracking on student learning. More specifically, they conducted an experimental evaluation of tracking in Kenya and found that tracking and streaming of pupils appears to have a positive and highly significant effect on test scores in both the short term and the long term. Students in tracking schools scored on average 0.18σ higher than students in nontracking schools, and continued to score 0.18σ higher even 1 year after the tracking program ended, suggesting longer-lasting impacts than those found in many other education interventions.

Most importantly, Duflo, Dupas, and Kremer found positive impacts for students at all quartiles of the initial test score distribution and cannot reject that students who started out above the median score gained the same as those below the median; those in higher achieving classes scored 0.19σ higher than the higher achieving control school students, while those in lower achieving classes scored 0.16σ higher than the lower achieving control school students. Additionally, lower-achieving students gained knowledge in basic skills, while higher-achieving students gained knowledge in more advanced skills, suggesting that teachers tailored their classes to the achievement level of their students. Finally, since students just below and just above the median baseline score were otherwise similar, but experienced a sharp change in the mean test score of their peers, the authors are able to use this regression discontinuity method to show that tracking did not cause adverse peer effects.

4.3.3 Technology-Enhanced Instruction

Perhaps the most promising class of interventions for improving the effectiveness of pedagogy involves the greater use of technology in the classroom. Improvements in technology have played a central role in increasing productivity in almost every sector of the economy, and it is therefore widely believed that greater use of technology in classrooms should be a promising way to rapidly improve education outcomes in developing, and developed, countries. In particular, given that the basic nature of instruction has not changed in many decades, many "technology optimists" believe that technology-enhanced instruction has the potential to create "disruptive innovation," and thereby sharply improve education outcomes.

Some of possible channels of impact include: (1) cost-effective replication and scaling up of high-quality instruction using broadcast technology (such as radio and television-based instruction); (2) overcoming limitations in teacher knowledge and training, for example to teach more advanced concepts in science and mathematics or to teach a new language (such as English) for which there is growing demand but a limited supply of teachers with the requisite competence; (3) providing supplemental instruction at home; (4) better engaging children in the learning process through the use of interactive modules (such as educational games and puzzles); (5) customizing individual student learning plans; and (6) shortening the gap between the time when students' attempt to answer questions and demonstrate understanding of concepts and the time when the teacher provides feedback to students via corrected homework or exams. These interventions also range from being quite inexpensive, such as radio-based instruction, to very expensive, such as individual laptops for students under the "One Laptop per Child" (OLPC) initiative.

While promising, the high-quality evidence to date on the impact of technology-enhanced instruction on learning outcomes in developing countries suggests a need for caution. First, there is only one study that examines the impact of computers and information technology on students' time in school. Cristia et al. (2014) used a difference-in-differences approach to examine the impact of an intervention in Peru that increased access to computers and the internet to students in public secondary schools in Peru. Their results, summarized in the second row of Table 8, show no significant effect of the program on either repetition or dropout rates.

There is much more evidence on the impact of technology-enhanced instruction on student learning; the third and fourth rows of Table 9 indicate that there are 17 estimates from nine different studies. Yet these results point to widely varying magnitudes of impact, perhaps varying more than almost any other intervention considered in this chapter, including estimates that are significantly negative and others that are significantly positive. Further caution is in order because *all* but one of the studies discussed below are based on RCTs, with high contextual internal validity of the estimated impacts but uncertain external validity. Thus, the differences in estimated impacts are quite striking and point to the importance of context, and perhaps more importantly to the importance of program design in creating effective programs of technology-aided instruction.

Studies that find significant positive impacts of computer-aided learning (CAL) include Banerjee et al. (2007) and Linden (2008), both of which examined interventions in India, and Lai et al. (2011), Yang et al. (2013), Lai et al. (2013), and Mo et al. (2014), all four of which are studies of programs in China. Banerjee et al. (2007) found that a CAL program that provided 2 hours per week of computer-based math instruction in two cities in Western India was particularly effective at improving math scores, with very large positive effects of 0.48σ at the end of 2 years of the program, but that the gains were not long-lasting (the effects fell to 0.10σ 1 year after the program). They also report that the CAL intervention was not as cost effective as a remedial tutoring program in the same setting (the latter was discussed in Section 4.3.1).

Linden (2008) found positive effects of 0.29σ from an out-of-school CAL program after 1 year of the program. Lai et al. (2011) estimated the impact of a CAL program in schools for migrant children in Beijing, China, that provided two 40-minute sessions of remedial math instruction per week to children in grade 3; after 1 year their math scores were 0.14σ higher than those of children in the control group, a statistically significant effect. Yang et al. (2013) also studied a CAL program in three different provinces in China and find modest (0.12σ) but significant positive effects on test scores. Lai et al. (2013) estimated the impact of a CAL program similar to that implemented in Beijing, but it focused on Chinese language and was implemented in a remote Western province (Qinghai); they found significant increases in both Chinese (0.20σ) and math (0.22σ) after 1 year. Mo et al. (2014) examined a similar CAL program that provided remedial math instruction to boarding school students in another area of rural China and found that it led to significant improvements in math test scores for both 3rd-grade (0.25σ) and 5th-grade (0.26σ) students.

Studies that found no impact include Barrera-Osorio and Linden (2009) in Colombia, and Beuermann et al. (2013) in Peru. Barrera-Osorio and Linden studied a school-level program that provided computers and teacher training to randomly selected schools in Colombia and found no impact of the program on test scores in either Spanish or math. The authors argue that the lack of impact was because of poor implementation, with the teachers failing to incorporate the new technology effectively in their teaching.

Even more striking are the results in Beuermann et al. (2013), who studied the impact of the OLPC program in Peru using a large-scale randomized evaluation. They found that while the program increased the ratio of computers to students in schools from 0.12 to 1.18 in treatment schools, there was no impact on either school enrollment or test scores in math and language. The results are striking both because of the intensity of the program, with each child getting an individual laptop, and because children were permitted to take the laptop home, which allowed for a much more intense immersion in technology and greater access than any other study.

Finally, two studies found a negative impact: Linden (2008) and Malamud and Pop-Eleches (2011). Linden found that, in contrast with the positive impacts found from an after-school supplemental CAL program, a CAL program that was implemented in-class and thus substituted for regular instruction had a strong negative impact (−0.55σ) on test scores. The authors interpret these results as being driven by the difficulty of effectively modifying pedagogy within the classroom to incorporate technology, which could lead to a worsening of outcomes if effective preexisting instructional patterns are disrupted.

Malamud and Pop-Eleches (2011) used an RDD to study the impact of providing vouchers for purchasing computers to the families of middle-school students in Romania. They found that students who received the voucher had significantly lower GPA's (the results are not reported in standard deviations). They also found a possible explanation for this finding: students who received vouchers to buy a computer reported spending more time playing games and less time reading or doing homework.

These cautionary results are especially relevant for education policy, where it is tempting for politicians to want to scale up interventions like "computers for all" as a potential short-cut for addressing the challenges of education quality. Our summary of the evidence, as well as the brief discussion of the theoretical mechanisms, suggest that there are many good reasons to be excited about the *potential* for technology-enabled instruction to produce substantial improvements in students' learning outcomes. However, the evidence on the impact of greater use of technology in the classroom is mixed, and program impacts seem to depend crucially on the details of both the intervention and its implementation. In particular, it appears that the key success factor is the extent to which careful thought goes into integrating effective pedagogical techniques with technology. Much more, and much more careful, research is needed (on both process and impacts) before committing resources to scaling up these programs — especially those involving expensive investments in hardware — with scarce public funds.

4.3.4 Reading Intensive Pedagogy and Reading Materials

A final pedagogy intervention for which there is a high-quality study is a reading program in the Philippines. More specifically, Abeberese et al. (2014) conducted a randomized evaluation program of a reading program for grade 4 students in the Philippines. This program provided age-appropriate reading materials and trained teachers to incorporate reading into their teaching. Four months after the start of the program the reading scores of the students in the program schools were 0.13σ higher, although this fell to 0.06σ seven months after the start of the program. Both estimates are statistically significant. In contrast, there was no impact of the program on mathematics scores.

Overall, the evidence on pedagogy suggests that the quality of classroom pedagogy may be a key determinant of the extent to which increases in school inputs are translated into improvements in learning. In particular, there may be very large returns from focusing on foundational learning for *all* children in developing country school systems with a large number of first-generation learners. It is interesting to note that countries that strongly outperform their predicted test scores conditioning on levels of income (such as Vietnam and Cuba) feature egalitarian education systems that try to ensure basic skills for all students as opposed to identifying and focusing on very high achievers. This may be an example where a macro-level "systems" approach and a micro-level "interventions" approach may be yield similar insights.

4.4 Governance

The fourth critical determinant of education outcomes — in addition to policies to stimulate household demand for education, provision of school inputs, and pedagogical practices — is the quality of governance of the education system. We use the term governance to refer to a broad set of characteristics of education systems that relate to how efficiently they are managed. These include goal setting, personnel policy (hiring, training, retention, and promotions), accountability and monitoring, and performance management. More broadly, governance also includes decentralization, the extent of choice and competition in school markets, and the regulatory structure for private schools. In this section, we first discuss a body of evidence which suggests that poor governance in developing country education systems may be a first order constraint in translating inputs into outcomes, and then discuss evidence on policy options to improve school governance.

4.4.1 Measuring Governance

Perhaps the clearest indicator of weak governance in education is outright corruption whereby fiscal appropriations for education do not reach the targeted schools and communities. A striking instance of such corruption is provided by Reinikka and Svensson (2004), who showed that 87% of central government funds allocated to a school capitation grant (for nonwage expenditures) in Uganda never reached the schools, and that the median school in their representative sample had not received *any* of the funds. They show that most of the funds were captured by local officials and politicians. In a different context, Ferraz et al. (2012) document the impact of education corruption on students' educational outcomes in Brazil. Using data from an independent federal audit of local municipal finances they show that, conditional on a wide variety of student and municipality characteristics, municipalities where these audits detected corruption in education had test scores that were 0.35σ lower and also had higher dropout rates.

Another striking measure of weakness of school and teacher governance in developing countries is the high rate of teacher absence from schools. Chaudhury et al. (2006) present results from a multicountry study where enumerators made unannounced visits to public schools to measure teacher attendance and activity, and report an average teacher absence rate of 19%, with teacher absence rates of 25% in India and 27% in Uganda. In India, Kremer et al. (2005) report that not only were 25% of teachers absent from work, but another 25% were in school but not teaching and thus only about half of the teachers were found to be actually engaged in teaching activities. Muralidharan et al. (2014) present results from a nationally representative panel survey that revisited the rural villages surveyed by Kremer et al. (2005),[23] and find only a modest reduction in teacher

[23] The panel study did not include the urban areas surveyed by Kremer et al. (2005).

absence rates, from 26.3% to 23.7%.[24] They also calculate that the fiscal cost of teacher absence in India is $1.5 billion *each year*, highlighting the large costs of poor governance in education.

In addition to these two nationally representative studies, several other studies have also noted the high rates of teacher absence in India. Duflo et al. (2012b) find teacher absence rates in excess of 40% in informal schools run by an NGO in Rajasthan. Muralidharan and Sundararaman (2011, 2013a,b) and Muralidharan (2012) regularly document teacher absence with multiple unobserved visits to a representative sample of rural government-run primary schools in Andhra Pradesh, finding teacher absence rates ranging from 24% to 28% over the 5-year period from 2005–06 to 2009–10.

A recent literature has tried to measure management quality in schools using standardized surveys to codify management practices, following the method developed by Bloom and Van Reenen (2007). Bloom et al. (2015) measure variation in management practices across schools in several countries and find that: (a) average school management quality was poorest in the developing countries in their sample; (b) the variation in country-level management practices in education was higher than in other sectors such as healthcare and manufacturing; (c) management scores are positively correlated with students' test scores (though the test scores are in levels, rather than value-added); (d) there are large disparities in management practice scores even within countries — especially between government-run public schools and those with more managerial autonomy; and (e) having better management scores is not correlated with the demographic characteristics of students, but is correlated with the quality of governance (measured as greater accountability for student performance to an outside body), and with the degree of school leadership. The descriptive evidence summarized above all points to the likely importance of improving the governance of education systems in developing countries for improving outcomes. We now turn to evidence on specific ways of improving governance.

4.4.2 Monitoring

The most basic policy tool to reduce teacher absence is to increase the extent of monitoring and oversight of schools. This can include administrative (top-down) monitoring as well as community-based (bottom-up) monitoring. While there are no experimental studies of just improving monitoring (without also linking the improved monitoring to the introduction of teacher incentives), Muralidharan et al. (2014) use a nationally representative village-level panel dataset from India (collected in 2003 and then again in 2010) on teacher absence to study the correlations between changes in various school

[24] The absence rate of 25% given above includes both the rural and the urban sample, whereas the absence rate in the rural sample in 2003 was 26.3% (for the villages in the panel data set).

and management characteristics in this period and changes in teacher absence.[25] They find that increasing the probability of a school having been inspected in the past 3 months from 0 to 1 is correlated with a 7 percentage point reduction in teacher absence (or 30% of the observed absence rates). This estimate is similar in both cross-section and panel estimates, bivariate as well as multiple regressions, and with and without state or district fixed effects. Using the most conservative of these estimates, Muralidharan et al. (2014) calculate that increasing inspections and monitoring could be over 10 times more cost effective at increasing teacher–student contact time (through reduced teacher absence) than hiring additional regular teachers.

The remainder of this section summarizes the literature on the impact of school governance policies on students' educational outcomes. Results for studies that examine the impact of such policies on students' time in school are summarized in Table 10, and Table 11 does the same for students' test scores.

Evidence on the impact of monitoring on time in school is scarce and not encouraging, as seen in the first row of Table 10. Only two studies have considered this impact, those by Duflo et al. (2012b) and Banerjee et al. (2010). Both studies found insignificant effects of monitoring on student attendance.

The evidence of the impact of monitoring on student learning is only somewhat more encouraging. The first row of Table 11 presents the results for four experimental (RCT) studies. In only one of the four studies was there a significantly positive impact of monitoring on students' test scores. This is the study of Duflo et al. (2012b), who conducted a randomized evaluation of an intervention that monitored teacher attendance in informal

Table 10 Summary of impacts on time in school of governance interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Monitoring (all RCTs)	0 (0)	1 (1)	1 (1)	0 (0)	2
School-based management					
RCTs	0 (0)	7 (3)	5 (3)	1 (1)	3
Other high-quality studies	0 (0)	2 (1)	1 (1)	0 (0)	2
Teacher performance pay (RCT)	0 (0)	1 (1)	2 (1)	0 (0)	1
Private school (vouchers)					
RCTs	0 (0)	1 (1)	1 (1)	1 (1)	2
Other high-quality studies	0 (0)	1 (1)	0 (0)	0 (0)	1
Single-sex school (high quality)	0 (0)	0 (0)	0 (0)	1 (1)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

[25] This paper is not included in the studies summarized in Tables 10 and 11 in the following paragraphs because the outcome variable is teacher absence, not students' educational outcomes.

Table 11 Summary of impacts on test scores of governance interventions

	Negative, significant	Negative, insignificant	Positive, insignificant	Positive, significant	Total studies
Monitoring (all RCTs)	0 (0)	1 (1)	4 (3)	1 (1)	4
School-based management					
RCTs	0 (0)	9 (3)	7 (3)	2 (2)	5
Other high-quality studies	0 (0)	1 (1)	1 (1)	1 (1)	2
Teacher performance pay					
RCTs	0 (0)	1 (1)	2 (1)	5 (2)	3
Other high-quality studies	0 (0)	0 (0)	1 (1)	1 (1)	1
Contract teachers (all RCTs)	0 (0)	0 (0)	0 (0)	3 (2)	2
Private school (vouchers)					
RCTs	0 (0)	0 (0)	3 (2)	2 (2)	3
Other high-quality studies	0 (0)	2 (1)	2 (1)	0 (0)	2
Diagnostic feedback to teachers (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1

Figures are number of estimates; figures in parentheses are number of papers/studies; statistical significance is at 10% level.

schools in Rajasthan (India) using cameras with time-date stamps to record teacher and student attendance. The program not only monitored teachers, but also paid teacher salaries as a function of the number of valid days of attendance. They found that this program reduced teacher absence by half, but structural estimates of a model of labor supply suggest that the mechanism for this result was not the "monitoring" per se, but rather the incentives tied to the attendance. In contrast, no significant impact was found by Muralidharan and Sundararaman (2010), who experimentally studied the impact of a program that provided schools and teachers with low-stakes monitoring and feedback; they found that this program had no impact on either teacher attendance or test scores. These results suggest that while "monitoring" is an important tool in reducing teacher absence, "low-stakes" monitoring is unlikely to be very effective, and that it is "high-stakes" monitoring, that is monitoring with positive (negative) consequences for teacher presence (absence), which is more likely to be effective.

A different way to improve monitoring of schools is to increase the amount of "bottom-up" monitoring through the community. The evidence here is less encouraging. Banerjee et al. (2010) conducted an experimental evaluation of the impact of a community mobilization program to improve school quality in rural areas of the Indian state of Uttar Pradesh; they found no impact of various programs to build community involvement in schools in that state on community participation, teacher effort, or learning outcomes. This is consistent with the study of Muralidharan et al. (2014) which, using village-level panel data, found that the correlations between "bottom-up" measures of governance and monitoring, such as the frequency of PTA meetings, and teacher absence are negative but typically not significant, and that the magnitude of the correlations is always lower than that of the correlations between "top-down" inspections and teacher absence.

There is some positive evidence on the impact of community-based information campaigns (aimed at improving bottom-up monitoring) but the interventions have typically been quite intensive. Pandey et al. (2009) conducted an experimental evaluation of an information campaign to improve parental participation in village education committees (VEC's) in three states in India and found positive impacts on both process measures as well as learning outcomes, but the estimated impacts on learning outcomes were generally statistically insignificant. Moreover, the intervention was an intensive one that involved 8–9 village-level meetings in just 2 months. Meetings were advertised in advance with audio tapes, and the meetings included high-quality videos that were used to explain the rights of VECs as well as facilitators to answer questions.[26] While it may be possible to scale up such information interventions in a less intensive and thus more cost-effective way using mass media such as television and radio, there is limited evidence on the effectiveness of such "light-touch" information campaigns on improving school governance.

4.4.3 School-Based Management

Another approach to improve monitoring and accountability of schools and teachers is to decentralize more management authority to schools and communities — an approach that is broadly referred to as "school-based management" (SBM). The theory of change associated with this approach is to empower communities to take charge of their schools and in particular to make teachers more accountable to them. Several reforms based on this approach have been attempted around the developing world, but the empirical evidence on its success is both limited and mixed. Five high-quality studies, three RCTs and two non-RCT studies, have examined the impact of SBM on children's time in school. These five studies have produced 16 estimates, of which all but one are statistically insignificant, as seen in Table 10. Turning to student learning, seven high-quality studies have estimated the impact of SBM on students' test scores, as seen Table 11, of which five are RCTs and two are non-RCT studies. Most of the 21 estimates from these seven studies are statistically insignificant, but three estimates from three different studies are significantly positive. We discuss these studies further below.

Pradhan et al. (2014) conducted an experimental evaluation of a series of interventions that aimed to enhance community participation in school management in Indonesia. The specific interventions included implementing elections for school-committee

[26] This is similar to evidence from Uganda where researchers found large positive impacts of an information-based intervention to improve the functioning of community health centers (Bjorkman and Svensson, 2009). Note, however, that this intervention included specific measures to solve the collective action problem such as a joint meeting between communities and providers to identify changes that would be made by providers and communities, and to have similar follow up meetings as well.

memberships, providing grants and training to existing school committees, and facilitating "linkage" meetings between school committees and village councils. They found no significant impact on time in school measures and test scores for most of the interventions, except for a significantly positive impact on test scores of the "linkage" intervention, which facilitated the collaboration between the school committee and the village council. The authors suggest that this may be because the "linkage" intervention incorporated stakeholders from a more powerful community institution (the village council).

Beasley and Huillery (2014) also use a randomized experiment to evaluate a parent-empowerment program in Niger. The program provided grants to school committees to encourage parent participation in school management. However, the program had no impact on either time in school measures or test scores. The authors suggest that a possible explanation for the lack of impact is that parents do not have enough knowledge and information to make effective decisions to improve educational quality.

Lassibille et al. (2010) and Glewwe and Maïga (2011) both present experimental evaluations of the AGEMAD program in Madagascar, which aimed to strengthen school management at the district, subdistrict, school and teacher levels; both studies found no impact on student test scores of these interventions. The final experimental evidence on SBM is provided by a study in Kenya by Duflo et al. (2015), who found that training school management committees to evaluate the performance of contract teachers and to have influence on the renewals of contract teachers' contracts had a significantly positive impact on the performance of contract teachers and on students' test scores.

There are also difference-in-differences studies of two SBM programs in Mexico. Gertler et al. (2012) conducted a difference-in-differences analysis of the impact of the AGE program that empowered parents to improve school quality; they found that it reduced failure and grade repetition rates, but had no impact on dropout rates (they did not have data on test scores). They also found no impact in poor communities, suggesting that the AGE program may increase inequality across schools. Santibanez et al. (2014) also used a difference-in-differences strategy to evaluate the PEC-FIDE, another SBM program in Mexico, and found that the program had no general impact on students' test scores or time in school measures. The authors suggest that schools are more likely to allocate grants based on immediate benefits rather than improving structural governance. A final study by Yamauchi (2014), which combined difference-in-differences with propensity score matching, found a significantly positive impact of an SBM program on students' test scores in the Philippines. Note, however, that this intervention included not only SBM but also additional resources to implement the program, so the significantly positive effect reflects both the switch to SBM and the additional resources to the schools.

Overall, the results on monitoring suggest that there may be significant problems of both collective action and asymmetry in power between teachers and communities, especially in historically disadvantaged areas with low levels of education, that may make community-based monitoring less effective than top-down administrative monitoring.

This is a result that is consistent with the experimental findings of Olken (2007) in the context of monitoring corruption in Indonesia, and is also seen in comparisons of top-down and bottom-up school monitoring in the same context (Muralidharan et al., 2014). To the extent that community-based monitoring programs are effective, they seem to involve either intensive interventions, or specific components to mitigate the collective action problem such as coordination meetings with village councils or authority over specific aspects of school functioning (such as renewal of contract teachers).

4.4.4 Teacher Performance Pay

A common feature of the pay structure for teachers (and other civil-service employees) is the use of fixed salary schedules with little or no component of pay that is linked to performance. Since the effort exerted by teachers is a key determinant of education quality, a natural set of policy options to enhance governance in education would be to consider linking the compensation of teachers, as well as education administrators, to measures of performance. Reasons for the status quo of little or no link to performance include difficulties in measuring productivity of individual teachers, as well as concerns that linking pay to performance on measurable attributes of a job will lead to diversion of effort away from socially valuable tasks that may not be as well measured (Baker, 1992; Holmstrom and Milgrom, 1991). Nevertheless, the demonstrated low levels of teacher effort in developing countries (manifested by high rates of absence) have led both policymakers and researchers to consider the possibility that introducing performance-linked pay for teachers may improve outcomes. Four high-quality studies have been conducted on this topic in recent years in developing countries. As seen in Table 10, one of these studies examined the impact of this type of education policy on students' time in school. All four studies examined the impact of this type of intervention on test scores (see lines 4–5 in Table 11), most of which have found significantly positive impacts. These studies are discussed in detail in the following paragraphs.

Muralidharan and Sundararaman (2011) present experimental evidence on the impact of a program in the Indian state of Andhra Pradesh that provided bonus payments to teachers based on the average improvement of their students' test scores in independently administered learning assessments (with a mean bonus of 3% of annual pay). They found that at the end of 2 years of the program, students in incentive schools performed significantly better than those in control schools by 0.27σ and 0.17σ in math and language tests, respectively. Students in incentive schools also performed better on subjects for which there were no incentives, suggesting positive spillovers between improved performance on math and language and the untested subjects (science and social studies).

This study featured two different bonus payment interventions, and teachers were randomly assigned to each; the first intervention based bonuses on the average performance of groups of teachers, while the second based bonuses on each teacher's individual performance. In the first year, the treatment schools with group and individual-level teacher bonuses did equally well (with both having a significantly positive impact), but at the end of 2 years, students in individual-incentive schools scored better than those in group-incentive schools (0.33σ vs. 0.22σ for math, and 0.24σ vs. 0.09σ for language), though these differences were not statistically significant. Finally, the performance pay programs were implemented as a part of a larger set of experimental evaluations costing the same amount, which allowed the authors to compare the relative effectiveness of input- and incentive-based approaches to improve learning outcomes in the same setting. They find that the teacher incentive schools performed significantly better than other randomly chosen schools that received additional schooling inputs of a value similar to the cost of the teacher incentive program.

Second, Muralidharan (2012) presents evidence from a long-term follow up in the same setting, where teacher performance pay program was extended for 5 years to a subsample of the original schools, and finds that students who completed all of their 5 years in primary school under the individual teacher incentive program performed significantly better than those in control schools by 0.54σ and 0.35σ on math and language tests, respectively. The group teacher incentive program also had positive (and mostly significant) effects on student test scores, but the effect sizes were always smaller than those of the individual incentive program, and were not significant at the end of primary school for the cohort exposed to the program for 5 years. The paper estimates that the individual teacher performance pay program would be around 15 to 20 times more cost effective (including administrative costs) at improving learning outcomes than the default policy of reducing pupil-teacher ratios by hiring more teachers (even assuming the most generous estimates of the impact of pupil-teacher ratio reductions on test scores obtained in the same setting).

The third and final experimental study is Glewwe et al. (2010), who conducted an experimental evaluation of a teacher incentive program in Kenya that provided school-level group incentives using prizes for high-achieving schools. The prizes were awarded, using a tournament design, to the schools that had the best average student test scores and also to those that had the highest average improvements (to reward both the best schools as well as the most-improved ones). The authors report that the incentive program led to teachers increasing efforts in test-preparation but not in activities that would increase long-term learning (such as reduced absence rates). They found that students in treatment schools performed better on high-stakes tests but not on low-stakes tests, and also that these gains dissipated after the incentive program ended. They interpret their results as suggesting that teacher incentives may not be an effective strategy to promote long-term learning. Nevertheless, there are two important caveats. The first is that we now know that all interventions appear to have significantly high rates of test-score decay (see Andrabi et al., 2011) and that there may be important long-term gains in human capital even when test score gains decay (Chetty et al., 2011). Second, the group-nature of the incentive program (across 12 teachers) may have induced free riding and weakened the incentives faced by individual teachers (as seen in Muralidharan, 2012).

Further evidence in favor of the impact of performance pay for teachers on learning outcomes is from a nonexperimental study by Contreras and Rau (2012). They used a difference-in-difference procedure to evaluate the SNED program in Chile, which provided teacher bonus payments based on students' test scores and was rolled out in a scaled up way across all public schools. Their estimates indicate that this program led to a large (0.29σ) and significant increase in students' mathematics test scores.

Two other findings are relevant to this discussion of teacher performance pay. First, incentives can be based on teacher attendance rather than students' test scores. In particular, as discussed earlier, Duflo et al. (2012b) found that paying teachers on the basis of the number of days they are present in the school (as opposed to a flat salary that does not depend on performance) led to a halving of teacher absence rates (from 42% to 21%) and significant increases in student test scores (by 0.17σ). Thus, even if performance incentives on the basis of test scores is difficult to implement, it may be possible to increase student learning by using incentives on intermediate measures of effort, such as attendance, that are easier to measure and also less subject to concerns of multi-tasking. Second, it is worth noting that positive impacts of teacher performance pay on student learning outcomes have also been found in a developed country, Israel, for both group and individual teacher incentives (Lavy, 2002, 2009).

Taken together, these results suggest that even modest changes to compensation structures to reward teachers on the basis of objective measures of performance (such as attendance or increases in student test scores) can generate substantial improvements in learning outcomes at a fraction of the cost of a "business as usual" expansion in education spending. However, not all performance pay programs are likely to be effective, so it is quite important to design the bonus formulae well and to make sure that these designs reflect insights from economic theory. For instance, in the Andhra Pradesh experiment studies by Muralidharan and Sundararaman (2011), there was no reduction in teacher absence in the incentive schools, but teachers report higher levels of teaching activity (including extra classes) in incentive schools, suggesting that the incentives did not have any impact on the extensive margin of effort (attendance) for which there was no direct reward, but that they did have an impact on the intensive margin of teacher effort conditional on attending school. In contrast, Duflo et al. (2012b) found sharp increases in teacher attendance (which was monitored and rewarded), but no increase on teaching activity among those who were present, which is consistent with teacher activity being an unrewarded activity.

4.4.5 Teacher Contractual Structure

A different way to motivate teachers to exert effort would be to make employment contracts subject to periodic renewal, with contracts not being renewed for underperforming teachers. Widely used in the private sector, the performance-contingency of the employment contract can be a powerful motivator for effort. However, the majority of teachers tend to be employed in the public sector in both developing and developed countries, and public sector employment contracts typically feature lifetime tenure after a very short probationary period (if any).

Nevertheless, in recent years many developing countries have started to employ new teachers on short-term renewable contracts. The use of such "contract teachers" has been driven by a combination of lack of enough qualified teachers to match the needs of rapidly expanding school systems, the high cost of hiring them, and the reluctance of qualified teachers to serve in rural areas where the needs of the expanding education system are the greatest. Governments in several developing countries have responded to this challenge by staffing teaching positions with locally hired teachers on fixed-term renewable contracts; these teachers are not professionally trained and are paid much lower salaries than those of regular teachers.[27] The growing use of contract teachers in public schools has been one of the most significant trends in providing primary education in developing countries in the last two decades. Contract teachers comprise a third of public-school teachers across twelve countries in Africa (Bourdon et al., 2010) and their share among all public-school teachers in India grew from 6% in 2003 to 30% in 2010 (Muralidharan et al., 2014).

The use of contract teachers has been controversial. Supporters consider the use of contract teachers to be an efficient way to expand education access and quality to a large number of first-generation learners, and argue that contract teachers face superior incentives compared to tenured civil-service teachers. Opponents argue that using underqualified and untrained teachers may staff classrooms but will not produce learning outcomes, and that the use of contract teachers deprofessionalizes teaching, reduces the prestige of the entire profession, and reduces the motivation of all teachers (Kumar et al., 2005).

While there is very little evidence on the impact of changing the employment contract structure of teachers while holding all else constant, as seen in Table 11 there is evidence from two high-quality studies on the impact of "contract teachers" on students' test scores, although the specific ways in which such teachers differ from regular civil-service teachers vary across contexts. In the first study, Duflo et al. (2015) present results from an experimental evaluation of a program in Kenya that provided a randomly selected set of schools with an extra contract teacher. The extra contract teacher was provided to the first grade, and was used to reduce class size from around 80 to around 40. Half the students were randomly assigned to the contract teacher and the other half to the

[27] Contract teacher schemes have been used in several developing countries including Benin, Burkina Faso, Cambodia, Cameroon, Chad, Congo, India, Indonesia, Kenya, Madagascar, Mali, Nicaragua, Niger, Senegal, and Togo. See Duthilleul (2005) and Bourdon et al. (2010) for reviews of contract teacher programs. Contract teachers have also been widely employed in several states of India, and are also referred to as "para-teachers" (see Govinda and Josephine, 2004).

regular civil-service teacher. This design helps the authors distinguish between the effects of a class size reduction when taught by a regular teacher (measured by comparing student performance in classes with regular teachers across treatment and control schools — since the former schools would have half the class size), and the effects of being taught by a contract teacher (measured by comparing classrooms in treatment schools that were taught by contract teachers with those that were taught by regular teachers).

As explained above (Section 4.2.3), Duflo et al. (2015) found that simply reducing class sizes had no significant impact on test scores. More relevant for teacher contract structure is that they found that students who had the reduced class sizes *and* were also taught by a contract teacher scored significantly higher (0.29σ, averaged across subjects) than those in control schools. Even more relevant is that they found that holding class size constant, students taught by contract teachers scored significantly higher than those taught by civil-service teachers even though the contract teachers are paid much lower salaries.

The second high-quality study is that of Muralidharan and Sundararaman (2013a), who present experimental evidence from a program that provided an extra contract teacher to 100 randomly chosen government-run rural primary schools in the Indian state of Andhra Pradesh. At the end of 2 years, students in schools with an extra contract teacher performed significantly better than those in comparison schools by 0.16σ and 0.15σ in math and language tests, respectively. They also found that contract teachers were significantly less likely to be absent from school than civil-service teachers (16% vs. 27%). Finally, they implemented four different nonexperimental estimation procedures (using both within- and between-school variation as well as variation over time in pupil-teacher ratios in the same school) and found that they can never reject the null hypothesis that contract teachers are at least as effective in improving student learning as regular civil-service teachers. In fact, their point estimates typically suggest that the contract teachers are more effective than regular teachers, who are more qualified, better trained, and paid salaries five times higher than those of contract teachers.

It is also relevant to this discussion to highlight that all three of the studies on teaching at the right level, which were discussed in the previous section on pedagogy and found large positive effects on student learning outcomes of that intervention, used volunteer/informal/contract teachers with minimal formal training who were paid stipends that were at most one fifth of the salary of regular teachers. These results suggest that the superior work incentives of contract teachers may more than make up for their lack of formal teacher training. They also suggest that the binding constraint in translating increased education spending into improved learning outcomes may not be teacher training and qualifications (as is commonly believed) but teacher effort, which is (relatively) weak for civil-service teachers with lifetime employment security because there is no reward under the status quo for either effort or performance (and, conversely, few consequences for poor performance). Most relevant for policy perhaps is the fact that the results suggest that similar education outcomes can be had at much lower cost (since contract teacher salaries are much lower than those of civil-service teachers), and that it may be possible to significantly expand education quality for a given level of spending by making more use of contract teachers.

Nevertheless, scaling up the insights from this body of evidence into policy is non-trivial because of political challenges. This is best highlighted by Bold et al. (2013) who conduct a similar study in Kenya to the one conducted by Duflo et al. (2015) and find that the contract teacher program that they studied had a positive impact on student learning when implemented by a nonprofit partner, but that it had no impact when implemented by the government (partly because the program itself was not implemented). In practice, the scaling up of contract teacher programs has been politically challenging because of pressure from contract-teacher unions (especially when a large number of contract teachers are hired) to get "regularized." Muralidharan (2016b) discusses these practical challenges in more detail and presents an approach to potentially address them by making contract teachers the first stage of a performance-based career ladder.

4.4.6 Private Schools, Vouchers, and Public-Private Partnerships

Another important trend in primary education in developing countries over the past two decades has been the rapid growth of private schools, with recent estimates showing that private schools now account for over 20% of total primary school enrolment in low-income countries (Baum et al., 2014). The growing market share of fee-charging private schools is especially striking as it is taking place within a context of increased spending on public education and near universal access to free public primary schools. This raises important questions regarding the effectiveness of private schools in these settings and the optimal policy response to their growth.

Opponents of the growth of private schooling argue that it has led to economic stratification of education systems and has weakened the public education system by causing the middle class to leave the public-school system. They also worry that private schools compete by cream-skimming students, and attract parents and students on the basis of superior average *levels* of test scores, but that they may not be adding more value to the marginal applicant.[28] Others contend that private schools in developing countries have grown in response to failures of the public-schooling system, that they are more accountable and responsive to parents, that they have better management practices that ensure better teacher selection and accountability, and that the revealed preference of parents suggests that they are likely to be better than public schools. Supporters of private

[28] This concern is supported by several studies across different contexts, which find that highly demanded elite schools do not seem to add more value to student learning (see Zhang (2014) in China, Lucas and Mbiti (2014) in Kenya, Cullen et al. (2006) in Chicago, and Abdulkadiroglu et al. (2014) in Boston and New York).

schools recognize that their main weakness is that they are not accessible that to the poor who cannot afford to pay fees, and argue therefore that policymakers should be more open to voucher-like models that combine public funding with private provision of education.[29]

Thus, understanding the relative impact of public and private schools on learning outcomes is a first-order question for both research and policy. There are five high-quality studies of school voucher programs in developing countries that defrayed the cost of attending private schools. As seen in Table 10, three of these studies examined the impact of vouchers on time in school. Of the four estimated impacts, one is significantly positive and the other three are statistically insignificant. All five high-quality studies examined the impact of voucher programs on student learning. The summary information in Table 11 shows that two of the three RCT studies found a significant impact, but neither of the other high-quality studies found a significant impact. The following paragraphs discuss these studies in more detail.

Angrist et al. (2002) and Angrist et al. (2006) study the short and medium-term effects the PACES program in Colombia that provided vouchers (allocated by lottery) to students to attend private secondary schools.[30] They found that voucher winners performed significantly better both 3 and 7 years after receiving the voucher. More specifically, Angrist et al. (2002) find that the voucher winners scored 0.16σ higher on math and reading tests after 3 years, and that they completed more years of school and had lower rates of grade repetition. Angrist et al. (2006) find that the voucher winners have significantly higher high-school graduation rates (5.6 percentage points higher on a base of 25–30%), and that they scored 0.2σ higher on college-entrance exams (after controlling for differential attrition in the long-term follow-up study).

However, while both of these studies suggest that the PACES voucher program was highly effective, it may not be possible to interpret these results as reflecting only the differential "productivity" of private schools because the PACES program not only allowed students to use the vouchers to attend private schools, but it also allowed vouchers to be topped up by parents (to attend a better school than they could have afforded without a voucher), and required students to maintain minimum academic standards to continue receiving the voucher. Thus while the results point to the effectiveness of the PACES program, the estimates reflect a combination of private school productivity, additional education spending, and student incentives.

Muralidharan and Sundararaman (2015) present experimental evidence on the impact of a school-choice program in the Indian state of Andhra Pradesh that featured a unique

[29] See Tooley and Dixon (2007), Muralidharan and Kremer (2008), Goyal and Pandey (2009), and Tooley (2009).

[30] In Tables 10 and 11 we classify these two studies as RCTs because they are based on an actual lottery to determine eligibility for the vouchers. One could also classify them as "natural experiments."

two-stage randomization of the offer of a voucher (across villages as well as students). The design created a set of control *villages* that allows the authors to experimentally evaluate both the individual impacts of school choice (using the student-level lottery) as well as its aggregate effects including the spillovers onto nonapplicants and students who start out in private schools (using the village-level lottery).

At the end of 2 and 4 years of the school choice program, Muralidharan and Sundararaman (2015) find no difference between the test scores of lottery winners and losers on the two main subjects of Telugu (the native language of Andhra Pradesh) and math, suggesting that the large cross-sectional test-score differences in these subjects across public and private schools (of 0.65σ) mostly reflect omitted variables. However, they also find that private schools spend significantly less instructional time on Telugu (40% less) and math (32% less) than public schools, and instead spend more time on English, and on science and social studies. They also taught a third language, Hindi, which was not taught in public primary schools. When they conduct tests in these additional subjects after 4 years of the voucher program they find small positive effects of winning the voucher on English (0.12σ; p-value = 0.098), and science and social studies (0.08σ; p-value = 0.16), and large, positive effects on Hindi (0.55σ; p-value < 0.001).

Assuming equal weights across all subjects, they find that students who won a voucher had average test scores that were 0.13σ higher, a statistically significant effect, and that the average student who used a voucher to attend a private school scored 0.26σ higher ($p < 0.01$). However, this positive impact was driven mainly by Hindi (which was taught in private schools but not in public primary schools), and they find no impact of winning a voucher on average test scores excluding Hindi. However, even without assuming equal weights across subjects, Muralidharan and Sundararaman (2015) are able to infer that private schools were more productive than public schools because they were able to achieve similar Telugu and math test scores for the lottery winners with substantially less instructional time, and use the additional time to generate large gains in Hindi test scores. Further, the annual cost per student in the public-school system is over three times the mean cost per student in the private schools in the sample. Thus, students who win a lottery to attend private schools have slightly better test scores (better on Hindi and same on other subjects) even though the private schools spend substantially lower amounts per student.

Finally, two studies of Chile that were not RCTs found insignificant effects of attending a private school on Spanish and math test scores in that country. Lara et al. (2011) used a difference-in-differences approach to estimate the impact of vouchers to attend private schools on the test scores of students in grade 10 (all of whom were in public primary schools but some of whom used vouchers to move into private secondary schools). They find a very small positive impact (0.02σ) that is not statistically significant. Hsieh and Urquiola (2006) also use a difference-in-differences approach to examine the impact of using vouchers to enroll in a private school on the test scores of students in grades

4 and 8. They find negative but statistically insignificant effects on mathematics and language (Spanish) test scores.

Overall, the results from the three RCTs suggest that private schools were more productive (by being able to deliver similar learning outcomes in math and language at lower financial and time cost), but not necessarily more effective at raising test scores. Thus, from a policy perspective an important open question is to understand how public and privately managed schools would perform in a setting where the value of the voucher was set equal to the per-student spending in public schools. The two studies from Chile, where vouchers ensured similar levels of spending at private and public schools, suggests little effect, but Chile is very different from India, so it is difficult to generalize from one of these countries to the other. Further evidence from both countries, as well as others, would allow us to measure the extent to which the apparent greater productivity of private schools in India can translate into better absolute learning outcomes, and whether this finding extends to other countries.

4.4.7 Single-Sex Schools

A final governance intervention is single-sex schools, for which there is only one study. Jackson (2012) used regression discontinuity and difference-in-differences estimates to estimate the impact of single-sex schools on enrollment (as measured by taking a national examination) of grade 10 students in Trinidad and Tobago. He found, as seen in Table 10, that students who attend such schools are 7.3 percentage points more likely to take the national exam.

Overall, the summary of the evidence reviewed suggests that developing country education systems often feature weak governance, as seen by high rates of teacher absence and low levels of effort. Changing education system governance at scale is not easy given the political constraints involved. Nevertheless, the evidence suggests that interventions that improve governance of school systems — especially with regard to teacher effort and accountability — may have a large positive impact on learning outcomes in developing countries.

5. INTERPRETING THE EVIDENCE, AND IMPLICATIONS FOR FUTURE RESEARCH

As discussed in Section 3, an important goal for the empirical research on education in developing countries has been to estimate a series of partial derivatives of an education production function. Combining estimates of marginal returns to various possible education interventions with data on marginal costs would then allow researchers and policymakers to rank various interventions by the highest marginal return on investment. This is an implicit goal for many individual studies and typically an explicit goal for papers that aim to review the evidence. Given limited budgets and several competing uses for these funds, such a ranking could be used to help policymakers prioritize across competing potential investments. Further, reallocating public education expenditure away from ineffective forms of spending to more effective ones could potentially improve education outcomes even within a given budget.

In this section, we discuss the extent to which the evidence summarized in the previous section yields clear conclusions that would allow policymakers to use this evidence to improve the effectiveness of public spending on education. In attempting to draw such broad conclusions, we highlight the many challenges in interpreting the evidence, and discuss the implications for policy as well as for future research.

5.1 Challenges in Interpreting the Evidence

5.1.1 Production Function Parameters Versus Policy Parameters

The discussion in Section 3 highlighted the value of exogenously varying $Q_{i,t}$ to estimate the causal impact of $Q_{i,t}$ on education outcomes. Note however, that even random assignment of $Q_{i,t}$ may not yield the corresponding production function parameter $\boldsymbol{\beta}$ in Eq. (5).[31] This is because the production function parameter $\boldsymbol{\beta}$ is a partial derivative $\left(\partial A_{i,t}/\partial Q_{i,t}\right)$ holding *other inputs constant*. In practice, other inputs at the school or household level may endogenously respond to exogenous changes in $Q_{i,t}$, and the estimated parameter should therefore be more accurately interpreted as a policy parameter, which is a total derivative $\left(dA_{i,t}/dQ_{i,t}\right)$ that accounts for reoptimization by agents (students, parents, teachers, and school administrators) in response to an exogenous change in $Q_{i,t}$.

The extent to which the estimate from a well-identified evaluation of a change in $Q_{i,t}$ reflects reoptimization will depend critically on the duration of the study. A clear illustration is provided by Das et al. (2013), who studied a randomly assigned school grant program in India over a 2-year period and found significant positive effects on test scores at the end of the first year, but found no effect in the second year even though the grant was provided again in the second year, and was spent on very similar items in both years (books, school supplies, and classroom learning materials). They show that the most likely explanation for this result is that household spending on books and school supplies did not change across treatment and schools in the first year (when the school grant was unanticipated), but that households in treatment schools sharply cut back their spending on these categories in the second year (when the school grant was anticipated and could be accounted for in household decision making), and that this reduction offset around 80% of the per-student value of the grant. The authors therefore argue that the "first year" effect of the program is more likely to represent the "production function" effect of providing the school grant (since other factors did not have time to adjust), whereas the "second year" effect is closer to the "policy parameter" (which reflects household reoptimization).

[31] $Q_{i,t}$ is one of the $\boldsymbol{X}_{i,t}$ variables in Eq. (5) that are school inputs.

This example highlights the importance of measuring as many intermediate inputs as possible to have a better idea about the mechanisms of program impact. However, in practice, it will be difficult to measure *all* possible intermediate inputs, and the extent to which they may have changed in response to the exogenously varied treatment. Thus, it is perhaps most accurate to interpret the "causal estimate" of $\boldsymbol{\beta}$ from experimental studies as the "policy effect" of $Q_{i,t}$ at the point when the outcomes are measured. While one could argue that experimental studies estimate production function parameters over short time periods, before agents have had an opportunity to reoptimize their behavior, it is: (a) difficult to confirm that this is true on every dimension of potential behavior modification and (b) much less likely to be true over longer time horizons.[32]

Note that this limitation is also present for nonexperimental methods, and is therefore not a criticism of experimental estimates per se. But it an important limitation to highlight because experimental estimates are often implicitly interpreted as representing the production function parameters of estimated equations such as Eq. (5).

One advantage of well-identified evaluations using large administrative data sets (based on, for example, RDDs) is that it may be possible to observe the policy effects over longer time horizons at a much lower marginal cost than in experimental studies (since the cost of conducting follow up surveys on experimental samples can be quite large, and the challenge of differential attrition grows over time). A good example of this is provided by Bharadwaj et al. (2013), who measured the impact of early childhood interventions several years later using administrative data in both Chile and Norway. Another good example is that of Pop-Eleches and Urquiola (2013), who present regression-discontinuity based estimates of the impact of going to a better school and also collect extensive data on household reoptimization in response to access to a better school. The main point, however, is that it is important to recognize that over time the estimates of program impact are more likely to represent policy effects as opposed to estimates of production function parameters.

5.1.2 External Validity

The challenges of external validity of even well-identified studies are well known (see Deaton, 2010, and the references cited therein, for further discussion). Even well-identified experimental estimates of program impacts are an estimate of the program interacted with a set of unobserved covariates — which are likely to be quite different across different contexts. These challenges are magnified in the context of writing a chapter such as this, which seeks to summarize evidence across a range of developing countries

[32] While the discussion may suggest that experimental estimates may be lower bounds of production function parameters and upper bounds of policy parameters, this need not be true if the unmeasured inputs are complements to the experimental intervention as opposed to substitutes (as was the case in Das et al., 2013).

that are located on different continents. This problem is well known among academic researchers, but can be understated in "systemic reviews" that compare interventions and estimates across contexts (see Pritchett and Sandefur, 2013, for a discussion). There is no good solution to this problem beyond conducting more studies and gathering more evidence by replicating evaluations of interventions in many settings. Unfortunately, the academic incentives to conduct such replication studies are limited.

5.1.3 Comparability of Interventions

Further, even when there are multiple studies of a similar intervention, there are almost always variations in the specific details of the intervention that make comparisons difficult. Take the case of teacher performance pay. While there are multiple high-quality studies on the subject, no two studies have the same formula for how teachers will be paid bonuses! Some of the design details that vary include individual versus group incentives, tournaments versus piece rates, linear versus nonlinear bonus formulae, formulae based on students reaching thresholds (such as the fraction who pass a test) versus those that reward improvements for all students.[33] Similarly, interventions of technology-enabled learning vary from simply providing hardware to different forms of integration of technology into pedagogy, practice, assessment, and customizing of learning pathways.

It is no surprise then that there is considerable variation in the estimated impact of even the same program in different contexts. The ranges of effect sizes from experimental evaluations of the same intervention documented by McEwan (2015) show that the variations within intervention type are often larger than those between intervention types. How then should we make policy-relevant inferences from the body of evidence summarized in Section 4 and in the other review papers written in the past 3 years? The following section considers this question.

5.2 Mapping the Evidence to a Deeper Understanding of Developing Country Education

We suggest here that one way of proceeding from well-identified individual studies to a deeper understanding of how best to improve education in developing countries is to think about whether the interventions or programs studied were alleviating a binding constraint to improving education quality.[34] In particular, we may learn as much from understanding why well-intentioned interventions do not work as expected as we may from interventions that are "successful."

[33] Indeed the question of optimal design of a teacher bonus system is by itself an important theoretical question that is not easy to answer (see Barlevy and Neal, 2012, for a discussion).

[34] The approach here is similar to that proposed by Dani Rodrik in the context of country-level growth diagnostics. See Rodrik (2008) for a discussion.

5.2.1 Understanding Nonimpact

In theory (as laid out in Section 3), finding that an intervention had a zero effect (or did not have a significant impact) should simply mean that the estimate of the associated $\boldsymbol{\beta}$ in Eq. (5), and thus the marginal impact of increasing $Q_{i,t}$, is insignificantly different from zero. In practice, however, it is important to distinguish between four different interpretations of a zero result.[35] These include: (a) poor implementation of the intervention, including corruption or administrative failures, which are often the binding constraint in many developing countries; (b) substitution away from other inputs provided by agents (including governments, schools, teachers, and households) in response to the treatment; (c) the intervention may help some students, but may not alleviate a binding constraint to education outcomes in the context being studied; and (d) absence of complementary inputs/reforms that may be needed for the intervention to be effective. Note that reasons (c) and (d) are consistent with the interpretation that the marginal impact of increasing $Q_{i,t}$ on outcomes is zero in a production function sense, but reasons (a) and (b) are not. Further, the distinction between (c) and (d) also matters for policy because the policy implication of (c) may be not to prioritize increasing $Q_{i,t}$ until the binding constraint is reduced or removed, whereas the policy implication of (d) would be to increase $Q_{i,t}$ as long as the complementary input is also increased.

These possibilities are illustrated across four different randomized evaluations of the impact of providing books and materials to students. (In some cases, these were directly provided as books, while in other cases the input was a grant to schools that was spent primarily on books and related materials.) Each of the four studies finds zero average impacts of providing books and materials, but each points to different possible reasons for the zero effects. Sabarwal et al. (2014) find no impact on test scores from the provision of textbooks to schools in Sierra Leone and attribute this to the fact that schools stored the textbooks instead of distributing them to students (which is a form of poor implementation). Das et al. (2013), described above, also find no net impact on test scores from the provision of a school grant (that was spent mostly on books and related materials) in India, but attribute it to households offsetting the intervention by reducing their own spending on these inputs.

Glewwe et al. (2009) also find no impact on test scores from providing textbooks to students in Kenya. But they do find positive impacts on students with the highest baseline test scores, and they suggest that their results are consistent with the fact that the majority of children could not read the English language textbooks to begin with, and thus could not benefit from the textbooks (whereas those who could read *did* benefit). Thus, in this case, the nonimpact is interpreted as suggesting that textbooks did not alleviate the binding constraint to learning in this context (which was the lack of sufficient reading skills).

[35] A fifth possibility, that the impact has no effect under almost any circumstances, is discussed below.

Finally, in ongoing work by one of us (Muralidharan), we again also find no impact on test scores from the provision of a large capitation grant to schools in Tanzania (the largest item that the grant was spent on was textbooks). However, this study was explicitly designed to test for complementarities with teacher effort (which was boosted by a separate intervention that paid teachers bonuses based on student performance) using a cross-cutting design with a sample size large enough to test for complementarities, and we find that the interaction effect of the school grant and teacher performance pay was significantly positive. In other words, the school grant on its own had no impact, but had a significant impact when provided in conjunction with a teacher performance pay intervention. Thus, it is likely that the performance pay treatment contributed to teachers making more effective use of the additional materials, but it is also true that having the materials allowed teachers to significantly improve student outcomes relative to teachers who increased effort due to performance-linked pay but had few or no textbooks to work with.

The broader point here is that these experiments, all of which found zero results, provide useful results in and of themselves, and yield an important policy conclusion that the marginal impact of providing only books and learning materials to students may be very low. On the other hand, the fact that four papers with the same result point to four different reasons for this nonimpact suggest that a "black box" experiment on its own may yield limited insights into the nature of the education production function and the true binding constraints to learning. Many more insights can be gained by collecting data on intermediate processes and inputs to better understand the factors that explain the observed "reduced form" zero estimates of program impact.

Of course, it is also possible that the true impact of a particular input is zero, with none of the above qualifications being applicable. A good example is provided by de Ree et al. (2015), who conducted an experimental evaluation of the impact of an Indonesian policy that unconditionally doubled teacher pay, and find that there was zero intensive margin impact of the salary doubling on student test scores. (They do not study the potential extensive margin benefits of improved teacher quality over time.) In this case, the program was well implemented, there was no measured crowding out of any other inputs, and there was no heterogeneity in the impacts across either teacher or student characteristics. While it is conceivable that some complementary intervention may have caused the salary doubling to have had a positive impact, the policy was implemented without any such intervention, and represents what was actually done. Thus, finding that the impact of a program that had a very large fiscal cost was zero is of value in itself, and is directly policy relevant given the number of settings where such policies are contemplated.

5.2.2 Drawing the Right Lessons From "Effective" Interventions

As discussed in Section 5.1.3, even when there are multiple high-quality studies of the same intervention, the details of the interventions often differ in important ways. Thus, even in cases where there is evidence to suggest that an intervention is "effective," we argue that it is more useful to focus on the principle illustrated by the successful intervention rather than a more mechanical conception of identifying interventions that "work." Overall, our assessment of the evidence to date is that interventions that expand schooling inputs may be less effective at improving learning outcomes than those that improve the effectiveness of pedagogy within classrooms and/or improve governance and the accountability of schools and teachers.

For instance, the fact that multiple studies find positive impacts of supplemental instruction targeted at the current learning level of the student (as opposed to the level of the textbook for the grade that the child is in) combined with evidence on flat learning trajectories from longitudinal data on learning outcomes (Muralidharan and Zieleniak, 2014) suggest that a key binding constraint to improving education quality in low-income countries is that students who are not functionally literate or numerate at the end of the second year of primary school may be learning very little in school in the subsequent years even if they stay enrolled. This may be an important reason for why expensive expansions in inputs and resources may be having little impact on learning outcomes, while inexpensive supplemental instruction programs are so effective at doing so (Banerjee and Duflo, 2011; Pritchett and Beatty, 2012).

Thus, the evidence to date suggests that the lack of foundational learning may be a critical binding constraint for education systems in low-income countries, and that policymakers would do well to prioritize giving attention to this constraint. The next step should be to identify specific models by which such foundational learning may be most effectively provided, and to iteratively pilot and evaluate such models to inform decisions on scaling up. The series of studies conducted on evaluating different models of "Teaching at the Right Level," which are summarized in Banerjee et al. (2015), is a good illustration of such an effort.

Similarly, the evidence strongly suggests that poor governance and accountability in education systems in low-income countries may be a critical constraint to improving learning outcomes. In particular, there is compelling evidence to suggest that pivoting public education spending from a business as usual input-augmentation strategy to an "inefficiency reduction" strategy may substantially improve education outcomes without increasing total spending. Two stark comparisons are provided by considering alternative ways of increasing teacher compensation, and of reducing class size.

De Ree et al. (2015) show that an unconditional doubling of teacher salaries in Indonesia had no impact on student learning outcomes, whereas Muralidharan and Sundararaman (2011) and Muralidharan (2012) show that a teacher performance pay program was highly effective, even though it spent only 3% of annual salary on the bonuses.[36] Similarly, Duflo et al. (2012b) show how changing the *structure* of teacher

[36] They show that the administrative cost of implementing the program was equal to at most another 3% of pay.

pay (from a flat salary to one that depends on the number of days attended) can substantially improve teacher attendance, even while not increasing the *level* of pay very much. These results provide a striking example of how pivoting public expenditure on teacher salaries (the largest component of education spending) from unconditional raises to performance-linked payments could sharply increase the efficiency of education spending.

Another example is provided by Muralidharan et al. (2014) who show that reducing teacher absence by investing in better school monitoring could be over ten times more cost effective at reducing the effective student-teacher ratio in schools (net of teacher absence) than the default policy of hiring more teachers. Thus, reallocating existing education expenditure away from less effective to more effective policies may substantially increase the total factor productivity (TFP) of publicly provided education in low-income countries.

Again, the specific details of how best to improve governance will depend on contextual factors, and the evidence presented in Section 4 should not be interpreted as calling for a universal scale up of any specific policy for improving governance. Rather, as the discussion above suggests, the evidence should best be interpreted as pointing to the likely large returns to investing in improving governance relative to a business as usual input-augmentation strategy.

5.3 Cost-Effectiveness and Making Individual Studies More Comparable

As discussed earlier, a key motivation for the empirical research on education in developing countries in the past decade has been to identify the relative cost-effectiveness of alternative ways to spend scarce public resources on education. Indeed, our discussion above has been motivated by cost-effectiveness and the possibility of improving education outcomes in a fiscally neutral way by reallocating spending from less cost-effective policies to more cost-effective ones. Nevertheless, analyzing the cost-effectiveness of every study reviewed in this chapter requires both the estimation of marginal benefits of different programs (the focus in the discussion so far) *and* data on the costs of these programs. Unfortunately, individual studies do not typically report cost data in a comparable way, which makes it challenging for us (and others) to present the evidence in this format.

Organizations such as the Abdul Latif Jameel Poverty Action Lab (J-PAL) are working to conduct such analyses, at least for experimental studies, but this is still challenging because studies do not typically report cost data in a systematic way, and often use customized measures of learning outcomes that are difficult to compare.[37] While the current default approach is to report per student cost per standard deviation of test scores improved of an intervention, this approach still has several shortcomings. First, the

[37] The authors of this chapter have served as former and current education program co-chairs of J-PAL, and the discussion reflects our experiences from attempting this exercise.

standard deviation of a test score can be sensitive to the range of question difficulty on the test, and the representativeness of the student sample in which the test was carried out. Second, the framework is not easy to apply to report on interventions that do not improve test scores, but reduce costs (technically, these would be infinitely cost-effective interventions on the margin). Third, the studies are typically carried out across several different contexts, which makes such cost-effectiveness comparisons difficult to interpret. Fourth, program costs are typically higher in smaller scale experiments than in scaled up settings where the fixed cost of program design can be amortized over a larger number of schools. The following paragraphs provide some results on cost-effectiveness, but the wide variation in results and the problems just mentioned imply that they should be regarded as exploratory rather than conclusive.

The three studies that found significant impacts of merit-based scholarships also report the amounts of the scholarships, although they do not report administrative and other costs of those programs. The Kenya girls' scholarship program studied by both Kremer et al. (2009) and Friedman et al. (2011) paid a total of $38 per scholarship recipient ($19 per year, including payments to parents, for 2 years) which implies a cost of about $6 per student since only about 15% of students received scholarships. The former study found impacts of 0.27σ at the end of the 2 year program, which implies a cost of $2 per 0.1σ increase, and the latter study found an impact of 0.20σ four to five years after the program, which implies a cost of $3 per 0.1σ increase. The more recent study by Blimpo found impacts of about 0.26σ after 1 year for all three variants of the incentive program, but the (average) payments to students varied by the intervention: the individual incentive cost about $9 per student, which implies a cost of $3 per 0.1σ increase; the team incentive cost $3 per student, which implies a cost of $1 per 0.1σ increase; and the team tournament incentive cost about $6 per year, which implies a cost of $2 per 0.1σ increase.

Three studies of CCT programs found significantly positive impacts on test scores: the studies by Baird et al. (2011) and Baird et al. (2013) of the same program in Malawi, and the study of Barham et al. (2013) on a CCT program in Nicaragua. The lowest cost intervention in the Baird et al. (2011) study provided payments of $5 per month for 10 months per year, and thus cost $50 per year. The payments were provided over a period of 2 years so the total cost was $100 per student. The impacts of the program on English and math scores, respectively, were 0.14σ and 0.12σ. Averaging over these impacts implies a cost of $77 per 0.1σ increase in student test scores. This is for girls who were already in school when the program began; Baird et al. (2013) examine the impact of the program on girls who were not in school when the program began. The impacts for these girls were slightly higher, 0.13σ and 0.16σ for English and math, respectively, but that paper did not investigate the size of the impact by the size of the grant, and so the estimates are for an average grant of $10 per month; averaging over these impacts implies a cost of $138 per 0.1σ increase in student test scores. Finally, Barham et al. (2013) evaluated a CCT program in Nicaragua that made educated-related payments of $133 per household per year. A typical household had one or two children in the program, so the cost per child could be \$133 or \$67 per year. The impact measured is of 3 years in the program, so total costs are \$399 (one eligible child in household) or \$200 (two eligible children in household) for the duration of the program. The estimated impact, averaged over language and math, was 0.20σ. Even assuming the lower cost of \$200 this implies a cost of \$100 per 0.1σ increase in student test scores.

Two studies of the introduction of computers found significantly positive effects and also had information about the cost of the intervention. Banerjee et al. (2007) found that providing four computers per school for children to share in primary schools in India increased math test scores by 0.48σ after 2 years, but 1 year after the program ended the impact was only 0.09σ. The program cost \$15 per student per year, so \$30 for 2 years. Using the estimated effect of 0.48σ the cost was \$6 per 0.1σ increase in student test scores, but using the "long-run" estimate of 0.09σ the cost increases to \$33 per 0.1σ increase in student test scores. Linden (2008) also provides cost estimates for the two computer interventions that he considered. The one that was successful, which took place out of school and so did not reduce instructional time during the school day, increased students' test scores by 0.29σ after 1 year and cost only \$5 per student per year. This implies that the cost of increasing test scores was only about \$2 per 0.1σ increase in student test scores. Overall, the cost-effectiveness of interventions that provide computers to students varies widely, depending on the specifics of the intervention (recall that many had no effect, and some even had negative effects) and, for at least one study, on how long the effects persists after the child is no longer in the program.

A final study to consider is the teacher incentive program evaluated by Muralidharan and Sundararaman (2011), which after 2 years increased students' test scores by 0.28σ when the teacher incentives were at the individual teacher level and by 0.15σ when the incentives were at the group (of teachers) level. The cost of the former was \$222 per school, or about \$3 per student, while the cost of the latter was \$133 per school, or about \$2 per student (the average school had about 80 students). These costs imply that both types of teacher incentive programs cost only about \$1 per 0.1σ increase in student test scores.

These cost figures must be interpreted with caution, because most of them do not include administrative costs. Moreover, for the CCT interventions and the merit-based scholarship interventions much of the cost was not a cost to society as a whole but rather a redistribution of funds from taxpayers to program beneficiaries, and this redistribution may well have had poverty reduction benefits that should be included as benefits when making comparisons. Thus much more work needs to be done to compare the cost-effectiveness of these interventions. Yet it is clear that even among successful programs there is very wide variation in cost-effectiveness, which implies that more evidence is needed in order to improve students' educational outcomes in a way that maximizes those improvements given limited resources for achieving that goal. Indeed, having more comparable measures of costs and benefits would be very useful for education research and policy, and it would also be useful for funders of education research to create standardized templates for reporting costs and benefits that authors of individual studies should be encouraged to fill out to enable such comparisons (even though they should be interpreted with caution).

6. CONCLUSION

Improving education in developing countries is likely to promote both aggregate economic growth and the "inclusiveness" of this growth. As a result, national governments as well as international development organizations have prioritized and invested considerable resources in improving education attainment and outcomes in developing countries, and they continue to do so each year. This focus has resulted in sharp increases in school enrollment and completion rates in developing countries, and it is a reflection of this priority that these rates are substantially higher than what today's high-income countries had achieved at a comparable level of income per capita. At the same time, there is growing evidence that the extent to which extra years in school translates into additional human capital is much less than policymakers had hoped.

The past decade has seen a sharp increase in the number of high-quality empirical studies on the impacts of various education policies on improving enrollment, attendance, and learning outcomes. This chapter has summarized this evidence, and it has also attempted to interpret the body of evidence to date to guide both policymakers as well as researchers and funders of future research.

For policymakers, the main message from the summary of the evidence is that a "business as usual" approach of expanding spending on education (most of which is spent on increasing fairly standard inputs into education) is unlikely to have much impact on improving learning outcomes beyond that predicted by the cross-country relationship between per capita income and learning outcomes. On the other hand, interventions that focus on improved and more effective pedagogy (especially providing foundational literacy and numeracy skills to the millions of first-generation learners who are falling behind the pace of curriculum), and on improved governance of the education system (especially teacher performance and accountability) are likely to yield considerably greater returns on increased spending.

For researchers, we hope that this chapter provides a useful framework for organizing future research efforts, and helps direct attention to the areas where the knowledge gaps are particularly large. In particular, one point worth highlighting is that much of the high-quality empirical research in education in developing countries has been done by economists, and as a result the topics on which there is more evidence tend to be topics of interest to economists (such as household demand, information, inputs, and incentives). However, some of the most promising avenues for improving education in low-income settings may involve improving the design and delivery of classroom instruction. Thus, we believe that a fruitful area of future research is for researchers trained in rigorous empirical methods (especially in running field experiments) to collaborate with experts in curriculum and pedagogy to improve the empirical evidence base on effective pedagogy in low-income settings. Much more work is needed on the more "standard" topics as well, but it is likely to be more fruitful to focus attention on the optimal design of interventions (including many incentive-based ones) and testing alternative designs rather than simply testing whether some interventions work. In addition, we hope to have highlighted the importance of collecting intermediate data on mechanisms and context in order to shed more light not just on *whether* an intervention works, but also on *why* it may or may not have worked in the setting where it was studied.

Finally, for funders and international organizations, we urge that more attention be paid to the creating of "research public goods" including IRT-linked question item banks (that allow the results of different interventions in many parts of the world to be compared on a common scale), the creating of standardized templates for reporting cost information, and to encourage recipients of research funding to use these common platforms to enable greater comparability and improve the ability of individual studies to contribute to enhanced understanding of education in developing countries.

ACKNOWLEDGMENTS

We thank Amy Damon, Suzanne Wisniewski, and Bixuan Sun for collaboration in putting together the studies reviewed in this chapter. We also thank James Sayres and Eric Stephens for research assistance. Finally, we thank Eric Hanushek, Lant Pritchett, and the UK Department for International Development (DFID) for comments on preliminary versions of this chapter.

REFERENCES

Abdulkadiroglu, A., Angrist, J., Pathak, P.A., 2014. The elite illusion: achievement effects at Boston and New York exam schools. Econometrica 82, 137–196.

Abeberese, A.B., Kumler, T.J., Linden, L.L., 2014. Improving reading skills by encouraging children to read in school: a randomized evaluation of the Sa Aklat Sisikat reading program in the Philippines. J. Hum. Resour. 49 (3), 611–633.

Adrogue, C., Orlicki, M.E., 2013. Do in-school feeding programs have an impact on academic performance? The case of public schools in Argentina. Educ. Policy Anal. Arch. 21 (50), 1–23.

Alderman, H., Kim, J., Orazem, P.F., 2003. Design, evaluation, and sustainability of private schools for the poor: the Pakistan urban and rural fellowship school experiment. Econ. Educ. Rev. 22, 265–274.

Alderman, H., Gilligan, D.O., Lehrer, K., 2012. The impact of food for education programs on school participation in northern Uganda. Econ. Dev. Cult. Chang. 61 (1), 187–218.

Ambler, K., Aycinena, D., Yang, D., 2014. Channeling Remittances to Education: A Field Experiment Among Migrants from El Salvador. NBER Working Paper # 20262 [published as Ambler, Aycinena and Yang (2015)].

Ambler, K., Aycinena, D., Yang, D., 2015. Channeling remittances to education: a field experiment among migrants from El Salvador. Am. Econ. J. Appl. Econ. 7 (2), 1–27.

Andrabi, T., Das, J., Khwaja, A., Zajonc, T., 2011. Do value-added estimates add value? Accounting for learning dynamics. Am. Econ. J. Appl. Econ. 3 (3), 29–54.
Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in Colombia: evidence from a randomized natural experiment. Am. Econ. Rev. 92 (5), 1535–1558.
Angrist, J., Bettinger, E., Kremer, M., 2006. Long-term educational consequences of secondary school vouchers: evidence from administrative records in Colombia. Am. Econ. Rev. 96 (3), 847–862.
ASER, 2014. Tenth Annual Status of Education Report. ASER Centre, New Delhi.
Attanasio, O., Fitzsimons, E., Gomez, A., Gutierrez, M.I., Meghir, C., Mesnard, A., 2010. Children's schooling and work in the presence of a conditional cash transfer program in rural Colombia. Econ. Dev. Cult. Chang. 58 (2), 181–210.
Attanasio, O.P., Meghir, C., Santiago, A., 2012. Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. Rev. Econ. Stud. 79, 37–66.
Baez, J.E., Camacho, A., 2011. Assessing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia. World Bank Policy Research Paper No. 5681, World Bank, Washington, DC.
Baird, S., McIntosh, C., Ozler, B., 2011. Cash or condition? Evidence from a cash transfer experiment. Q. J. Econ. 26 (4), 1709–1753.
Baird, S.J., Chirwa, E., de Hoop, J., Özler, B., 2013. Girl Power: Cash Transfers and Adolescent Welfare. Evidence from a Cluster-Randomized Experiment in Malawi. NBER Working Papers 19479, National Bureau of Economic Research.
Baird, S., Hicks, J.H., Kremer, M., Miguel, E., 2015. Worms at Work: Long-Run Impacts of a Child Health Investment. UC Berkeley.
Baker, G., 1992. Incentive contracts and performance measurement. J. Polit. Econ. 100 (3), 598–614.
Banerjee, A., Duflo, E., 2011. Poor Economics. MIT Press, Cambridge, MA.
Banerjee, A., Cole, S., Duflo, E., Linden, L., 2007. Remedying education: evidence from two randomized experiments in India. Q. J. Econ. 122 (3), 1235–1264.
Banerjee, A., Banerji, R., Duflo, E., Gelnnerster, R., Khemani, S., 2010. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. Am. Econ. J. Econ. Policy 2 (1), 1–30.
Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., et al., 2015. Teaching at the Right Level: Evidence from Randomized Evaluations in India. MIT, Cambridge, MA.
Banerji, R., Berry, J., Shotland, M., 2013. The Impact of Mother Literacy and Participation Programs on Child Learning: Evidence from a Randomized Evaluation in India. Abdul Latif Jameel Poverty Action Lab (J-PAL), Cambridge, MA.
Barham, Tania, Macours, Karen, Maluccio, John A., 2013. More Schooling and More Learning?: Effects of a Three-Year Conditional Cash Transfer Program in Nicaragua After 10 Years. Inter-American Development Bank.
Barlevy, G., Neal, D., 2012. Pay for percentile. Am. Econ. Rev. 102 (5), 1805–1831.
Barrera-Osorio, F., Filmer, D., 2013. Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches. World Bank Policy Research Working Paper 6541.
Barrera-Osorio, F., Linden, L., 2009. The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. World Bank Policy Research Working Paper #4836, Impact Evaluation Series No. 29.
Barrera-Osorio, M.B., Linden, L., Perez-Calle, F., 2011. Improving the design of conditional transfer programs: evidence from a randomized education experiment in Colombia. Am. Econ. J. Appl. Econ. 3, 167–195.
Baum, D., Lewis, L., Lusk-Stover, O., Patrinos, H., 2014. What Matters Most for Engaging the Private Sector in Education: A Framework Paper. The World Bank, Washington, DC.
Beasley, E., Huillery, E., 2014. Willing But Unable: Short-Term Experimental Evidence on Parent Empowerment and School Quality. J-PAL Working Paper.
Becker, G.S., 1962. Investment in human capital: a theoretical analysis. J. Polit. Econ. 70, 9–49.
Behrman, J.R., Parker, S.W., Todd, P.E., 2009. Schooling impacts of conditional cash transfers on young children: evidence from Mexico. Econ. Dev. Cult. Chang. 57 (3), 439–477.
Behrman, J.R., Parker, S.W., Todd, P.E., 2011. Do conditional cash transfers for schooling generate lasting benefits? A five-year followup of PROGRESA/Oportunidades. J. Hum. Resour. 46 (1), 93–122.

Behrman, J.R., Gallardo-Garcia, J., Parker, S.W., Todd, P.E., Velez-Grajales, V., 2012. Are conditional cash transfers effective in urban areas? Evidence from Mexico. Educ. Econ. 20 (3), 233–259.

Bellei, C., 2009. Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. Econ. Educ. Rev. 28, 629–640.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., Pouliquen, V., 2015. Turning a shove into a nudge? A "labeled cash transfer" for education. Am. Econ. J. Econ. Policy 7 (3), 86–125.

Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. J. Polit. Econ. 75, 352–365.

Betts, J., 2011. The economics of tracking in education. In: Hanushek, E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, vol. 3. North Holland, Amsterdam, pp. 341–381.

Beuermann, D., Cristia, J., Cruz-Aguayo, Y., Cueto, S., Malamud, O., 2013. Home Computers and Child Outcomes: Short-Term Impacts from a Randomized Experiment in Peru. NBER Working Paper Series #18818.

Bharadwaj, P., Loken, K., Nielson, C., 2013. Early life health interventions and academic achievement. Am. Econ. Rev. 103 (5), 1862–1891.

Bjorkman, M., Svensson, J., 2009. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. Q. J. Econ. 124 (2), 735–769.

Blimpo, M.P., 2014. Team incentives for education in developing countries: a randomized field experiment in Benin. Am. Econ. J. Appl. Econ. 6 (4), 90–109.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. Q. J. Econ. 122 (4), 1351–1408.

Bloom, N., Lemos, R., Sadun, R., Van Reenen, J., 2015. Does management matter in schools. Econ. J. 125 (May), 647–674.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., Sandefur, J., 2013. Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. Working Paper 321Center for Global Development, Washington, DC.

Borkum, E., 2012. Can eliminating school fees in poor districts boost enrollment? Evidence from South Africa. Econ. Dev. Cult. Chang. 60 (2), 359–398.

Borkum, E., He, F., Linden, L., 2012. School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program. Working Paper, Abdul Latif Jameel Poverty Action Lab, MIT.

Bourdon, J., Frölich, M., Michaelowa, K., 2010. Teacher shortages, teacher contracts and their impact on education in Africa. J. R. Stat. Soc. Ser. A 173, 93–116.

Burde, D., Linden, L.L., 2013. Bringing education to Afghan girls: a randomized controlled trial of village-based schools. Am. Econ. J. Appl. Econ. 5 (3), 27–40.

Carnoy, M., Gove, A., Marshall, J., 2007. Cuba's Academic Advantage: Why Students in Cuba Do Better in School. Stanford University Press, Palo Alto, CA.

Chaudhury, N., Hammer, J., Kremer, M., Muraldhiran, K., Rogers, F.H., 2006. Missing in action: teacher and health worker absence in developing countries. J. Econ. Perspect. 20 (1), 91–116.

Chaudhury, N., Parajuli, D., 2010. Conditional cash transfers and female schooling: the impact of the female school stipend programme on public school enrolments in Punjab, Pakistan. Appl. Econ. 42 (28), 3565–3583.

Chay, K., McEwan, P., Urquiola, M., 2005. The central role of noise in evaluating interventions that use test scores to rank schools. Am. Econ. Rev. 95 (4), 1237–1258.

Chen, Y., Jin, G.Z., 2012. Does health insurance coverage lead to better health and educational outcomes? Evidence from rural China. J. Health Econ. 31 (1), 1–14.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How does your kindergarten classroom affect your earnings: evidence from Project Star. Q. J. Econ. 126, 1593–1660.

Chin, A., 2005. Can redistributing teachers across schools raise educational attainment: evidence from operation blackboard in India. J. Dev. Econ. 78 (2), 384–405.

Conn, K., 2014. Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Rigorous Impact Evaluations. Columbia University, New York.

Contreras, D., Rau, T., 2012. Tournament incentives for teachers: evidence from a scaled-up intervention in Chile. Econ. Dev. Cult. Chang. 61 (1), 219–246.

Cristia, J., Czerwonko, A., Garofalo, P., 2014. Does technology in schools affect repetition, dropout and enrollment? Evidence from Peru. J. Appl. Econ. 17, 89–112.
Cullen, J.B., Jacob, B.A., Levitt, S., 2006. The effect of school choice on participants: evidence from randomized lotteries. Econometrica 74, 1191–1230.
Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., Sundararaman, V., 2013. School inputs, household substitution, and test scores. Am. Econ. J. Appl. Econ. 5 (2), 29–57.
de Janvry, A., Finan, F., Sadoulet, E., 2012. Local electoral incentives and decentralized program performance. Rev. Econ. Stat. 94 (3), 672–685.
Deaton, A., 2010. Instruments, randomization, and learning about development. J. Econ. Lit. 48 (2), 424–455.
Deaton, A., 2013. The Great Escape: Health, Wealth, and the Origins of Inequality. Princeton University Press, Princeton, NJ.
Duflo, E., 2001. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. Am. Econ. Rev. 91 (4), 795–813.
Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. Am. Econ. Rev. 101, 1739–1774.
Duflo, E., Dupas, P., Kremer, M., 2012a. School governance, teacher incentives, and pupil-teacher ratios: experimental evidence from Kenyan primary schools. No. w17939, National Bureau of Economic Research. (published as Duflo, Dupas and Kremer (2015)).
Duflo, E., Hanna, R., Ryan, S.P., 2012b. Incentives work: getting teachers to come to school. Am. Econ. Rev. 102 (4), 1241–1278.
Duflo, E., Dupas, P., Kremer, M., 2015. School governance, teacher incentives, and pupil-teacher ratios: experimental evidence from Kenyan primary schools. J. Public Econ. 123, 92–110.
Duthilleul, Y., 2005. Lessons Learnt in the Use of 'Contract' Teachers. International Institute for Educational Planning, UNESCO, Paris.
Evans, D., Popova, A., 2015. What Really Works to Improve Learning in Developing Countries: An Analysis of Divergent Findings in Systematic Reviews. Policy Research Working Paper 7203, The World Bank, Washington, DC.
Ferraz, C., Finan, F., Moreira, D.B., 2012. Corrupting learning: evidence from missing federal education funds in Brazil. J. Public Econ. 96 (9-10), 712–726.
Filmer, D., Schady, N., 2014. The medium-term effects of scholarships in a low-income country. J. Hum. Resour. 49 (3), 663–694.
Fiszbein, A., Schady, N., et al., 2009. Conditional Cash Transfers: Reducing Present and Future Poverty. The World Bank, Washington, DC.
Friedman, W., Kremer, M., Miguel, E., Thornton, R., 2011. Education as Liberation?. No. w16939, National Bureau of Economic Research.
Galiani, S., McEwan, P.J., 2013. The heterogeneous impact of conditional cash transfers. J. Public Econ. 103, 85–96.
Gertler, P.J., Patrinos, H.A., Rubio-Codina, M., 2012. Empowering parents to improve education: evidence from rural Mexico. J. Dev. Econ. 99, 68–79.
Gitter, S., Barham, B., 2008. Women and targeted cash transfers in Nicaragua. World Bank Econ. Rev. 22 (2), 271–290.
Glewwe, P., Kassouf, A.L., 2012. The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. J. Dev. Econ. 97 (2), 505–517.
Glewwe, P., Kremer, M., 2006. Schools, teachers, and education outcomes in developing countries. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education: Volume 1. North Holland, Amsterdam.
Glewwe, P., Maïga, E., 2011. The impacts of school management reforms in Madagascar: do the impacts vary by teacher type? J. Dev. Effect. 3 (4), 435–469.
Glewwe, P., Miguel, E., 2008. The impact of child health and nutrition on education in less developed countries. In: Schultz, T.P., Strauss, J. (Eds.), Handbook of Development Economics, Volume 4. Elsevier, Amsterdam.
Glewwe, P., Kremer, M., Moulin, S., Zitzewitz, E., 2004. Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. J. Dev. Econ. 74 (1), 251–268.

Glewwe, P., Kremer, M., Moulin, S., 2009. Many children left behind? Textbooks and test scores in Kenya. Am. Econ. J. Appl. Econ. 1 (1), 112–135.
Glewwe, P., Ilias, N., Kremer, M., 2010. Teacher incentives. Am. Econ. J. Appl. Econ. 2 (3), 205–227.
Glewwe, P., Hanushek, E., Humpage, S., Ravina, R., 2013. School resources and educational outcomes in developing countries. In: Glewwe, P. (Ed.), Education Policy in Developing Countries. University of Chicago Press, Chicago.
Glewwe, P., Park, A., Zhao, M., 2014. A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China. Working Paper, Abdul Lateef Jamil Poverty Action Lab, MIT, Cambridge, MA. Forthcoming, Journal of Development Economics.
Govinda, R., Josephine, Y., 2004. Para-Teachers in India: A Review. National Institute of Educational Planning and Administration, New Delhi.
Goyal, S., Pandey, P., 2009. How Do Government and Private Schools Differ? Findings from Two Large Indian States. World Bank, Washington, D.C.
Handa, S., 2002. Raising primary school enrolment in developing countries: the relative importance of supply and demand. J. Dev. Econ. 69, 103–128.
Hanushek, E.A., Woessmann, L., 2015. The Knowledge Capital of Nations: Education and the Economics of Growth. MIT Press, Cambridge, MA, p. 262.
Hidalgo, D., Onofa, M., Oosterbeek, H., Ponce, J., 2013. Can provision of free school uniforms harm attendance? Evidence from Ecuador. J. Dev. Econ. 103 (2013), 43–51.
Holmstrom, B., Milgrom, P., 1991. Multi-task principal-agent analysis: incentive contracts, asset ownership, and job design. J. Law Econ. Org. 7, 24–52.
Hsieh, C.-T., Urquiola, M., 2006. The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. J. Public Econ. 90 (8), 1477–1503.
Huan, W., Chu, J., Loyalka, P., Tao, X., Shi, Y., Qu, Q., Yang, C., Rozelle, S., 2014. Can School Counseling Reduce School Dropout in Developing Countries?. REAP Working Paper #275.
Jackson, C.K., 2012. Single-sex schools, student achievement, and course selection: evidence from rule-based student assignments in Trinidad and Tobago. J. Public Econ. 96 (1–2), 173–187.
Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. Q. J. Econ. 25 (2), 515–548.
Kazianga, H., de Walque, D., Alderman, H., 2012. Educational and child labour impacts of two food-for-education schemes: evidence from a randomized trial in rural Burkina Faso. J. Afr. Econ. 21 (5), 723–760.
Kazianga, H., Levy, D., Linden, L.L., Sloan, M., 2013. The effects of "girl-friendly" schools: evidence from the BRIGHT school construction program in Burkina Faso. Am. Econ. J. Appl. Econ. 5 (3), 41–62.
Kremer, M., 1993. The O-ring theory of economic development. Q. J. Econ. 108, 551–575.
Kremer, M., Muralidharan, K., Chaudhury, N., Rogers, F.H., Hammer, J., 2005. Teacher absence in India: A snapshot. J. Eur. Econ. Assoc. 3, 658–667.
Kremer, M., Miguel, E., Thornton, R., 2009. Incentives to learn. Rev. Econ. Stat. 91 (3), 437–456.
Kremer, M., Brannen, C., Glennerster, R., 2013. The challenge of education and learning in the developing world. Science 340, 297–300.
Krishnaratne, S., White, H., Carpenter, E., 2013. Quality Education for All Children? What Works in Education in Developing Countries. International Initiative for Impact Evaluation (3ie) Working Paper 20, New Delhi.
Kumar, K., Priyam, M., Saxena, S., 2005. The trouble with para-teachers. Frontline 18.
Lai, F., Luo, R., Zhang, L., Huang, X., Rozelle, S., 2011. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing. Rural Education Action Program Working Paper 228.
Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, X., Boswell, M., Rozelle, S., 2013. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China. Rural Education Action Program Working Paper.
Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D., Mann, V., 2013. The support to rural India's public education system (STRIPES) trial: a cluster randomised controlled trial of supplementary teaching, learning material and material support. PLoS One. 8 (7), e65775.

LaLonde, R.J., 1986. Evaluating the econometric evaluations of training programs with experimental data. Am. Econ. Rev. 76, 604–620.

Lara, B., Mizala, A., Repetto, A., 2011. The effectiveness of private voucher education: evidence from structural school switches. Educ. Eval. Policy Anal. 33 (2), 119–137.

Lassibille, G., Tan, J.-P., Jesse, C., Nguyen, T., 2010. Managing for results in primary education in Madagascar: evaluating the impact of selected workflow interventions. World Bank Econ. Rev. 24 (2), 303–329.

Lavy, V., 2002. Evaluating the effect of teachers' performance incentives on pupils' achievements. J. Polit. Econ. 110 (6), 1286–1318.

Lavy, V., 2009. Performance pay and teachers' effort, productivity and grading ethics. Am. Econ. Rev., 99 (5), 1979–2011.

Li, T., Han, L., Zhang, L., Rozelle, S., 2014. Encouraging classroom peer interactions: evidence from Chinese migrant schools. J. Public Econ. 111, 29–45.

Linden, L., 2008. Complement or Substitute? The Effect of Technology on Student Achievement in India. J-PAL Working Paper.

Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., Zhang, L., Shi, Y., Chu, J., Rozelle, S., 2013. Can information and counseling help students from poor rural areas go to high school? Evidence from China. J. Comp. Econ. 41, 1012–1025.

Lucas, A.M., Mbiti, I., 2014. Effects of school quality on student achievement: discontinuity evidence from Kenya. Am. Econ. J. Appl. Econ. 6, 234–263.

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., Yue, A., Zhao, Q., Martorell, R., 2012. Nutrition and educational performance in rural China's elementary schools: results of a randomized control trial in Shaanxi Province. Econ. Dev. Cult. Chang. 60 (4), 735–772.

Malamud, O., Pop-Eleches, C., 2011. Home computer use and the development of human capital. Q. J. Econ. 126, 987–1027.

McEwan, P., 2013. The impact of Chile's school feeding program on education outcomes. Econ. Educ. Rev. 32, 122–139.

McEwan, P., 2015. Improving learning in primary schools in developing countries: a meta-analysis of randomized experiments. Rev. Educ. Res. 85 (2), 353–394.

Miguel, E., Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica 72 (1), 159–217.

Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S., Brinton, C., 2013. School dropouts and conditional cash transfers: evidence from a randomised controlled trial in rural China's junior high schools. J. Dev. Stud. 49 (2), 190–207.

Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M., Rozelle, S., 2014. The Persistence of Gains in Learning from Computer Assisted Learning (CAL): Evidence from a Randomized Experiment in Rural Schools in Shaanxi Province in China. REAP Working Paper 268.

Munshi, K., Rosenzweig, M., 2006. Traditional institutions meet the modern world: caste, gender and schooling choice in a globalizing economy. Am. Econ. Rev. 96 (4), 1225–1252.

Muralidharan, K., 2012. Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India. JPAL Working Paper.

Muralidharan, K., 2013. Priorities for primary education policy in India's 12th five-year plan. India Policy Forum 9, 1–46.

Muralidharan, K., 2016a. Field Experiments in Education in Developing Countries. UC San Diego.

Muralidharan, K., 2016b. A New Approach to Public Sector Hiring in India for Improved Service Delivery. UC San Diego.

Muralidharan, K., Kremer, M., 2008. "Public-Private Schools in Rural India". In: Peterson, P., Chakrabarti, R. (Eds.), School Choice International. MIT Press, Cambridge, MA.

Muralidharan, K., Prakash, N., 2013. Cycling to School: Increasing Secondary School Enrollment for Girls in India. NBER Working Paper #19305.

Muralidharan, K., Sundararaman, V., 2010. The impact of diagnostic feedback to teachers on student learning: experimental evidence from India. Econ. J. 120 (546), F187–F203.

Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: experimental evidence from India. J. Polit. Econ. 119 (1), 39–77.

Muralidharan, K., Sundararaman, V., 2013a. Contract Teachers: Experimental Evidence from India. JPAL Working Paper.
Muralidharan, K., Sundararaman, V., 2013b. The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. National Bureau of Economic Research Working Paper No. 19441. (published as Muralidharan and Sundararaman (2015)).
Muralidharan, K., Sundararaman, V., 2015. The aggregate effect of school choice: evidence from a two-stage experiment in India. Q. J. Econ. 130 (3), 1011–1066.
Muralidharan, K., Zieleniak, Y., 2014. Chasing the Syllabus: Measuring Learning Trajectories in Developing Countries with Longitudinal Data and Item Response Theory. UC San Diego.
Muralidharan, K., Das, J., Holla, A., Mohpal, A., 2014. The Fiscal Costs of Weak Governance: Evidence from Teacher Absence in India. NBER Working Paper 20299.
Murnane, R., Ganimian, A., 2014. Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations. NBER Working Paper 20284.
Newman, J., Pradhan, M., Rawlings, L.B., Ridder, G., Coa, R., Evia, J.L., 2002. An impact evaluation of education, health and water supply investments by the Bolivian social investment fund. World Bank Econ. Rev. 16 (2), 241–274.
Olken, B., 2007. Monitoring corruption: evidence from a field experiment in Indonesia. J. Polit. Econ. 115 (2), 200–249.
Olken, B.A., Onishi, J., Wong, S., 2014. Should aid reward performance? Evidence from a field experiment on health and education in Indonesia. Am. Econ. J. Appl. Econ. 6 (4), 1–34.
Orkin, K., 2013. The Effect of Lengthening the School Day on Children's Achievement in Ethiopia. YL-WP119.
Oster, E., Thornton, R., 2011. Menstruation, sanitary products, and school attendance: evidence from a randomized evaluation. Am. Econ. J. Appl. Econ. 3 (1), 91–100.
Ozier, O., 2014. Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming. World Bank Policy Research Working Paper 7052, World Bank.
Pandey, P., Goyal, S., Sundararaman, V., 2009. Community participation in public schools: impact of information campaigns in three Indian states. Educ. Econ. 17 (3), 355–375.
Pitt, M., Rosenzweig, M., Hassan, M., 2014. Human capital investment and the gender division of labor in a brawn-based economy. Am. Econ. Rev. 102 (7), 3531–3560.
Pop-Eleches, C., Urquiola, M., 2013. Going to a better school: effects of behavioral responses. Am. Econ. Rev. 103 (4), 1289–1324.
Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., Artha, R.P., 2014. Improving educational quality through enhancing community participation: results from a randomized field experiment in Indonesia. Am. Econ. J. Appl. Econ. 6 (2), 105–126.
Pridmore, P., Jere, C., 2011. Disrupting patterns of educational inequality and disadvantage in Malawi. Compare: J. Comp. Intern. Educ. 41 (4), 513–531.
Pritchett, L., Beatty, A., 2012. The Negative Consequences of Over-Ambitious Curricula in Developing Countries. Harvard Kennedy School.
Pritchett, L., Sandefur, J., 2013. Context matters for size: why external validity claims and development practice do not mix. J. Global. Develop. 4 (2), 161–197.
de Ree, J., Muralidharan, K., Pradhan, M., Rogers, H.F., 2015. Double for Nothing: Experimental Evidence on the Impact of an Unconditional Teacher Salary Increase on Student Performance in Indonesia. National Bureau of Economic Research Working Paper 21806.
Reinikka, R., Svensson, J., 2004. Local capture: evidence from a central government transfer program in Uganda. Q. J. Econ. 119 (2), 679–705.
Rodrik, D., 2008. The New Development Economics: We Shall Experiment, But How Shall We Learn?. Working Paper, Kennedy School of Government, Harvard University, Cambridge, MA.
Sabarwal, S., Evans, D.K., Marshak, A., 2014. The Permanent Input Hypothesis: The Case of Textbooks and (No) Student Learning in Sierra Leone. Policy Research Working Paper, No. WPS 7021, World Bank Group, Washington, DC.
Santibanez, L., Abreu-Lastra, R., O'Donoghue, J.L., 2014. School based management effects: resources or governance change? Evidence from Mexico. Econ. Educ. Rev. 39, 97–109.

Schultz, P.T., 2004. School subsidies for the poor: evaluating the Mexican Progresa poverty program. J. Dev. Econ. 74, 199–250.
Sylvia, S., Luo, R., Zhang, L., Shi, Y., Medina, A., Rozelle, S., 2013. Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China. Econ. Educ. Rev. 37, 1–12.
Tan, J.-P., Lane, J., Lassibille, G., 1999. Student outcomes in Philippine elementary schools: an evaluation of four experiments. World Bank Econ. Rev. 13 (3), 493–508.
Todd, P., Wolpin, K., 2003. On the specification and estimation of the production function for cognitive achievement. Econ. J. 113 (485), F3–F33.
Todd, P.E., Wolpin, K.I., 2006. Assessing the impact of a school subsidy program in mexico: using a social experiment to validate a dynamic behavioral model of child schooling and fertility. Am. Econ. Rev. 96, 1384–1417.
Tooley, J., 2009. The Beautiful Tree: A Personal Journey into How the World's Poorest People Are Educating Themselves. Penguin, New Delhi.
Tooley, J., Dixon, P., 2007. Private education for low-income families: results from a global reseach project. In: Srivastava, P., Walford, G. (Eds.), Private Schooling in Less Economically Developed Countries: Asian and African Perspectives. Oxford Studies in Comparative Education. Symposium Books. Oxford, UK.
Urquiola, M., 2006. Identifying class size effects in developing countries: evidence from rural Bolivia. Rev. Econ. Stat. 88 (1), 171–177.
Urquiola, M., Verhoogen, E., 2009. Class-size caps, sorting, and the regression discontinuity design. Am. Econ. Rev. 99 (1), 179–215.
Yamauchi, F., 2014. An alternative estimate of school-based management impacts on students' achievements: evidence from the Philippines. Policy Research Working Paper No. WPS 6747; Impact Evaluation Series No. IE 113, World Bank Group, Washington, DC.
Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F., Rozelle, S., 2013. Computers and the academic performance of elementary school-aged girls in China's poor communities. Comput. Educ. 60 (2013), 335–346.
Yi, H., Song, Y., Liu, C., Huang, X., Zhang, L., Bai, Y., Ren, B., Shi, Y., Loyalka, P., Chu, J., Rozelle, S., 2014. Giving kids a head start: the impact and mechanisms of early commitment of financial aid on poor students in rural China. J. Dev. Econ. 113, 1–15.
Zhang, H., 2014. The Mirage of Elite Schools: Evidence from Lottery-Based School Admissions in China. Chinese University of Hong Kong, Hong Kong.

This page intentionally left blank