

# Improving Education Outcomes in South Asia

## Findings from a Decade of Impact Evaluations

*Salman Asim*

*Robert S. Chase*

*Amit Dar*

*Achim Schmitten*



**WORLD BANK GROUP**

Social Protection and Labor Global Practice Group

July 2015

## Abstract

There have been many initiatives to improve education outcomes in South Asia. Still, outcomes remain stubbornly resistant to improvements, at least when considered across the region. To collect and synthesize the insights about what actually works to improve learning and other education outcomes, this paper conducts a systematic review and meta-analysis of 29 education-focused impact evaluations from South Asia, establishing a standard that includes randomized control trials and quasi-experimental designs. It finds that while there are impacts from interventions that seek

to increase the demand for education in households and communities, those targeting teachers or schools and thus the supply-side of the education sector are generally much more adept at improving learning outcomes. In addition, interventions that provide different actors with resources and those that incentivize behavioral changes show moderate but statistically significant impacts on student learning. A mix of input- and incentive-oriented interventions tailored to the specific conditions on the ground appears most promising for fostering education outcomes in South Asia.

---

This paper is a product of the Social Protection and Labor Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The corresponding author may be contacted at [rchase@worldbank.org](mailto:rchase@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

**Improving Education Outcomes in South Asia:  
Findings from a Decade of Impact Evaluations**

Salman Asim<sup>a</sup>, Robert S. Chase<sup>a,\*</sup>, Amit Dar<sup>a</sup> and Achim Schmillen<sup>a</sup>

JEL-Classification I25; I21; O15

Keywords: systematic review; meta-analysis; impact evaluations; education outcomes; South Asia

<sup>a</sup> The World Bank. \* Corresponding author; address: Robert S. Chase; The World Bank Group; 1818 H St. NW, Washington DC; email: [rchase@worldbank.org](mailto:rchase@worldbank.org).

Acknowledgements: The authors thank Martin Rama and Jesko Hentschel for guidance and Mohan Prasad Aryal, Harsha Aturupane, Tara Beteille, Shantayanan Devarajan, Sangeeta Goyal, Elisabeth King, Tobias Linden, Matthew Morton, Shinsaku Nomura, Florencia Pinto, Lant Pritchett, Dhushyanth Raju, Venkatesh Sundararaman, Huma Ali Waheed and other World Bank colleagues for helpful comments and suggestions.

## **1. Introduction and Motivation**

Promoting learning for all and improving the quality of education and ultimately learning outcomes are vital objectives both for individuals and for countries as a whole. As people and governments across South Asia long have recognized, education is essential to living a fuller and more productive life. For people regardless of their poverty status, education is an end in itself: having the capability to read, to compute, to understand the world around us opens opportunities and life chances that otherwise would be closed, helping people to live fuller lives. Quality education can empower people to imagine and achieve what they thought out of reach, contributing to their own welfare and that of society in ways that they previously had not imagined. Further, from the perspective of individual productivity, education generates economic returns, increasing the capacity of individuals to promote improved livelihoods, manage shocks and take advantage of and generate new economic opportunities. As such, quality education is particularly valuable for those nearly 507 million people living in extreme poverty, surviving on less than 1.25 US dollars per day and who have few other assets beyond their human potential. Given these individual benefits and aggregating them, providing learning for a country's populace can have positive impacts for a country as whole. For all countries but particularly for countries with large populations, where economic opportunities are likely focused on labor intensive output, economic growth depends on the how productive is the population, productivity that depends on quality education and learning for all.

Quality education and learning are particularly important at this point in South Asia's development trajectory. South Asian countries not only have enormous populations living in poverty, where effectively their human potential is their only asset, but also they face huge demographic challenges that mean that achieving quality education will be particularly valuable in the coming decades. With a very young population, more than one million young people every month are now entering the labor market and looking for jobs. If wisely harnessed, this tide of young workers can produce a demographic dividend that will promote economic opportunities. However, if these large numbers of young labor market entrants do not have the human capital and employment opportunities to find productive and fulfilling livelihoods, those demographic changes will generate political and social pressures as they become inadequately occupied. To lift South Asia's large numbers of poor people out of poverty, increase the productivity of the population as a whole, and promote opportunities for young labor market entrants to find jobs, South Asia faces a serious imperative to provide quality education to its population.

Recognizing education as a crucial contributor to development, South Asian countries have made important progress to increase educational attainment. Considering the second Millennium Development Goal of providing universal primary education for all, countries in South Asia have made genuine, impressive progress to improve education access for their people. South Asia's net enrollment rate was 75 percent in 2000 and rose to 89 percent by 2010. The number of children between the ages of eight and 14 years that are out of school fell from 35 million to 13 million between 1999 and 2010. Sri Lanka and the Maldives have consistently enrolled almost all their children in primary schools. Bhutan and India have recently made significant progress by increasing enrollment

rates steadily to about 90 percent of children aged six to 14 years. In Pakistan the primary net enrollment rate jumped from 58 percent to 74 percent between 2000 and 2011, though that is still lower than the regional average. Moreover, according to data from the UNESCO Institute for Statistics, between 2000 and 2010 South Asia's lower secondary enrollment rate increased from about 44 percent to 58 percent.

In addition, there has been significant movement towards gender parity as well as some success in drawing the more marginalized into school. Bangladesh and Sri Lanka both now have more girls than boys in grades six to twelve; in India the percentage of girls in secondary school went up from 60 percent in 1990 to 74 percent in 2010.

While there has been progress to improve the access to education in the region, access for all remains elusive, particularly in getting the disadvantaged and marginalized into school, and in particular into post-basic education. Still, the key challenging policy agenda now is on enhancing the quality of education and making progress towards improving learning outcomes – the ultimate goal of any educational system and (as e.g. shown by Hanushek and Woessmann, 2012, and Hanushek, 2013) a stronger driver of economic growth than years of schooling. A recent report on challenges, opportunities, and policy priorities in school education in South Asia, Dundar et al. (2014), established that mean student achievements in mathematics, reading, and language are very low throughout the region, except maybe for Sri Lanka. Mean student achievement in arithmetic tends to be particularly low. For example, India's National Council of Educational Research and Training found that only a third of grade five students could compute the difference between two decimal numbers. Similarly, in rural Pakistan, only 37 percent of grade five students can divide three-digit numbers by a single-digit number, while in rural Bangladesh more than a third of fifth graders do not even have grade three competencies (World Bank, 2012). Mean student achievement in reading and language is low in most South Asian countries, too. The Annual Status of Education Report found, for instance, that less than half of fifth graders in rural India were able to read a grade two text in their native language. This meant they were already three years behind in grade-appropriate competency. In rural Pakistan, the situation seems not to be much better. According to the South Asian Forum for Education Development, only 41 percent of grade three students are able to read a sentence in Urdu or Sindhi. A recent national assessment of both rural and urban learning competencies in Bangladesh showed that only 25 percent of grade five pupils had attained the reading achievement expected of their grade (World Bank, 2012). Moreover, within countries mean levels tend to be low but variances high. Thus, a small proportion of students can meet international benchmarks while the rest perform very poorly (Dundar et al., 2014).

While the progress to date on access to education is notable, there are several reasons why the next set of challenges of improving quality and reaching the most disadvantaged children will be even more difficult. First, the two challenges are jointly determined, for if primary education is not viewed as providing quality learning that leads to improved life choices, then families will not view it as important to make other sacrifices, like paying for uniforms and text books or foregoing other earnings behaviors, to send children to school. Further, while it is relatively straightforward to measure whether

or not children have access to education, it is much more subjective and difficult to measure whether they are learning in school. Further, while it is essential to have school inputs available to students, i.e., classrooms that they can learn in, teachers that they can learn from, and materials that they can interact with, there is no clear recipe for how these inputs should be combined to inspire learning in children. And even if there were clear accepted measures of learning and consensus on the appropriate pedagogy to combine education inputs for quality education, there are obviously differences between children in which learning strategies are most appropriate for which students. As South Asia education policy shifts focus from access to quality and reaching the most disadvantaged potential learners, a large number of difficult challenges arise.

Aware of these open questions and challenges, this paper seeks to contribute to this crucial education and development issue of what approaches are most effective to improve learning in schools in South Asia and to continue expanding school enrollment so that all children will be able to complete a full course of primary education.

Across the region, there is growing recognition of the need to improve the quality of education so that people can enjoy broader life opportunities, become more productive and lift themselves out of poverty or out of risk of falling into poverty. Further, there is a continued imperative to get the most disadvantaged and hard to reach young people into school. However, while awareness of these challenges is a vital precondition to address them, a large and unresolved policy question is how to improve education quality and get all children into school in South Asia.

Dundar et al. (2014) point to the need for a multi-pronged strategy for improving education quality that includes initiatives within and outside the education sector. The first priority is to focus explicitly on measuring student achievement and progress, significantly strengthening student learning assessments. This needs to happen regularly, consistently, and rigorously, and in benchmarking national learning outcomes against international standards. The second priority is to ensure that young children get enough nutrition. Evidence worldwide, from both developing and developed countries shows that investing in early-life nutrition, with appropriate coverage and age targeting, is critical to offset life-long learning disadvantages, and can be a highly cost-effective investment in the quality and efficiency of education. Third, raising teacher quality is essential to improving education quality. Higher and clear standards must be enforced, along with providing support to teachers to improve their quality. Merit-based recruitment policies need to be enforced along with measures to enhance teacher accountability and address the problem of teacher absenteeism. Fourth, governments need to use financial incentives to boost education quality – potentially linking them to need and student performance. Finally, governments cannot possibly afford to improve educational quality by themselves. They need to partner with the private sector, including non-governmental organizations, in this effort. Governments should encourage greater private-sector participation by easing entry barriers and encouraging well-designed public-private partnerships in education.

While primary education enrollments in South Asia and across the world have been increasing over the past two decades, another important trend has been affecting our capacity to address development challenges: development researchers have increasingly supported and analyzed innovations that seek

to improve development outcomes using impact evaluations. Impact evaluations assess changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, program or policy (Baker, 2000). In practice, it is often extremely challenging to reliably estimate the outcomes attributable to an intervention because this necessitates an answer to the question of what would have happened to those receiving the intervention if they had not in fact received the program. Over the last decade, innovative analytic techniques that fall under the broad category of experimental and quasi-experimental designs have made it possible to make a lot of progress on this challenge. The resulting impact evaluations have generated highly rigorous evidence of whether innovations really make the differences they intend to make.

Randomized control trial (RCT) studies represent the highest standard of impact evaluation evidence in development. This economics research methodology derives from techniques used in the physical sciences. Overall, the approach is to isolate a specific innovation, working to ensure that the innovation is the only systematic difference between the general population and the “treatment” group that participates in the experiment. The central design feature of the RCT approach is that the treatment group is randomly selected, so that, for the purposes of statistical analysis, treatment and control groups are indistinguishable, save for the fact that the treatment group is subjected to the innovation whose impact the study is evaluating. By comparing development outcomes in the treatment group versus the control group, research that uses experimental methods is able to rigorously isolate whether the tested innovations make an attributable difference.

Quasi-experimental designs are likewise highly rigorous approaches to gathering evidence that isolates the impact of development innovations. While these methods do not randomly assign participation in treatment and control groups, they use econometric techniques to exploit features of the process of how and where innovations were implemented that introduced random variation. Taking advantage of this variation, researchers using quasi-experimental methods are able to rigorously isolate whether the tested innovations make an attributable difference. Quasi-experiment designs are sometimes seen as slightly less rigorous than randomized control designs. At the same time, they offer a number of advantages. In particular, quasi-experiment designs are often cheaper to implement than RCTs and offer the possibility to evaluate programs after their introduction using existing data. Besides, with quasi-experiments, there are sometimes fewer concerns about the external validity of results as this type of design is frequently used to analyze interventions introduced on either the national level or at least in a large geographical area.

Experimental and quasi-experimental evidence has been applied with increasing prevalence in countries around the world, considering innovations in different sectors. Within that growing set of experimental and quasi-experimental design impact evaluations, there are a relatively large number of experimental and quasi-experimental impact evaluations that have tested and gathered high quality evidence on education in South Asia. As presented below in Section 3, there are 29 distinct studies that document the results of impact evaluations of education interventions in South Asia that reached our standard for rigor. This provides a rich body of high-quality analytic literature which we seek to bring together in this paper.

While South Asia has been at the forefront of the movement to rigorously evaluate education-related interventions, a substantial number of such impact evaluations have also been conducted for other regions. Recently, this body of literature has been reviewed by several literature reviews with different regional or even global perspectives. Together with a group of more verbal or narrative reviews (for instance by Glewwe et al., 2011; Kremer, Brannen and Glennerster, 2013; and Murnane and Ganimian, 2014), four works stand out because – similar to this paper – they combine a systematic literature review with a rigorous meta-analysis of the available evidence to investigate what kind of interventions are most effective in improving education outcomes. One of these four reviews (by Petrosino et al., 2012) centers primarily on school enrollment and attendance. The other three (by Krishnaratne, White and Carpenter, 2013; Conn, 2014; and McEwan, 2014) are mainly concerned with students’ learning outcomes. Another differentiating factor among the four reviews is that Petrosino et al. (2012), Krishnaratne, White and Carpenter, (2013) and McEwan (2014) consider evidence from all over the developing world while Conn (2014) concentrates on Sub-Saharan Africa, a region that shares many similarities with South Asia but also exhibits a number of important differences. Reassuringly, the results of our systematic review and meta-analysis are generally very consistent with the findings from the four other methodologically comparable works. Nevertheless, since our paper and the four other relevant works differ from each other with regard to their exact methodology, coverage and categorization of interventions, in Section 3 we will not only document the findings of our own systematic literature review and rigorous meta-analysis but subsequently also discuss each of the four other works in turn. This discussion will highlight similarities between their main findings and ours as well as those instances where education-related interventions have apparently had different impacts in South Asia than in other parts of the developing world.

The rest of this paper is structured as follows. In section 2, it outlines a conceptual framework structured according to actors that operate to provide learning in schools. It then summarizes in Section 3 the evidence that is available, applying a clear standard of evidence. From this systematic review, the paper then conducts a meta-analysis that allows for a careful, rigorous summary of the findings from existing impact evaluation studies, allowing the reader to understand whether the findings of individual studies jointly lead to significant implications. From this meta-analysis, we draw conclusions in section 4 for what insights emerge from this literature about the most promising approaches to address the crucial challenges which South Asia faces to strengthen its education system for all.

## **2. Conceptual Framework**

In this section we develop and present a basic framework that helps to understand available evidence for how to improve education outcomes in South Asia. With this goal in mind, we seek to frame and organize the factors that influence children’s learning and other education-related outcome variables.

Overall, we can consider a results chain that involves multiple inputs to education, several categories of outputs, and ultimately, different measures of learning outcomes. Inputs can come in several



categories, ranging, for instance, from the physical building and furnishing of classrooms, hiring teachers to educate children, providing students and teachers with learning materials and curricular innovations, enabling households to create the preconditions that ready children to attend and learn in school (such as through improved nutrition or early childhood development), and building household and community awareness of the importance and relevance for educating children. Outputs can involve enrollment of girls and boys in different grades, class sizes and proportions of teachers per student, and increasing the proportion of teachers present in school. Education outcomes can similarly be measured in multiple ways, including learning as measured by test scores on different subjects or progress to next education levels. Besides, they might also include improvements in earnings or other measures of welfare. Further, for many of the linkages between inputs and outputs and outcomes, there are relationships and accountability systems that create incentives for the actors involved. For example, while teachers are central actors to generate improved learning in the classroom, the contracts under which teachers are hired have crucial implications for how teachers teach. When considering the evidence available to address efforts to improve education outcomes in South Asia, there are a range of different inputs, outputs, and outcomes, each of which is mediated through different principal agent relationships. Given these multiple inputs, outputs, outcomes and accountability relationships, there are no simple, linear results chains that adequately capture the extensive evidence we seek to summarize, though it is useful to keep in mind how each of the inputs affects outputs and outcomes and to measure those impacts.

Moreover, it is possible to systematically organize available education innovations, all of which seek to change the inputs to education. Those innovations can be distinguished along three dimensions, represented schematically in **Figure 1**.

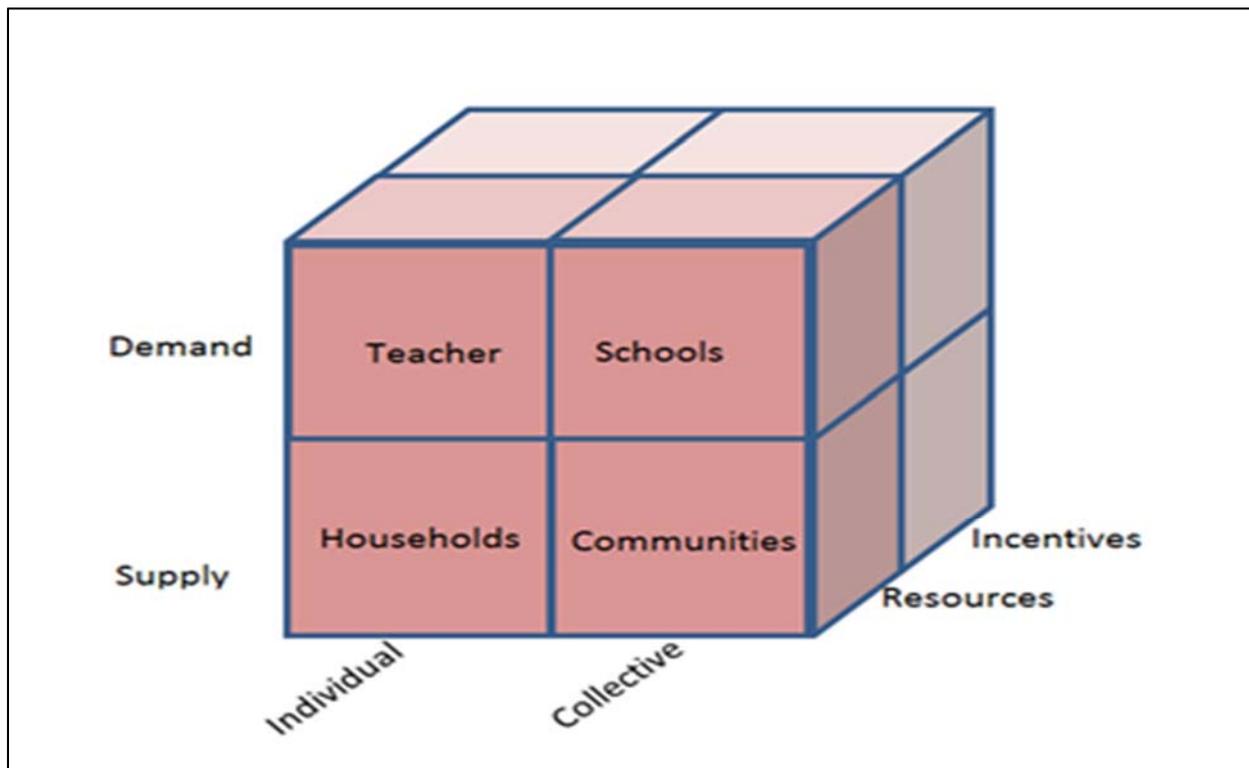
The first dimension is supply versus demand for education. Education outcomes depend on having educational services supplied through provision of physical facilities, teachers that have incentives to educate children, curricula and materials for children to interact with, and all of these provided in a way that can be sustained over time. A significant proportion of education interventions focus on the supply of education. But to achieve improved learning for children, children need to want to attend school and learn, usually given their families' demand for education. In turn, the relative value that people within the local neighborhood or community place on education vis a vis other important calls on their time and resources will influence students enthusiasm for learning within the classroom. As opposed to the supply of education, several interventions seek to support the demand for education, generating that enthusiasm that allows children to avail themselves of the educational opportunities that are offered.

A second dimension for organizing education interventions is whether or not they operate to affect individuals alone or groups of individuals. Given how education systems bring together individuals that then operate largely as a collective, it is also useful to differentiate between those interventions and research that focus on the individual within that system or on the collective of those individuals.

Innovations to improve education also operate on a third dimension. They may specifically provide resources to education actors or they may seek to influence those actors' incentives. For many years,

education programs focused on the former, identifying the lack of resources as the essential gap hindering learning. More recently, as South Asia’s learning outcomes have stagnated despite increased resources being devoted to the education sector this has increased awareness of the importance to understand the incentives that service delivery actors face. The World Development Report on “Making Services Work for the Poor” (World Bank, 2004) brought particular attention to those incentives, spurring more innovations in South Asia education that seek to improve incentives of key education actors. The third dimension of Figure 1 highlights the distinction between innovations focused on resources and those focused on incentives. However, it is often more difficult to disentangle innovations along this dimension, for resources are often applied to promote incentives for particular actors, and the incentives that actors face undoubtedly affect how or whether they use those incentives.

**Figure 1 – Typology of Actors Affecting Education Outcomes**



With these three dimensions of supply vs. demand, individual vs. collective, resources vs. incentives, we can organize education innovations into eight categories, illustrated by the cells of the cube in Figure 1. The front face of the cube is particularly useful for understanding education innovations. Individuals responsible for supplying education services are teachers whereas schools are collective entities that supply such services. The smallest unit of analysis for generating demand for education is

the household, while the collective interest of the community also has an impact on demand for education. These four categories – teachers, schools, households and communities – define the primary sets of actors that education innovations in South Asia seek to influence and are the primary focus of this analysis of the lessons learned from education impact evaluations.

Some clear examples illustrate the framework. A program that hires more teachers to work in schools affects the supply of education, focusing on the individual level. It would fall in the upper left box of Figure 1. A school construction program that aims to expand access to education by building more schools supports the supply side of education, providing a collective resource. We would classify it as a school focused intervention, falling into the upper right box. A conditional cash transfers (CCTs) program provides cash to households on the condition that children in those households attend schools. This type of intervention seeks to increase the demand for education, focusing on the household as the primary actor. Another category of intervention supports community awareness of the value of education and the capacity of that collective group to comment on whether the school is performing well. This demand-focused community accountability mechanism would fall in the lower right box of Figure 1.

Further, a vouchers program, that aims to expand the choice of households to send a child to a school of preference, is a demand-focused intervention giving resources directly to the household to buy quality services of their choice. Such interventions would fall in the lower left box of Figure 1. Performance-pay for teachers generally link teacher's pay with teacher presence in schools; other variants include bonus payments to teachers for meeting performance targets linked to students' test scores. This is a supply-side intervention for strengthening incentives of teachers to deliver quality education, and would fall in the upper-left corner in Figure 1. Another possible intervention is performance-based subsidies to schools. In such cases, schools are expected to perform at a certain level to get the per-student subsidy. We would classify it as a supply-focused intervention improving incentives of service providers.

### **3. Evidence**

#### **3.1. Systematic Review**

This systematic review strives to identify, appraise and synthesize rigorous education-related impact evaluations for South Asia. It will serve as the basis of the ensuing meta-analysis (results of which will be discussed in the next section) and cover all research that fulfills the following three criteria: (i) the research evaluates the impact of one or more clearly-defined education-related interventions, (ii) it measures effects on enrollment, attendance, test scores or other education-related outcome variables, and (iii) it uses data for at least one South Asian country. Additionally, in order to be included in the systematic review, a number of strict quality criteria also have to be satisfied. In particular, all causal statements need to be based on evidence gained from an RCT or a credible quasi-experiment, a well-defined “business-as-usual” control group has to be present and an intervention's effects have to be reported in a way that is transparent, standardized and comparable to effects reported by other studies.

A rigorous search and appraisal of the available evidence combined with these strict quality criteria results in a set of 29 distinct studies that document the results of rigorous impact evaluations of education interventions in South Asia. The specific interventions evaluated by these studies as well as the countries covered and methods used are listed in Table A. Additionally, the table lists whether a specific intervention primarily (i) targets individuals or a collective, (ii) addresses education supply or demand, and (iii) provides resources or incentives. Based on these categories introduced in Section 2 above and the four “actors” derived from them the remainder of this section will discuss the literature identified through the systematic review. More specifically, it will discuss specific research questions, methods and most importantly results of individual studies structuring this discussion by whether an intervention targets individuals on the supply-side (teachers), collectives on the supply side (schools), individuals on the demand side (households) or collectives on the demand side (communities). In the discussion, the evidence on education interventions from the 29 most rigorous studies at the heart of the systematic review will be supplemented by select descriptive or non-(quasi-)experimental work on South Asia, evidence related to service delivery outside of education that offers lessons for the sector, and, high-quality research from beyond the South Asia region.<sup>1</sup>

**Table A – Rigorously Evaluated Education Interventions in South Asia**

Study	Country	Method	Intervention	Supply vs. demand	Individual vs. collective	Resource vs. Incentive
Afridi (2010)	India	DD	midday meals	demand	individual	incentive
Andrabi, Das and Khwaja (2013)	Pakistan	RCT	report cards with school and child test scores	demand	collective	incentive
Aturupane, Glewwe, Kellaghan, Ravina, Sonnadara and Wisniewski (2013)	Sri Lanka	DD	report cards	demand	collective	incentive
			school management structures	demand	collective	incentive
Banerjee, Banerji, Duflo, Glennerster and Khemani (2010)	India	RCT	information on existing institutions	demand	collective	incentive
			training community members in a testing tool for children	demand	collective	incentive
			training volunteers to hold remedial reading camps	supply	individual	resource
Banerjee, Cole, Duflo and Linden (2007)	India	RCT	hiring of young women to teach students lagging behind in basic literacy and numeracy skills	supply	individual	resource
			computer-assisted learning program focusing on math	supply	collective	resource

<sup>1</sup> Appendix A provides more background on the systematic review. In particular, it explicitly describes the search strategy, explains the inclusion criteria in greater detail, and presents statistics on countries covered, methodologies used and other features of the 29 studies that document rigorously evaluated education interventions in South Asia. A critical assessment of the available evidence is made in Appendix B.

Study	Country	Method	Intervention	Supply vs. demand	Individual vs. collective	Resource vs. Incentive
Banerji, Berry and Shotland (2013)	India	RCT	adult literacy classes for mothers	demand	individual	resource
			training for mothers on how to enhance their children's learning	demand	individual	resource
Barrera-Osorio and Raju (2010)	Pakistan	RDD	bonus payments to teachers	supply	individual	incentive
			sanctions to schools	supply	collective	incentive
Barrera-Osorio, Blakeslee, Hoover, Linden, Raju and Ryan (2013)	Pakistan	RCT	publicly funded private primary schools	supply	collective	resource
Berry and Linden (2009)	India	RCT	actively recruiting children to attend bridge classes	demand	individual	incentive
			peer attending bridge classes	supply	collective	incentive
Borkum, He and Linden (2013)	India	RCT	school libraries	supply	collective	resource
Burde and Linden (2013)	Afghanistan	RCT	placing a school in a village	supply	collective	resource
Chaudhury and Parajuli (2010a)	Pakistan	DDD / RDD	CCT for girls	demand	individual	incentive
Chaudhury and Parajuli (2010b)	Nepal	RCT	devolution of school management responsibility	demand	collective	incentive
Das, Dercon, Habyarimana, Krishnan, Muralidharan and Sundararaman (2013)	India	RCT	school grants	supply	collective	resource
Duflo, Hanna and Ryan (2012)	India	RCT	verifying teacher's attendance through photographs and partly basing teacher salary on attendance	supply	individual	incentive
He, Linden and MacLeod (2008)	India	RCT	English education curriculum (different implementations)	supply	collective	resource
He, Linden and MacLeod (2009)	India	RCT	literacy skills development program (different implementations)	supply	collective	resource
Jayaraman and Simroth (2011)	India	DDD	midday meals	demand	individual	incentive
Lakshminarayana, Eble, Bhakta, Frost, Boone, Elbourne and Mann (2013)	India	RCT	supplementary remedial teaching by community volunteer, provision of learning material and additional material support for some girls	supply	individual	resource

Study	Country	Method	Intervention	Supply vs. demand	Individual vs. collective	Resource vs. Incentive
Linden (2008)	India	RCT	computer assisted learning program (different implementations)	supply	collective	resource
Muralidharan and Prakash (2013)	India	DDD / DDDD	bicycles for girls	demand	individual	incentive
Muralidharan and Sundararaman (2010)	India	RCT	diagnostic tests and feedback to teachers and monitoring of classroom processes	supply	individual	incentive
Muralidharan and Sundararaman (2011) and Muralidharan (2012)	India	RCT	bonus payments to teachers based on improvement in students' test scores (individual- or group-based)	supply	individual	incentive
Muralidharan and Sundararaman (2013a)	India	RCT	school choice program featuring lottery-based allocation of school vouchers	demand	individual	resource
Muralidharan and Sundararaman (2013b)	India	RCT	extra contract teachers	supply	individual	resource
Parajuli, Acharya, Chaudhury and Thapa (2012)	Nepal	RCT	community-driven development program centered around income generating activities	demand	collective	resource
Rao (2014)	India	DD	mixing wealthy and poor students	supply	collective	incentive
Sarr, Dang, Chaudhury, Parajuli and Asadullah (2010)	Bangladesh	DD	grants to schools	supply	collective	resource
			grants to schools and education allowances to students	supply / demand	individual / collective	resource / incentive

Importantly, while all studies included in the systematic review are selected because they report interventions' impacts on education-related outcome variables like enrollment, attendance or test scores, many also evaluate whether the interventions affected output variables such as teachers' attendance or time spent on student-centric activities. Because of the importance to understand the mechanisms behind an intervention's success or failure to address outcome variables, these output variables will also be part of the subsequent discussions. The same is the case for the cost-effectiveness of different types of interventions. In many ways, a program's cost-effectiveness has much more policy relevance than its "pure" effectiveness. Therefore, whenever information on costs is available, it will be mentioned. It should be noted, however, that details on costs are only reported in 13 out of the 29

studies that form the basis of the systematic review. A list of these 13 studies and the reported costs associated with the interventions they document can be found in Table B.<sup>2</sup>

**Table B – Costs of Rigorously Evaluated Education Interventions in South Asia**

Study	Interventions	Costs
Andrabi, Das and Khwaja (2013)	report cards with school and child test scores	USD 1 per child / USD 20 per marginal child enrolled
Banerjee, Cole, Duflo and Linden (2007)	hiring of young women to teach students lagging behind in basic literacy and numeracy skills	USD 2.25 per child per year
Banerjee, Cole, Duflo and Linden (2007)	computer-assisted learning program focusing on math	USD 15.18 per child per year
Barrera-Osorio, Blakeslee, Hoover, Linden, Raju and Ryan (2013)	publicly funded private primary schools	USD 18 per child per year
Chaudhury and Parajuli (2010a)	CCT for girls	USD 36 per child per year
He, Linden and MacLeod (2008)	English education curriculum (different implementations)	USD 11.20-20.46 per child per year
Lakshminarayana, Eble, Bhakta, Frost, Boone, Elbourne and Mann (2013)	supplementary remedial teaching by community volunteer, provision of learning material and additional material support for some girls	USD 48-61 per child per 18 months
Linden (2008)	computer assisted learning program (different implementations)	USD 5 per child per year
Muralidharan and Prakash (2013)	bicycles for girls	USD 12 per child per year
Muralidharan and Sundararaman (2010)	diagnostic tests and feedback to teachers and monitoring of classroom processes	USD 6 per child per 24 months
Muralidharan and Sundararaman (2011) and Muralidharan (2012)	group-based bonus payments to teachers based on improvement in students' test scores	USD 2 per child per year
Muralidharan and Sundararaman (2011) and Muralidharan (2012)	individual-based bonus payments to teachers based on improvement in students' test scores	USD 3 per child per year
Muralidharan and Sundararaman (2013a)	school choice program featuring lottery-based allocation of school vouchers	savings of 102 USD per child per year
Muralidharan and Sundararaman (2013b)	extra contract teachers	USD 6 per child per 24 months

A first strand of research investigates the roles of interventions targeting individuals on the supply-side of the education sector, i.e. teachers. Studies in this literature address questions such as whether there is a relationship between standard teacher variables – like a teacher’s formal education or experience – and learning outcomes, if non-traditional types of teachers (e.g. contract teachers with relatively low salaries, few formal qualifications and little job security) are as effective as regular teachers, and if monitoring teachers’ performance and linking their pay to performance can have positive impacts on learning or other outcome variables. Teacher-focused interventions are motivated by the fact that teachers’ efforts in classrooms are central to learning and that throughout the South

<sup>2</sup> It should be noted that some interventions – for instance those that change the composition of schools or classes by mixing wealthy and poor students – might in principle not involve any direct costs.

Asia region, one of the constraints to improving education outcomes seems to be the amount of effort that teachers exert. For example, throughout India but particularly in its low income states, teacher absenteeism averages over 20 percent and the fiscal cost of teacher absence is estimated to be around 1.5 billion US dollars per year (Chaudhury et al., 2006, and Muralidharan et al., 2014).<sup>3</sup>

More tangibly, interventions targeting individuals on the supply-side of the education sector usually center on an improvement in resources or inputs allocated to teachers or are motivated by an incentives-based approach in the spirit of the 2004 World Development Report (World Bank, 2004). Additionally, some aim to change teachers' behaviors based on insights derived from behavioral economics. Maybe the most traditional input-based approach is to hire better-educated and more experienced teachers in the hope that this will improve students' learning outcomes. For a while, the conventional wisdom has been that such standard measures of teacher quality usually fail to influence learning outcomes. In fact, experimental or quasi-experimental evidence on this topic is rather scarce. But regressions controlling for observable variables and pupil fixed effects indeed show that in Punjab, Pakistan, variables such as teacher certification and experience had no bearing on students' standardized marks (though they were important determinants of teachers' pay). In contrast, teaching "process" variables like lesson planning, involving students through asking questions during class and quizzing them on past material significantly influenced children's learning (Aslam and Kingdon, 2011). Similar regressions for India demonstrate that measures such as a teacher's MA qualification, pre-service teacher training and a first division in the teacher's own Higher Secondary exam could actually have payoffs in terms of higher student achievement, but that these payoffs depended very much on the school environments. In this particular context, the level of unionization appeared to be an especially important variable (Kingdon and Teal, 2010).

One sub-group of interventions focused on individuals on the supply-side of the education sector for which positive influences on learning outcomes have consistently been found supplements regular teachers with volunteers. A study for Uttar Pradesh, India, demonstrates that training volunteers to hold remedial reading camps significantly increased the reading skills of camp participants. The same study showed that hiring a *balsakhi*, i.e. a young woman from the community, to teach students lagging behind in basic literacy and numeracy skills proved successful as well. The *balsakhi* intervention increased average test scores in treatment schools by 0.14 standard deviations in the first cohort exposed to the program and by 0.28 standard deviations in another group of children that was exposed to it one year later. The positive learning impacts were mostly due to large gains experienced by children at the bottom of the test-score distribution and at costs of about two US dollars per child per year the intervention was relatively economical (Banerjee et al., 2007). In a different randomized control trial in Andhra Pradesh, India, volunteers were again trained to hold remedial lessons. In addition to that, children were also provided with learning materials. At costs of between 48 US dollars

---

<sup>3</sup> Besides teachers' effort, a wide range of other aspects of teachers' behaviors might have a positive or negative influence on education outcomes. One of these aspects is whether teachers are fair instead of biased towards certain groups of students. In this context, an RCT designed to test for discrimination in grading in India found that teachers gave exams purportedly coming from lower cast students scores that were 0.03 to 0.08 standard deviations lower than those they gave to exams that had been labeled as coming from high cast students (Hanna and Linden, 2012).



and 61 US dollars per child over the 18 month treatment period, the combined intervention was one of the costlier education interventions in South Asia (at least among the group of interventions that have been rigorously evaluated and for which data on costs are available). At the same time, however, it led to very large learning effects: In treated schools, mean test scores in both mathematics and language went up significantly. On average, composite test scores in this group increased by 0.75 standard deviations (Lakshminarayana et al., 2013).<sup>4</sup>

Interventions that relied on volunteers to supplement regular teachers led to significantly positive learning effects also in regions other than South Asia. In Chile, for example, a three-month program of small group tutoring in vulnerable schools that used college students as volunteer teachers was established. This intervention increased fourth graders' performance in a reading test by between 0.08 and 0.09 standard deviations. Though treatment effects for the average student were only marginally statistically significant, students from low-performing and poor schools exposed to the intervention increased their performance in a reading test by between 0.15 and 0.20 standard deviations. Their self-perceptions as readers went up as well (Cabezas, Cuesta and Gallego, 2011).

Non-traditional types of teachers are also deployed in another sub-group of interventions focusing on individuals on the supply-side. This sub-group evaluates the effect of replacing or supplementing regular teachers by contract teachers with usually much lower salaries and only short-term contracts. For instance in Andhra Pradesh, India, some schools were each provided with an extra contract teacher, at a cost of about three US dollars per child per year. Subsequently, average test scores in treatment schools increased by about 0.15 standard deviations in mathematics and 0.13 standard deviations in language relative to a control group of schools that had not received extra contract teachers. Non-experimental tests showed that contract teachers were much less likely to be absent from school and as effective in improving students' learning as regular teachers even though the latter's resumes generally characterized them as much more qualified and better trained. In addition to that, average salaries of contract teachers amounted to only about a fifth of those of civil-service teachers (Muralidharan and Sundararaman, 2013b).

Evidence not derived from (quasi-)experimental interventions confirms that the deployment of contract teachers can positively influence learning outcomes: Fixed-effects panel regressions of the education production function for Uttar Pradesh and Bihar, India, indicate that the hiring of contract teachers – who were being paid just a third of the salaries of regular teachers with comparable observed characteristics – raised students' overall test scores by between 0.07 and 0.21 standard deviations (Atherton and Kingdon, 2010). Moreover, according to a set of regressions controlling for observables and school fixed effects, contract teachers' attendance as well as teaching activity were significantly higher compared to those of regular teachers in Uttar Pradesh and Madhya Pradesh, also in India. Also in this case, the employment of contract teachers was rather economical; their average salaries were one-fourth or less of those of civil service teachers (Goyal and Pandey, 2009). At the same time,

---

<sup>4</sup> If in addition to the training of volunteers and the provision of learning materials, material support was given to girls, this had no further positive effects on test scores.

an increased reliance on contract teachers faces substantial obstacles in the political economy realm. Powerful teachers unions regularly lobby for the regularization of contract teachers' contracts and even the courts have at times constrained the appointment of contract teachers.

Overall, irrespective of whether particular interventions focused on the deployment of non-traditional types of teachers put more emphasis on the provision of additional resources or on changing teachers' contracts and incentives, they appear very promising. The same can generally be said about programs centered on monitoring teachers' performance and linking their pay to performance. This sub-group of individual supply-side interventions directly follows an incentive-based framework, and the most straightforward ones have involved bonus payments for individual teachers. Others have introduced pay-for-performance models for larger groups of teachers. These group bonus payments (or sanctions) can be motivated either by arguments drawn from behavioral economics – such as a perception of what constitutes a “fair” remuneration scheme – or be based on political economy realities on the ground, in particular the prominent role of teachers unions generally opposed to individual-based pay-for-performance models.

Both types of bonus payments were evaluated in Andhra Pradesh, India. The rigorous impact evaluation found that while the payment of group bonus payments to teachers based on improvements in students' test scores cost as little as about two US dollars per child per year, it increased average test scores by 0.14 standard deviations after one year. Effects on test scores were quite persistent: The intervention raised test scores by 0.16 standard deviations after two years, 0.14 standard deviations after three years and 0.19 standard deviations after four years. After five years, effects were no longer statistically significant. Individual-based bonus payments were even more effective in improving learning outcomes, especially in the longer run. They increased composite test scores by 0.16 standard deviations after one year, 0.27 standard deviations after two years, 0.20 standard deviations after three years, 0.45 standard deviations after four years and 0.44 standard deviations after five years. All these effects were statistically significant. Moreover, at about three US dollars per child per year, costs were again relatively low. There was no evidence that teacher bonus payments had any adverse consequences (like more of teachers' efforts being channeled towards incentivized subjects as opposed to non-incentivized ones such as science and social studies). In fact, incentive schools performed significantly better than those in a separate group of randomly chosen schools that had received additional schooling inputs of similar value (Muralidharan and Sundararaman, 2011; Muralidharan, 2012).

A study for Punjab, Pakistan, however, could not find any statistically significant incentive effects on learning for marginal failers when bonuses were paid to teachers as a group based on their students' test scores (Barrera-Osorio and Raju, 2010).<sup>5</sup> Some caution regarding the promise of direct teacher incentives is also warranted due to the somewhat mixed evidence that exists for countries from outside of South Asia. In Kenya, for example, rewarding primary school teachers for improvements in their

---

<sup>5</sup> The study's authors conjecture that frequent revisions of the bonus eligibility criteria and a botched communication strategy regarding these criteria may have generated uncertainty and discouraged teachers' effort.

students' test scores only led to gains on the narrow outcomes linked to the monetary incentives as opposed to a broader-based increase in students' human capital. Important output variables like teacher attendance and homework assignment were also unaffected by the intervention while the number of test preparation sessions did increase in treatment schools (Glewwe, Ilias and Kremer, 2010).

Altering teachers' salary structures in a way that links at least part of their pay to students' performance fundamentally changes their contracts and incentives. Monitoring teachers' behavior and potentially basing part of their salary on readily observable characteristics like classroom attendance might be classified as a related but much lighter type of individual supply-side intervention that could face less political economy constraints. One rigorously evaluated education intervention in South Asia involved monitoring teachers' behaviors but no direct monetary incentive: In Andhra Pradesh, India, teachers were provided with feedback and results of diagnostic tests on their students' learning outcomes and classroom processes were monitored. At costs of about six US dollars per child per year this intervention was relatively economical and it apparently led teachers to exert more effort in the classroom – at least when they were being observed. However, it had no significant effect on students' test scores (Muralidharan and Sundararaman, 2010).

An impact evaluation for Udaipur and Rajasthan, India, showed that the monitoring of teachers might be more effective in improving learning outcomes if it is combined with monetary incentives. When teachers' attendance was verified through photographs and salaries were partly based on attendance, this lowered teacher absenteeism by 21 percentage points and increased average test scores by 0.17 standard deviations (Duflo, Hanna and Ryan, 2012).

Rather promising evidence on the effects of monitoring also exists for a very different type of service provider: the police. In Rajasthan, India, visits of decoy surveyors pretending to register crimes at police stations – which had initially been conceived of primarily as a method of data collection for an evaluation of the impacts of other policies – led to a statistically significant improvement in police performance (Banerjee et al., 2012). At the same time, evidence from the health sector suggests that in the long run the monitoring of service providers might lead to problems of its own. One impact evaluation for India, in particular, analyzed the effects of monitoring the presence of government nurses in public health facilities combined with steps to punish the worst delinquents. Initially, the monitoring system was very effective but after a few months the local health administration appeared to undermine the scheme by letting nurses claim an increasing number of “exempt days.” Eighteen months after its inception, the program had become completely ineffective (Banerjee, Glennerster and Duflo, 2008).

A second group of interventions targets not individual teachers but larger collectives on the supply-side of the education sector. In other words, interventions in this group focus on schools or sometimes school administrations or principals. Typical examples include the placement of a new school in a village, updates to the curriculum, or the better use of information technology. A majority of the relevant programs are motivated by a traditional view on education that focuses on inputs and/or conjectures that learning outcomes are bound to improve if children are given the opportunity to go

to school. In contrast, only a minority of rigorously evaluated interventions targeting schools explicitly tries to influence the incentives or behaviors of school administrations or principals.

Some of the most straightforward input-oriented interventions targeting collectives on the supply-side of the education sector provide schools with unconditional block grants. In Andhra Pradesh, India, the introduction of such school grants increased average test scores by 0.08 standard deviations in language and 0.09 standard deviations in mathematics – but these effects only occurred in the first year the school grants were administered. In the second year, when families anticipated the grants, they reacted by offsetting their own spending and so the grants no longer exhibited any significant learning effects (Das et al., 2013). Similarly, in Bangladesh, an intervention that provided grants to some schools did not increase treated children’s test scores relative to those of children in control schools that had not received grants. It should be noted, though, that the Bangladeshi grants increased average enrollment probabilities by between nine and 18 percent (Sarr et al., 2010).<sup>6</sup>

If school inputs are made in kind, this does not appear more promising for improving education outcomes than giving cash grants: In impact evaluation of a school library program in Karnataka, India, for instance demonstrated that the program had had no positive impacts on students’ scores on a language skills test administered after 16 months. Depending on the exact implementation, language learning effects were sometimes even negative. Besides, there was no significant effect on test scores in other subjects or on attendance rates (Borkum, He and Linden, 2013).

However, other input-oriented interventions targeting collectives on the supply-side of the education sector appear more promising. Among those with the largest documented effects on both enrollment and learning is the placement of a new school in a village. With respect to this type of intervention, an RCT for Afghanistan showed that placing a school in a village for the very first time on average increased girls’ enrollment by 52 percentage points; girls’ average composite test scores improved by an impressive 0.65 standard deviations. While boys’ enrollment rose by an equally impressive 35 percentage points and average test scores for this group increased by 0.40 standard deviations, the effects on girls were so large that they eliminated the gender gap in enrollment and dramatically reduced gender differences in test scores. It goes without saying that Afghanistan is somewhat of a special case in that schools are much scarcer in this country than in many other parts of South Asia. Therefore, it appears questionable whether the placement of a new school in a village would be similarly high in other contexts (Burde and Linden, 2013).

It should be noted, though, that strong evidence for positive enrollment and learning effects of placing a school in a village or also an urban area exists for Pakistan, too. In the city of Quetta in Baluchistan, for example, school subsidies for new private girls’ schools increased girls’ enrollment by around 33 percentage points. Boys’ enrollment rose as well. This was partly because boys were also allowed to attend the new schools and partly because many parents would not send their daughters to school

---

<sup>6</sup> The study by Sarr et al. (2010) is not an RCT but relies on a quasi-experimental design. Since participating schools tended to be located in particularly disadvantageous situations, the authors conjecture that their results might actually understate the school grants’ true learning effects.

without also educating their sons (Kim, Alderman and Orazem, 1999). Likewise, in the province of Punjab, the establishment of publicly funded low-cost private schools had large positive impacts on the number of students, teachers, classrooms, and blackboards (Barrera-Osorio and Raju, 2011). Finally, in underserved rural districts of the neighboring province of Sindh, publicly funded private primary schools – which came at an average cost of about 18 US dollars per attending child per year – increased school enrollment in treatment villages by 30 percentage points. They also caused test scores of children in treatment villages to be on average 0.62 standard deviations higher than those of kids in control villages. Providing bigger financial incentives to schools that recruited girls did not lead to a greater increase in female enrollment than equally compensating schools for the enrollment of both boys and girls (Barrera-Osorio et al., 2013).

Another subgroup of input-oriented interventions targeting schools that has been analyzed by impact evaluations aims to make better use of information technology or to expand and update curricula. One impact evaluation in Gujarat and Maharashtra, India, for instance investigated the effects of a computer-assisted mathematics learning program. At costs of about 15 US dollars per child per year the program was relatively costly. But the learning effects it produced were substantial as well: Relative to the control schools, average math scores in treatment schools increased by an additional 0.35 standard deviations in the first year, 0.47 standard deviations in the second year, and a still significant 0.10 standard deviations one year after students had left the program (Banerjee et al., 2007). The establishment of a different computer-assisted mathematics learning program in Gujarat again demonstrated that this type of intervention can have rather large learning effects for relatively moderate costs. When the program was administered out of school, it increased average math scores by 0.27 standard deviations at costs of about five US dollars per treated child per year. However, when the program was administered in school, it actually decreased average overall test scores by 0.48 standard deviations and average math scores by 0.57 standard deviations. The author of the impact evaluation on the program's effects explains that the in-school implementation of the program was generally done in place of regular classes and that, apparently, computers are poor substitutes for regular teachers. She emphasizes that one lesson from her study is the importance of understanding how new technologies and teaching methods interact with existing resources (Linden, 2008).

Evidence from outside of South Asia and sectors other than education underlines the importance of developing a profound understanding of how innovative technologies influence existing resources and vice versa. In Colombia, for instance, the integration of computers into the teaching of language in public schools had little effect on students' test scores and other outcomes. The absence of an effect seemed to be due to the failure to incorporate the computers into the educational process (Barrera-Osorio and Linden, 2009). Similarly, evidence from Peru shows that the distribution of laptops to students under the slogan "One Laptop per Child" had no significant impact on enrollment or test scores in math and language. At the same time, the intervention did have positive effects on general cognitive skills as measured by Raven's Progressive Matrices, a verbal fluency test and a coding test (Cristia et al., 2012). Mixed evidence on the potential for technology to improve learning outcomes also comes from Ecuador. In this South American country, the roll-out of computer-aided instruction in mathematics and language in primary schools had a positive impact on average mathematics test

scores – which increased by about 0.30 standard deviations on average and by much more for those students at the top of the achievement distribution – but no statistically significant effect on language test scores (Carillo, Onofa and Ponce, 2010).

With regard to sectors other than education, rigorous evidence also exists on the impacts of using innovative technology in social protection. For India, this evidence relates to the introduction of smart cards – which coupled electronic transfers with biometric authentication – in two government welfare schemes. Even though the new smart card system was not fully implemented on the ground, it resulted in faster, more predictable, and less corrupt payments process without adversely affecting program access (Muralidharan, Niehaus and Sukhtankar, 2014).

Concerning interventions that expand and/or update the curriculum, the introduction of a revised English education curriculum in Maharashtra, India, increased average overall test scores by 0.26 standard deviations if implemented outside of schools and by 0.35 standard deviations if implemented inside schools or with the help of computers. Weaker students benefited more from those implementations that included teacher-directed activities while stronger students profited more from more self-paced machine-based implementations.<sup>7</sup> Depending on the exact mode of implementation, the teaching of the revised English education curriculum led to relatively substantial additional costs of between eleven and 20 US dollars per child per year (He, Linden and MacLeod, 2008). The implementation of a new literacy skills development program in the same Indian province of Maharashtra also had positive learning effects: Average reading scores of children in the treatment group increased by between 0.26 and 0.70 standard deviations depending on the exact mode of implementation. Overall, the program seemed to be more effective as a supplement to existing courses rather than as a primary means of instruction (He, Linden and MacLeod, 2009).

The promising evidence on interventions aimed at expanding and updating the curriculum from South Asia is borne out by findings from outside the region. In the Philippines, in particular, the introduction of a short-term reading program designed to provide age-appropriate reading material, the training of teachers in its use and the support of teachers' initial efforts for about a month led to short-term improvements in students' average reading skills of 0.13 standard deviations. While the program also led treated students to read more on their own at home, it did not exhibit statistically significant spillover effects on test scores in other subjects (Abeberese, Kumler and Linden, 2014).

As mentioned above, only a few rigorously evaluated interventions explicitly try to influence the incentives faced by school administrations or principals. This is in spite of the limited but very promising evidence that exists on the effectiveness of this type of intervention. In particular, the study for Punjab, Pakistan that evaluated the effects of group-based bonus payments to teachers already mentioned above also analyzed the effects of sanctioning schools based on students' test scores. It documented that the threat of sanctions had very strong incentive effects on learning for marginal

---

<sup>7</sup> This is consistent with the existing evidence on the heterogeneous effects of computer-aided mathematics instruction in Ecuador mentioned above (Carillo, Onofa and Ponce, 2010).

failers. These incentive effects were as high as 0.66 standard deviations (Barrera-Osorio and Raju, 2010).

The evidence on interventions that change the compositions of classes or track students according to their ability is also relatively scarce. An impact evaluation from Kenya suggests that students at all levels of achievement may gain from this kind of tracking: in Kenya, high achievers benefited from having more high-achieving peers in their classes but tracking also benefited lower-achieving students because it allowed teachers to better teach to their level (Duflo, Dupas and Kremer, 2011). For South Asia, not a whole lot is known on the impacts of class composition, tracking and peer effects. Only two rigorous impact evaluations on the topic exist. These show that in Delhi, India, mixing wealthy and poor students had substantial positive effects on the social behaviors of wealthy students, at the cost of negative but relatively modest impacts on academic achievement (Rao, 2014) and how in Haryana, also in India, peer networks had significant impacts on enrollment and attendance patterns of children in a community-based education program (Berry and Linden, 2009).

While the interventions discussed so far primarily aimed to improve education outcomes in South Asia by addressing supply-side challenges and bottlenecks, at least implicitly most of them also had the demand side in mind. This is because an – often tacit – assumption underlying the supply-side interventions targeting individuals or collectives (i.e. teachers or schools) is that offering a higher quality education and training should improve the utility and returns associated with education. This in turn should incentivize households or families to send their children to school and reduce the risk that children drop out at an early age because spending time at school is not seen as a useful activity for building human capital.

In parallel, there also exists a set of impact evaluations on interventions directly targeting the demand side. The most straightforward of these interventions are concerned with the education decisions made by individuals or small units like households or families. Interventions in this group typically either provide households with inputs or resources that lower the direct or indirect costs of schooling or offer incentives if households keep their children in school. Typical examples include the provision of cash, free tuition, or goods meant to reduce the cost of schooling to families, mothers or school children under the condition that children enroll in or attend school. In Bangladesh, for instance, the program already mentioned above that provided grants to schools also included education allowances to students. Disappointingly, an impact evaluation showed that the education allowances were as ineffective as the school grants in influencing enrollment numbers or other education outcomes (Sarr et al., 2010). Similarly, and again in Bangladesh, the Female Secondary School Assistance Project (FSSAP) featured a free tuition policy for female secondary school students and, again, an impact evaluation could not find any significant impacts on children's educational attainment (Hong and Sarr, 2009).

The same impact evaluation that analyzed the free tuition policy for female secondary school students also evaluated a different component of FSSAP: a conditional cash transfer (CCT) program featuring stipends for female students. These stipends constituted the first CCT program implemented in the developing world and in contrast to the ineffective free tuition policy, they had positive effects on

education and other indicators. In particular, they contributed to raising girls' years of education by up to two years and each additional year of education led to an increase in the labor force participation rate of married women by between 2.4 and 5.3 percent. Moreover, the stipend program also raised the age of marriage of women by 1.4 to 2.3 years; there are indications that it impacted the age of marriage of men as well (Hong and Sarr, 2009).

Promising evidence on the enrollment effects of CCT programs that give monthly stipends to girls also exist for Pakistan. The Punjab Female School Stipend Program, in particular, has been one of the most extensively evaluated education programs in South Asia. The substantial body of research that exists on the program has documented quite an array of positive effects: There is strong evidence that stipends led to a substantial increase in female school enrollment and completion as well as to fewer teenage pregnancies and later marriages. In addition to that, somewhat weaker evidence exist that stipends had positive spillover effects on boys. At the same time, and mostly in reaction to the program's large enrollment effects, it seems that it also contributed to a rise in student-teacher ratios. Finally, from a cost-benefit perspective it is interesting to note that the baseline stipends amounted to a little over 36 US dollars per child per year. This is one of the highest costs recorded for any of the rigorously evaluated education-related interventions in South Asia for which comparable cost estimates are available. A higher benefit level in some areas relative to others had little effect on school enrollment and retention (Chaudhury and Parajuli, 2010a; Hasan, 2010; IEG, 2011; Barrera-Osorio and Raju, 2012).

In general, the literature on South Asia seems to suggest that CCT programs often increases enrollment but that their impacts on learning outcomes tend to be limited or non-existent. This conclusion is also supported by evidence from outside the region. In China, for example, the introduction of a CCT program with payments conditional on continued enrollment in schools reduced dropout rates by 60 percent but did not influence children's test scores (Mo et al., 2011). Similarly, in Mexico, the launch of a CCT program increased average schooling by more than two months. But the program had no significant impact on average achievement scores in reading, writing, and mathematics skills tests (Behrman, Parker and Todd, 2011). Evidence on how cash or even non-cash influence households' behaviors also exists for the health sector. In Rajasthan, India, the combination of immunization camps with modest, non-cash demand-side incentives proved much more effective in raising the rate of fully immunized children than an intervention that focused on the supply-side only (Banerjee et al., 2010b).

Similar conclusions as for CCT programs can be reached for other types of interventions targeting individuals on the demand side that provided households not with cash but with goods or services meant to reduce the cost of schooling (sometimes, these interventions are called conditional kit transfers or CKTs). In Bihar, India, for instance, providing female students with a bicycle increased girls' age-appropriate enrollment in secondary school by 30 percent and reduced the gender gap in age-appropriate secondary school enrollment by 40 percent. At expenses of about twelve US dollars per child per year, these enrollment gains were achieved in a relatively cost-effective way. Moreover, the bicycle program also increased the number of girls who appeared for the school-leaving exam.



However, it did not have an impact on the number of girls who actually passed this exam (Muralidharan and Prakash, 2013). Also in India, the introduction of a mid-day meal program instead of the monthly distribution of free food grains substantially increased primary school enrollment and/or attendance. But, again, it had no significant learning effects (Afridi, 2010; Jayaraman and Simroth, 2011).

The evidence on interventions that provide households with goods or services conditional on children's continued enrollment or attendance in school available for South Asia is broadly in line with what has been found for other regions or sectors. In Kenya, for instance, a program that paid for textbooks, classroom construction, and the uniforms that parents in Kenyan schools are required to purchase led to a sharp reduction in dropout rates. It had no statistically significant effects on test scores (Kremer, Moulin and Namunyu, 2003). More promising evidence on learning effects of in-kind transfers to households or families, though, comes from another intervention carried out in Kenya. This intervention centered on the distribution of school uniforms to children in poor communities. It was able to reduce school absenteeism by 44 percent for the average student and by 62 percent for students who had not previously owned a uniform. Moreover, the intervention raised test scores for recipients of school uniforms by 0.25 standard deviations in the first year after its inception (Evans, Kremer and Ngatia, 2009).

A different strand of individual-level demand-side interventions focuses not on children or household as a whole. Instead, the emphasis is on mothers. This is motivated by the assumption that mothers generally have a preference for helping their children learn but might often lack the skills or experience to do so in practice. Against this backdrop, a group of mothers in Bihar and Rajasthan, India, were (i) trained on how to enhance their children's learning scores, (ii) offered adult literacy classes or, (iii) received both interventions concurrently. In all three treatment groups, children's average math scores increased. While – at 0.04 to 0.06 standard deviations – learning effects were rather modest, the interventions also significantly increased women's empowerment, mother participation in child learning and the presence of education assets in the home. Effects on children's literacy scores were not statistically significantly different from zero (Banerji, Berry and Shotland, 2013).

In addition to that, a number of impact evaluations exist for a related group of interventions that tries to improve education outcomes by investing into children's or their families' health and well-being.<sup>8</sup> In particular, following the pioneering impact evaluation by Miguel and Kremer (2004) in Kenya, deworming has for the last ten years been seen as one of the most effective and cost-effective interventions for improving education outcomes. This view has led to scaled-up deworming projects in many parts of the world, including in Andhra Pradesh and Bihar, India. Further impact evaluations have also followed. One of these analyzed a randomized health intervention in Delhi, India, that combined the delivery of deworming drugs with the provision of iron supplementation to Indian preschool children. The impact evaluation found that this intervention led to weight increases among

---

<sup>8</sup> This group of interventions is not included in the systematic review and ensuing meta-analysis because even though impacts on education outcomes are measured, as interventions are not directly related to education but rather to health.

children in the treatment group and also to a rise in preschool participation rates by 5.8 percentage points (Bobonis, Miguel and Puri-Sharma, 2006). Promising evidence for certain health interventions' impacts on education outcomes also exists for China, where offering free eyeglasses to primary school students increased average test scores of students with poor vision who wore the eyeglasses for one year by between 0.15 and 0.22 standard deviations (Glewwe, Park and Zhao, 2012).

In yet another group of interventions centered on individuals on the demand side of the education sector, effects of increasing households' or children's incentives to learn, of active encouragements to have children attend school and of promoting school choice and non-state education providers have been evaluated. Incentives to learn offer cash or in-kind incentives to households in return for children's school achievements and have become increasingly prominent due to encouraging impact evaluation results. A randomized evaluation of a merit scholarship program in Kenya, for instance, showed substantial exam score gains for girls in the treatment group. In this group, teacher attendance also improved in program schools and there were even positive externalities for girls unlikely to win a merit scholarship in the first place (Kremer, Miguel and Thornton, 2009). For now at least, evidence from South Asia on the effects of this kind of intervention is scarce. One impact evaluation for Haryana, India, was not included in the final sample of the systematic review because it contained no "business-as-usual" control group. This impact evaluation might still offer interesting lessons, though: It analyzed the effects of offering free after-school classes and incentive to learn. The incentives to learn were randomly assigned either to a parent (in the form of cash) or the child (in the form of toy of equal value). The impact evaluation found that the incentives increased average reading scores by a very substantial 0.53 standard deviations. In the aggregate, it did not matter if the recipient was a parent or the child (Berry, 2014).

Evidence for South Asia is also rather scarce for interventions where households are actively encouraged to send their children to school though one data point for Haryana, India exists: The impact evaluation by Berry and Linden (2009) already mentioned above which mainly focused on peer networks among out-of-school children also tested whether actively recruiting children to attend bridge classes had direct effects on enrollment levels in such informal classes designed as a bridge to the formal school system. The impact evaluation found that the intervention increased participation in bridge classes among out-of-school children whose parents had earlier indicated an interest in sending their children to participate in the classes by 30 percentage points.

On the issue of promoting school choice and non-state education providers, a bit more evidence exists for South Asia. This has to be interpreted against the background that – as pointed out above – across the South Asia region there are instances where institutions other than the state play an important direct role in the education sector. For example, non-state actors provide up to 23 percent of primary education enrollment in Pakistan's Punjab province (Barrera-Osorio et al., 2013). At the secondary level in India, enrollment in the non-state sector exceeds 20 percent with even higher rates in some states (Linden, 2012). Given this prominence of non-state actors, a number of impact evaluations have asked if further promoting households' school choice has the potential to improve access to and quality of learning.

In Andhra Pradesh, India, in particular, an impact evaluation analyzed the launch of a school choice program featuring a lottery-based allocation of school vouchers. It found that the program increased average composite test scores of lottery winners by 0.13 standard deviations as compared to those of lottery losers. The test scores of those students that ended up attending private schools because of the lottery increased by 0.23 standard deviations. There was no evidence of either positive or negative spillover effects on public-school students who had not applied for the vouchers, or on students who had started out in private schools to begin with. Moreover, the mean cost per student in the private schools in the sample was less than a third of the cost in public schools. This led to savings of about 102 US dollars per child per year (Muralidharan and Sundararaman, 2013a).

Rather promising evidence on the use of school choice programs featuring non-state education providers also comes from Pakistan and from outside the South Asia region. One study shows that in Punjab, Pakistan, government schools on average required twice the resources to educate a child than low-cost private schools. At the same time, test scores of children studying in private schools were at least as high as those of children in public schools. At least partly, this discrepancy in learning outcomes seemed to be due to teachers in private schools exerting relatively greater effort (Andrabi et al., 2007). In Colombia, an impact evaluation analyzed the introduction of a school choice program featuring the lottery-based allocation of scholarships for private schools. It documented that the program induced lottery winners to achieve about 0.20 standard deviations higher average test scores than lottery losers three years after the lottery. Moreover, there was also some evidence that lottery winners were comparatively less likely to marry or cohabit as teenagers (Angrist et al., 2002).

Beyond direct incentives that teachers or school administrations face through the contracts or conditions of their employment and direct demand side interventions altering the incentives of children or their families, there are also accountability mechanisms related to collectives on the demand side of the education sector that may affect education outcomes in South Asia. Multiple studies consider the impact of right to information innovations and direct community involvement approaches on strengthening schools' accountability on learning outcomes. Often taking opportunities from what is happening in the region, interventions in this category range from the relatively straightforward provision of information to the school community about a school's performance to more comprehensive interventions combining different aspects of citizens' engagement and demand side incentives.

In Uttar Pradesh, India, information on existing educational institutions was provided to the community. This intervention was subject to a rigorous impact evaluation which showed that it had no statistically significant effect on children's school attendance or learning levels as measured by reading or mathematics test scores (Banerjee et al., 2010a).<sup>9</sup> Likewise, in Sri Lanka, giving a school community regular information on their school's performance through a "report card" had no effect

---

<sup>9</sup> However, since no impacts on intermediate input and output measures like community involvement or awareness and teacher effort could be detected either, doubts might exist about whether the intervention was actually implemented on the ground.

on test scores. Moreover, the intervention did not have any impact on teacher absences, homework assignments or other teacher- or school related input and output variables (Aturupane et al., 2013). In Pakistan, though, “report cards” with school and child test scores proved to be effective. Such report card increased mean test scores in treatment villages in the province of Punjab by 0.11 standard deviations, decreased private school fees by 20 percent and raised primary school enrollment by five percent. At an average cost of about one US dollar per child – i.e. about 20 US dollars per child additionally enrolled in school – the intervention was also rather inexpensive (Andrabi, Das and Khwaja, 2013).

In Karnataka, Madhya Pradesh and Uttar Pradesh, India, a community-based information campaign had decidedly mixed results. In some of the states and/or grades where it was undertaken, it had a large positive impact on teachers’ attendance and engagement in teaching activities (which increased by up to 27 percent). Significant effects on students’ reading and especially mathematics skills could also be documented for a number of states and/or grades. For example, in grades three and four in Uttar Pradesh, the share of children who were able to write words and sentences and to do divisions increased by 37 percent and 60 percent, respectively. In many other cases, however, effects of the community-based information campaign were not statistically significant (Pandey, Goyal and Sundararaman, 2008 and 2011).

Also in Uttar Pradesh, India, community members were trained in a testing tool for children. As was the case with the intervention consisting of the provision of information on existing educational institutions to the community already mentioned above and analyzed by the same study, the training of community members had no effect on community involvement, teacher effort or children’s school attendance and reading or math levels (Banerjee et al., 2010a).

More comprehensive interventions combine different aspects of citizens’ engagement and demand side incentives. In Sri Lanka, for instance, a program established school management structures and provided training and support services to increase the participation of parents and the community in school management. A rigorous impact evaluation showed that this program increased average mathematics test scores by 0.22 standard deviations. Discussions with stakeholders suggested that a range of processes – including better teacher and parental involvement with the children, both at school and in the home – had likely contributed to this outcome (Aturupane et al., 2013).

A similarly comprehensive program was initiated in Madhya Pradesh, India. It included the creation of awareness on child development issues, the strengthening of linkages between different service providers, the intensification of connections between the community, local governments and service providers, the facilitation of the formation of village resource groups, the development of integrated village-level action plans, and advocacy and lobbying with the administration for flexible resource allocation. An impact evaluation demonstrated that this comprehensive set of interventions significantly improved children’s learning achievements (Sankar, 2013). Finally, in Nepal, the devolution of school management responsibilities to communities also had a statistically significant impact on certain schooling outcomes related to education access and equity. For instance, the intervention reduced the share of out-of-school children by 14.5 percent points in the overall sample

and by 16.6 percentage points for disadvantaged castes or ethnic groups. At the same time, it failed to significantly improve learning outcomes (Chaudhury and Parajuli, 2010b).

A number of similarly comprehensive collective-demand-side interventions have been evaluated in countries from regions other than South Asia. In Mexico, for example, doubling grants to parent associations in indigenous and general schools reduced grade failure by 7.4 percent and grade repetition by 5.5 percent in grades one to three and led to overall improvements in learning outcomes of more than 0.2 standard deviations. Training parents in organizing themselves had effects that were even higher than what was found for these cash grants (Gertler, Patrinos and Rodríguez-Oreggia, 2012). In Indonesia, simple block grants and/or training aimed to reinforce existing school committees demonstrated limited or no effects. However, measures that fostered outside ties between school committees and other parties led to greater engagement by education stakeholders and to improved learning outcomes. Linking school committees with village councils improved average test scores by 0.17 standard deviations compared to the intervention consisting of block grants to school committees. At 0.22 standard deviations, improvements in test scores were even more substantial when linkages between school committees and village councils were combined with district-level trainings of school committee members (Pradhan et al., 2011).

Altogether, it appears that the “collective demand” route of accountability for schools appears relatively promising but that rather limited interventions relying mostly or exclusively on the provision of information might not be sufficient for eliciting significant improvements in the access and quality of education. In contrast, more comprehensive interventions that combine different aspects of citizen’s engagement and demand side incentives often have strong positive effects on enrollment, attendance and learning outcomes. Promising but very limited evidence also exists on comprehensive community-centric interventions that do not directly incentivize communities to hold education providers accountable but rather provide communities with resources. An impact evaluation of a community-driven development program in Nepal that centered on income generating activities showed that this program led to a 15 percentage point increase in the school enrollment rate among children aged six to 15. Other impacts included an increase of real per capita consumption by 19 percent and a 19 percentage points decline in the incidence of food insecurity (defined as food sufficiency for six months or less). The impact evaluation did not investigate whether the intervention had effects on learning outcomes (Parajuli et al., 2012).

Similar conclusions to those applicable to more or less comprehensive community-focused interventions in the education sector can also be drawn from evidence from other sectors, in particular public health. In Bangladesh, for example, the natural occurrence of arsenic in groundwater is an important public health concern since the long-term exposure to arsenic in drinking water has been linked to several health risks. The issue has been the subject of a number of interventions focusing on the community that have been evaluated by rigorous impact evaluations. In one intervention, the populations in treatment villages were provided with nuanced risk information about arsenic levels in drinking water. This had no effect on whether households switched from a contaminated well to a safer one (Bennewar et al., 2013). However, a different and arguably more comprehensive intervention

that combined water arsenic testing with education regarding health implications of chronic arsenic exposure and methods to limit exposure, reduced arsenic exposure provided there were arsenics-safe drinking water sources available (George et al., 2012).

The study on the performance of the police force in Rajasthan, India, already mentioned above also provides evidence pointing to the limited effectiveness of interventions that solely rely on the provision of information to collectives on the demand side: when community observers were randomly placed in certain police stations, this did not have any robust impacts on police performance or the public's perception of the police. The intervention's failure was likely due to constraints on local implementation (Banerjee et al., 2012).

### **3.2. Meta-Analysis**

A meta-analysis relies on statistical techniques to combine results from two or more individual interventions. The objective is to improve the precision of estimated treatment effects and to assess whether these treatment effects vary for different types of interventions (Egger, Davey Smith and Phillips, 1997). Thus, meta-analyses are an ideally suited tool to utilize the education interventions in South Asia reviewed in the last section to derive rigorous and robust policy recommendations.

In the context of a meta-analysis, four methodological issues are of paramount importance: (i) the selection of the underlying evidence, (ii) the grouping of individual interventions, (iii) the weighting of the evidence, and (iv) the selection of adequate outcome measures.

- (i) With regard to the selection of the underlying evidence, the meta-analyses described here are based on the set of studies listed in Table A. That is, only those 29 studies are included in the meta-analyses that evaluate the impact of a clearly-defined education-related intervention on an equally well-defined education-specific outcome variable, contain data for at least one South Asian country, and satisfy strict quality and reporting criteria.
- (ii) Based on the conceptual framework introduced in section 2, interventions are grouped according to the primary actor that is targeted by intervention. Therefore, the meta-analyses distinguish between four types of interventions: Those addressing teachers, schools, households and communities, respectively. Indirectly, this also allows an assessment of whether interventions addressing education supply (teachers and schools) or demand (households and communities) appear more successful in raising education outcomes and if programs targeting individuals (teachers and households) or those addressing collectives (schools and communities) have shown greater promise. The exact mappings of interventions into these categories are again listed in Table A.
- (iii) Standard errors of interventions' impacts are used for weighting the evidence. In other words, interventions for which impacts have been more precisely estimated are given larger weights as compared to those interventions where the exact size of the impact is less clear.
- (iv) Adequate outcome measures should be pertinent to the question under study, widely used in the relevant literature, and easily comparable across different interventions. For these reasons,

three distinct meta-analyses are undertaken. The first two are concerned with interventions' impacts on standardized test scores in children's native language and mathematics. Additionally, a meta-analysis for effects on overall or composite test scores is being done.

Of course, meta-analyses with different specifications than the one reported here would be conceivable. In particular, some of the studies mentioned in the last section list interventions' impacts on outcome variables other than children's test scores in mathematics, language or a composite measure of learning. Examples of alternative outcome variables used in the literature include standardized test scores in English, and school enrollment and attendance. Many of these other outcome variables are very pertinent for access and quality of education in South Asia. However, comparable evidence on how different interventions impact other outcome is available for such a small number of studies that it seems unfeasible to include them in a meta-analysis. The same caveat applies to meta-analyses of interventions' heterogeneous effects on different subgroups (like boys and girls) and of their costs. As mentioned above, a program's cost-effectiveness is a variable that in many ways has more policy relevance than its "pure" effectiveness. But as also pointed out above, cost estimates are available for less than half of the rigorously evaluated education interventions in South Asia (cf. Table B). What is more, there are multiple reasons why it is extremely challenging to directly compare cost estimates stemming from different countries, time spans and types of interventions (Dhaliwal et al., 2012). Therefore, the meta-analyses reported here will abstract from costs or only mention them in passing.

The main messages of the three meta-analyses of education interventions in South Asia are visualized in Figures 2 to 4.<sup>10</sup> These figures are so-called forest plots and contain two different pieces of information: On the left-hand side, all interventions included in the corresponding meta-analysis are listed. These are identified by the author(s) and year of the underlying study as well as the specific intervention and grouped by actor. On the right-hand side, these interventions' impacts on the outcome variable are visualized. For each individual intervention, the horizontal line gives a 95 percent confidence interval for the intervention's impact on the relevant outcome variable. Besides, the solid black diamond indicates the point estimate of each intervention's impact. Again for each intervention, the size of the grey rectangle is determined by the relative precision of the respective estimate. It thus represents the weight any particular data point is given in the meta-analysis.

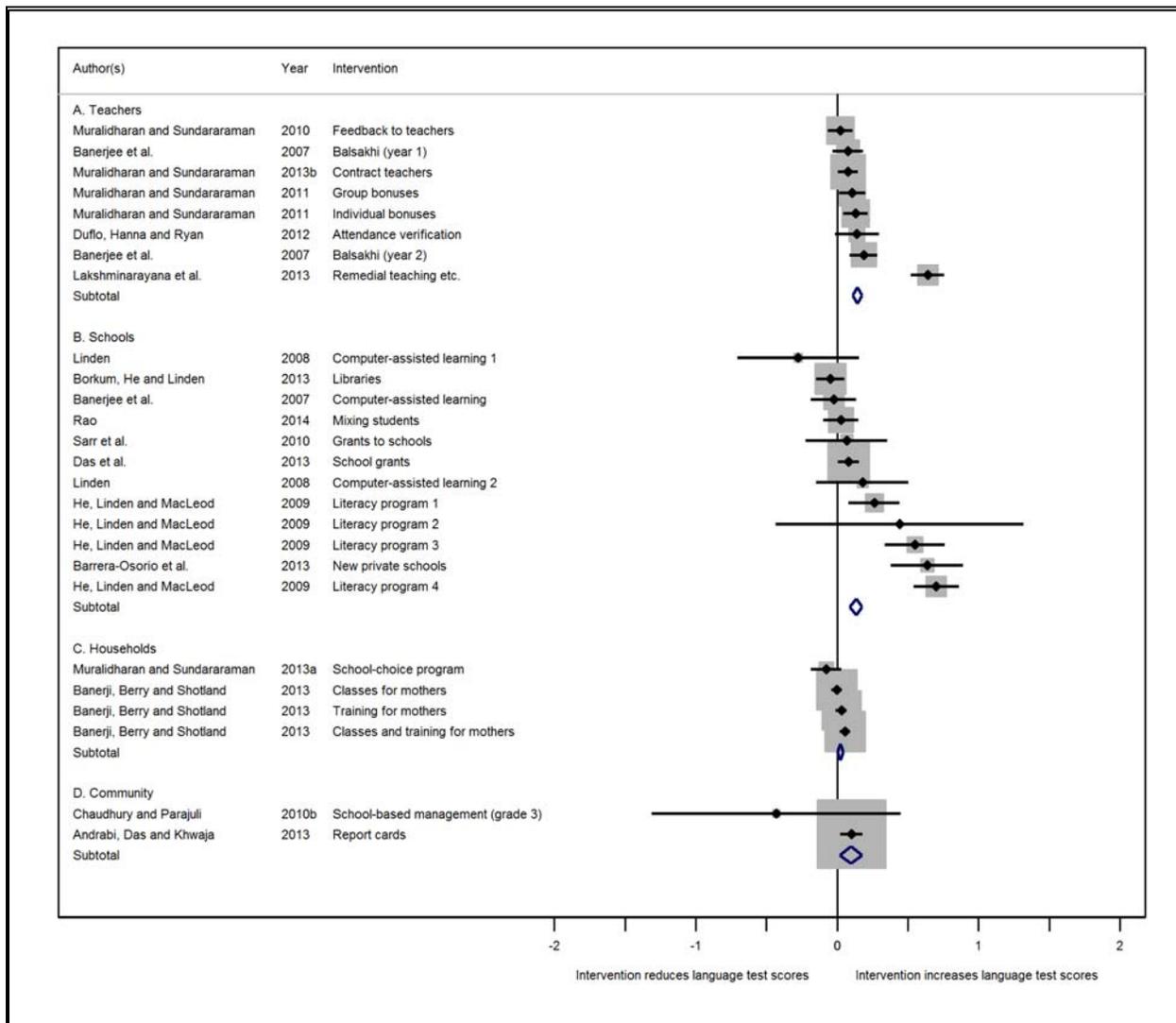
For each of the four groups of interventions (i.e. those targeting teachers, schools, households or communities), the row marked "Subtotal" lists and visualizes the results of the actual meta-analyses. The visualization of results again contains different elements: the center of the transparent diamond indicates the meta-analysis' best estimate of the impact of a particular group of interventions on the corresponding outcome variable. The spread between the diamond's left and right edge gives a 95 percent confidence interval for this impact. In case the diamond overlaps with the black vertical line originating at zero, the impact of a particular group of interventions on the outcome variable is not

---

<sup>10</sup> Details on the meta-analyses can be found in Appendix C which elaborates on the methodology underlying the meta-analyses and lists a table with detailed outputs.

statistically significantly different from zero (at least not on the five percent level of statistical significance). In contrast, if the diamond and the line do not overlap, the meta-analysis reveals that a particular group of interventions has a statistically significant impact on the outcome variable.

Figure 2 – Meta-Analysis of Interventions’ Impacts on Native Language Test Scores



**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates’ weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds’ left and right edges 95 percent confidence intervals; the sample includes both RCTs and quasi-experiments.

**Source:** World Bank staff calculations based on studies listed in Table A.

Figure 2 contains the results of the first meta-analysis. This meta-analysis synthesizes the impacts of interventions targeting teachers, schools, households or communities on children’s native language



test scores. Depending on the specific intervention, the exact native language used in the construction of the outcome variable varies. Examples of languages in which tests underlying the meta-analysis have been administered include Bangla (Sarr et al., 2010), Telugu (Muralidharan and Sundararaman, 2010) and Urdu (Andrabe, Das and Khwaja, 2013). As Figure 2 shows, the meta-analysis of interventions' impact on children's native language test scores contains eight interventions in the "teachers" category. Twelve interventions relate to schools or school administrations, four interventions center on households and two interventions address the community as a whole.

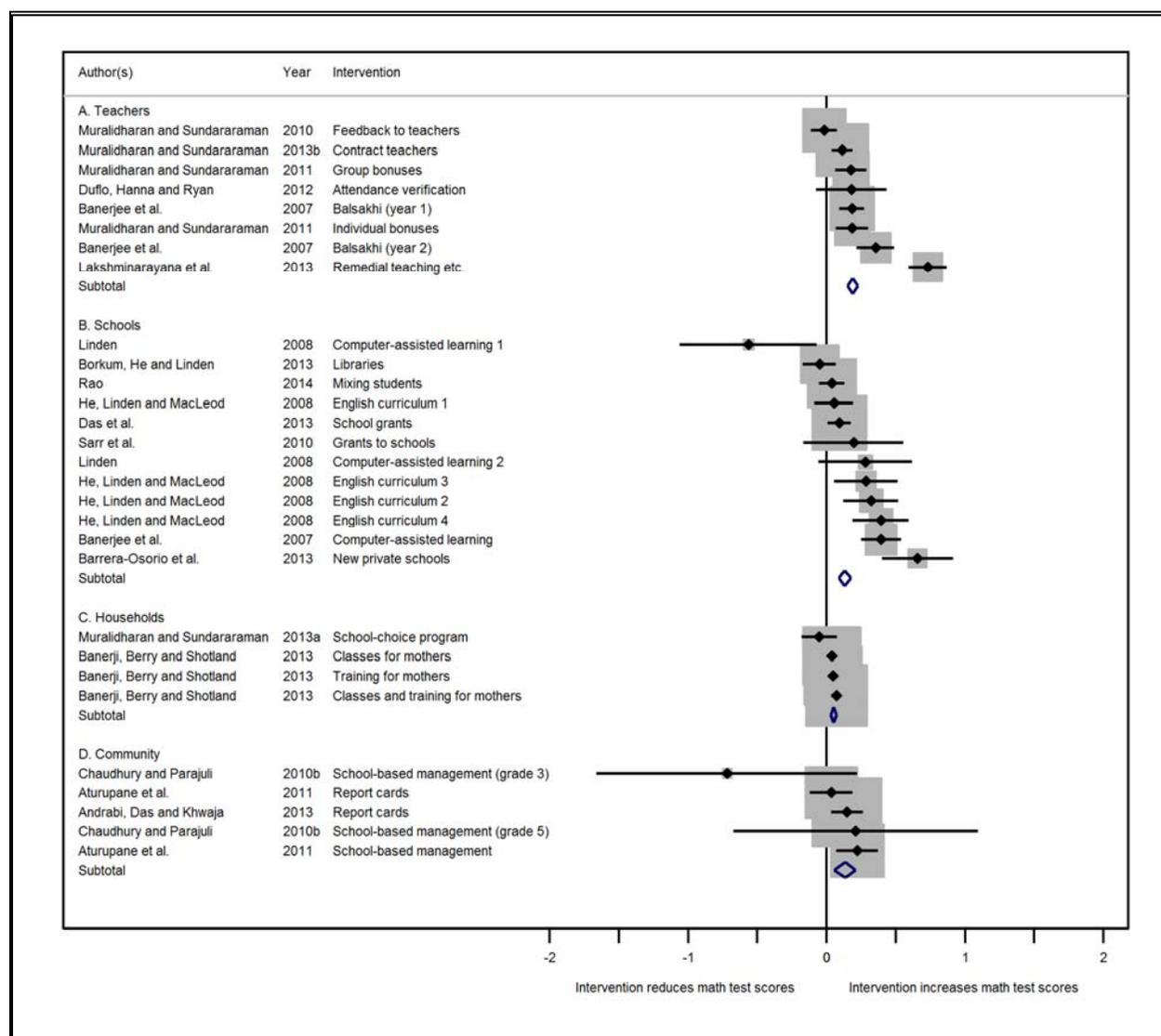
In terms of results, all four groups of interventions have a statistically significant impact on native language test scores: Overall, teacher-centric interventions increase average native language test scores by 0.14 standard deviations, interventions targeting schools raise them by 0.13 standard deviations and those addressing the whole community lead to an increase of average native-language test scores by 0.10 standard deviations. Interventions focusing on households are revealed to have only a small impact on average native-language test scores. Moreover, on the five percent level this impact is only just statistically significantly different from zero.

Among all 27 interventions included in the meta-analysis of impacts on native language test scores, the most effective ones are the establishment of publicly funded private primary schools (Barrera-Osorio et al., 2013) which raises test scores by 0.64 standard deviations, the provision of supplementary remedial teaching by community volunteers combined with the distribution of learning material and additional material support for some girls which leads to an increase in test scores of about again 0.64 standard deviations (Lakshminarayana et al., 2013), and one particular implementation of the introduction of a new English education curriculum which boosts native-language test scores by 0.70 standard deviations (He, Linden and MacLeod, 2009). The first and third of these interventions focus on schools; the second one is mainly a teacher-centric intervention.

In Figure 3, the results of a second meta-analysis are reported. This meta-analysis compiles the impacts of different interventions on children's mathematics test scores. As with the meta-analysis of impacts on native language test scores, this second meta-analysis is again based on exactly eight interventions directly addressing teachers' incentives or behaviors, twelve interventions related to schools or school administrations and four interventions centered on households. However, the number of interventions addressing the broader community is now larger. This is because more impact evaluations of interventions in this group analyze impacts on math test scores. As a result, Figure 3 lists five distinct interventions connected to right to information innovations, community involvement approaches, or similar programs.

Figure 3 shows that all four groups of interventions have a statistically significant influence on average mathematics test scores: Interventions focusing on teachers have the largest effect, raising math test scores by 0.19 standard deviations. But at 0.13 standard deviations, the effects on average math test scores of both school- and community-centric interventions are substantial as well. Interventions targeting households also increase math test scores in a statistically significant way, but the meta-analysis reveals that their impact is relatively small (0.05 standard deviations).

Figure 3 – Meta-Analysis of Interventions’ Impacts on Math Test Scores



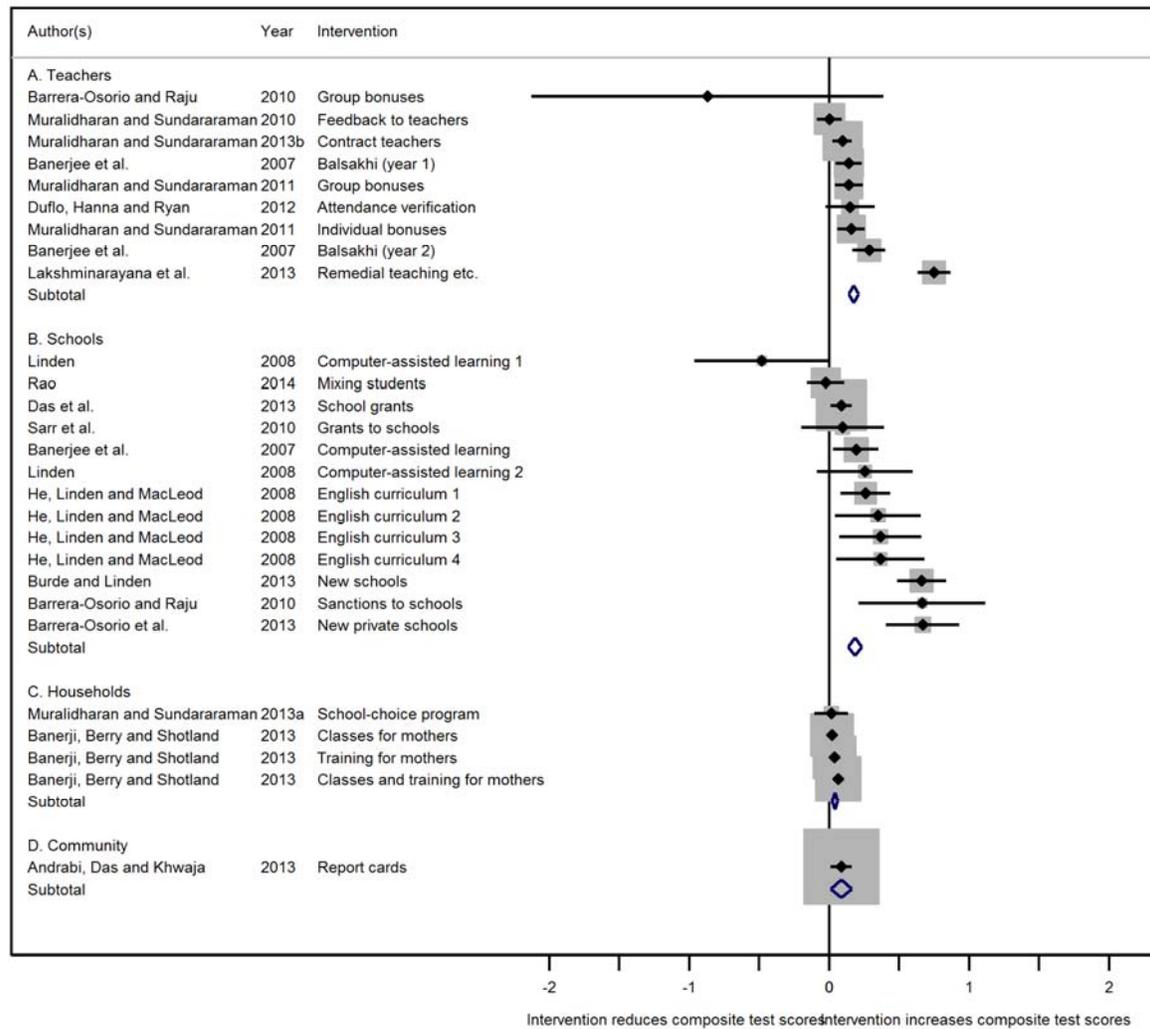
**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates’ weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds’ left and right edges 95 percent confidence intervals; the sample includes both RCTs and quasi-experiments.

**Source:** World Bank staff calculations based on studies listed in Table A.

As is the case for the meta-analysis of interventions’ impacts on native language test scores, the most promising interventions with regard to raising math test scores are the establishment of publicly funded private primary schools (Barrera-Osorio et al., 2013) and the provision of supplementary remedial teaching by community volunteers (Lakshminarayana et al., 2013). The establishment of private primary schools increases average test scores in mathematics by 0.66 standard deviations, and the provision of supplementary remedial teaching induces an average increase in math test scores of

0.73 standard deviations. Both these effects are even higher than those recorded for the respective interventions' impacts on native language test scores.

**Figure 4 – Meta-Analysis of Interventions' Impacts on Composite Test Scores**



**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates' weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds' left and right edges 95 percent confidence intervals; the sample includes both RCTs and quasi-experiments.

**Source:** World Bank staff calculations based on studies listed in Table A.

Finally, in Figure 4 the results of the third meta-analysis are depicted. This meta-analysis is concerned with different interventions' impacts on children's overall test scores. Overall test scores are defined

as composite learning scores, i.e. test scores that are constructed by combining standardized test scores from at least two different subjects. Depending on the intervention, the exact definition of composite learning scores varies. In particular, composite learning scores might or might not draw on native language and math scores used in one of the other two meta-analyses.

When it comes to overall learning levels, the meta-analysis visualized in Figure 4 draws on 27 interventions. Nine of these interventions fall into the “teachers” category, 13 belong to the “schools/school administrations” group, and four are centered on households. Only one intervention falls into the “community” category. This renders a meta-analysis of community-focused interventions on overall or composite learning outcomes futile.

Figure 4 demonstrates that all three groups of interventions for which a meta-analysis of impacts on children’s average overall test scores is feasible have a statistically significant impact on this outcome variable. What is more, for two out of three groups the impacts are rather substantial: Interventions focusing on teachers raise average overall test scores by 0.17 standard deviations and interventions targeting schools/school administrations increase them by 0.18 standard deviations. In contrast, interventions directly incentivizing households raise composite test scores by only 0.04 standard deviations. This is a relatively feeble effect but one that is still statistically significant.

Once more, the provision of supplementary remedial teaching by community volunteers (Lakshminarayana et al., 2013) and the establishment of publicly funded private primary schools (Barrera-Osorio et al., 2013) are among the most promising interventions. The first intervention raises average overall test scores by 0.75 standard deviations and the second one increases them by 0.67 standard deviations. Together with these teacher-centric interventions, two interventions that focus on schools or school administrations are also very effective. These are the placement of a school in a village for the first time (Burde and Linden, 2013) and the threatening of private schools with the withdrawal of public funds (Barrera-Osorio and Raju, 2010). At effects on overall learning scores of 0.66 standard deviations each, both interventions appear to be potent mechanisms for raising average overall test scores.

The main findings from the three meta-analyses of interventions’ impacts on native-language, mathematics and overall learning levels are summarized in Table C. In the table, interventions are again grouped by the four actors and the table also gives examples of especially promising interventions. What is more, Table C also facilitates a discussion of whether clear messages can be discerned for interventions’ impacts on learning outcomes if interventions are grouped not by the four actors but by the more fundamental categories introduced in the conceptual framework of section 2. In particular, the table allows an assessment of whether interventions addressing education supply (teachers and schools) or demand (households and communities) appear more successful in raising education outcomes and if programs targeting target individuals (teachers and households) or those addressing collectives (schools and communities) have shown greater promise.

On the first of these questions, one of the main results of the meta-analyses indeed is that supply-side interventions have the potential to induce moderate to important improvements in learning outcomes.

As Table C makes clear, teacher-centric interventions appear a bit more effective than school-focused programs in the case of mathematics skills. But apart from this qualification, both groups of supply-side interventions hold promise for improving learning levels irrespective of whether these learning levels are measured in terms of native language, mathematics or overall test scores. In contrast, demand-side interventions seem to be less effective in improving learning outcomes. In particular, meta-analyses of demand-side interventions that directly target children, households or families reveal that these have only weak impacts on native language, mathematics or overall test scores. This result confirms the more verbal discussion in the last section about how interventions targeting households or families appear to be quite appropriate for increasing school enrollment but much less suitable for improving learning outcomes. While more research is needed on interventions that aim to involve whole communities in holding education service providers accountable, the (relatively scarce) evidence that is already available is at least somewhat encouraging. In the case of language and math test scores, the meta-analyses summarized in Table C reveal impacts of community-focused interventions that are stronger than those found for interventions targeting households and in the same order of magnitude as those documented for either group of supply-side interventions. For composite test scores, a meta-analysis of community-centric interventions is not feasible at this point.

**Table C – Summary of Results of the Meta-Analyses**

Actor	Effects			Examples of especially promising interventions
	Language	Math	Composite	
<b>A. Teachers</b>	+	++	++	Supplementary remedial teaching by community volunteers
<b>B. Schools</b>	+	+	++	publicly funded private primary schools; revised English education curriculum
<b>C. Households</b>	(+)	(+)	(+)	<i>None</i>
<b>D. Community</b>	+	+		school management structures and training/support services for community members

**Notes:** o means no significant effect; (+) effect smaller than 0.05 s.d.; + effect between 0.05 and 0.15 s.d.; and ++ effect between 0.15 and 0.25 s.d.. A meta-analysis of community-focused interventions on overall learning outcomes is impossible because only one intervention falls into this category.

**Source:** World Bank staff calculations based on studies listed in Table A and Figures 3 to 5.

Instead of contrasting supply- and demand-side interventions, Table C can also be interpreted through the lens of whether programs targeting individuals or a collective appear more promising in promoting student learning. Partly because of the scarcity of rigorous evidence for the learning impacts of community-centric interventions, results from this exercise are less clear-cut than those contrasting supply- and demand-side interventions. This is especially the case for interventions' impacts on math test scores since teacher-centered interventions seem to be very effective in improving mathematics learning while this is not the case for household-focused interventions and interventions targeting schools and communities promise moderate impacts. Overall, though, the summary of different meta-analyses suggests that programs centered on collective bodies are more effective in raising learning outcomes than those targeting individuals.

The third way to categorize interventions that was introduced in section 2 relied on whether a specific intervention emphasized the provision of additional resources or aimed to alter individual or collective incentives. As outlined above, there is no direct correspondence between this third categorization and the four actors (teachers, schools, households and communities). For each of the actors, a particular intervention might provide resources or alter incentives. Of course, this does not render meta-analyses of native language, mathematics or overall learning outcomes with interventions grouped according to whether they primarily address inputs or incentives unfeasible. However, such meta-analyses fail to find any statistically significant differences in learning effects between the two groups of interventions.<sup>11</sup>

### 3.3. Discussion

As already mentioned in the introduction, the last ten years have seen a dramatic rise in the number of rigorous impact evaluations of education-related interventions in developing countries. While in South Asia we have the unique benefit of having relatively plentiful and high-quality impact evaluations, our paper is not the first to combine a systematic literature review with a rigorous meta-analysis of the available evidence to investigate what kind of interventions are most effective in improving education outcomes. In fact, as also already spelled out in the introduction, four methodologically similar reviews exist. One of these four reviews (by Petrosino et al., 2012) centers primarily on school enrollment and attendance. The other three (by Krishnaratne, White and Carpenter, 2013; Conn, 2014; and McEwan, 2014) are mainly concerned with students' learning outcomes. Another differentiating factor among the four reviews is that Petrosino et al. (2012), Krishnaratne, White and Carpenter, (2013) and McEwan (2014) consider evidence from all over the developing world while Conn (2014) concentrates on Sub-Saharan Africa, a region that shares many similarities with South Asia but also exhibits a number of important differences.

Reassuringly, the results of our systematic review and meta-analysis are generally very consistent with the findings from the four other methodologically comparable works. Nevertheless, there are instances where education-related interventions have apparently had different impacts in South Asia and other parts of the developing world.

The systematic review and meta-analysis of Petrosino et al. (2012) includes the universe of studies that (i) assess the impact of an intervention on primary or secondary school outcomes, (ii) use an experimental or quasi-experimental approach, (iii) are conducted in a developing country, (iv) include at least one quantifiable education-related outcome measure (like enrollment, attendance, dropout, or progression), (v) are published or made available before December 2009, (vi) and include data from

---

<sup>11</sup> Appendix C contains the forest plots of three meta-analyses of interventions' impacts on composite test scores where interventions are not grouped by actor but by whether they primarily target individuals or a collective, address education supply or demand, or provide resources or incentives. In addition to that, the appendix also shows outputs of meta-analyses that rely only on impact estimates derived from RCTs or group studies by whether or not at least one author was affiliated with the World Bank at the time of the study's publication.

1990 or beyond. Interventions are categorized into five groups: (i) economic interventions which include conditional and unconditional cash and/or food transfers, vouchers for private schools, microfinance opportunities, school fee reductions or eliminations, school scholarships or fellowships, and material support such as school uniforms; (ii) educational programs and practices providing services or materials to students, providing extra teachers, reducing class sizes, providing incentives to teachers, monitoring teacher attendance, training teachers, empowering and funding parent school associations, or covering more comprehensive school reform and improvement efforts; (iii) interventions related to health care and nutrition; (iv) building programs and infrastructure improvements; and (v) interventions providing information or training to students, parents or whole communities.

In their meta-analysis, Petrosino et al. (2012) mainly concentrate on two outcome variables, enrollment and attendance. They find that the largest impacts on these two variables are achieved by interventions entailing the construction of new schools or the improvement of school infrastructure. Relatively large enrollment and attendance effects can also be documented for interventions related to health care and nutrition, and economic practices and programs. In contrast, impacts are generally found to be small for those interventions that provide information or training or for those that reform educational practices or programs. Overall, the results of the systematic review and meta-analysis by Petrosino et al. (2012) are very consistent with our findings for South Asia. In particular, evidence for South Asia discussed above confirms that the establishment of new schools can dramatically increase enrollment levels. Besides, we also conclude for South Asia that at least some of the interventions categorized by Petrosino et al. (2012) as belonging to the group of economic practices and programs – like CCT programs – have a demonstrated ability to significantly increase school enrollment and attendance.

The main objective of the systematic review and meta-analysis by Krishnaratne, White and Carpenter (2013) is to find out whether education-related interventions are effective in improving learning outcomes in developing countries. The respective meta-analysis is based on 75 experimental or quasi-experimental studies from all over the developing world. Its overall conclusion is that education-related interventions *do* work. When it comes to a more thorough analysis of this result, “while this review provides a relatively detailed narrative of the variation in program effects within the body of the paper, the statements on ‘effective’ or ‘ineffective’ interventions in the conclusions section are based on relatively little data” (Conn, 2014, p. 10). Moreover, the focus of Krishnaratne, White and Carpenter (2013) is clearly on a more verbal discussion of different interventions’ impacts on the one hand and the meta-analysis of education-related interventions’ overall effect on the other hand. In contrast, a rigorous meta-analysis that compares the relative impacts on learning levels of different types of education-related interventions is only touched upon. This makes it difficult to compare our detailed findings for different groups of interventions with those of Krishnaratne, White and Carpenter (2013).

Many more details on the relative effectiveness of different types of interventions in improving learning outcomes are provided by the systematic review and meta-analysis by McEwan (2014). This review and meta-analysis is based on 77 interventions that fulfill the following criteria: (i) an

intervention has to be implemented in a developing country and conducted in grades one to eight or ages six to 14; (ii) children or groups of children have to be randomly assigned to a treatment or a business-as-usual control group (quasi-experiments are not covered); (iii) results have to be reported for at least one continuously measured learning outcome variable in language, reading, mathematics, or a composite score; and (iv) sufficient data has to be reported so that a treatment's effect size and standard error in the full experimental sample can be calculated. The 77 interventions that fulfill these criteria are grouped according to three categories and eleven subcategories. The categories differentiate between (i) instructional interventions; (ii) interventions related to health or nutrition; and (iii) interventions aiming at altering individuals' incentives. Instructional interventions are further divided according to whether a program relates to computers or technology, teacher training, class size or composition, instructional materials or monetary grants. Sub-categories for interventions related to health or nutrition administer the intake of food, beverages, and/or micronutrients, and deworming drugs. Finally, interventions that aim at altering individuals' incentives rely either on contract or volunteer teachers, student or teacher performance incentives, innovations in school management or supervision, or on informational treatments.

McEwan (2014) finds that, on average, monetary grants and deworming treatments have insignificant effects on learning. Nutritional treatments, treatments that disseminate information, and treatments that improve school management or supervision have small but generally significant learning effects. The largest mean effects on learning are achieved with computers or instructional technology, teacher training, smaller classes, smaller learning groups within classes, or the grouping of students according to their ability. Contract or volunteer teachers, student and teacher performance incentives, and the introduction of new instructional materials are similarly found to be very effective in improving students' learning levels. While our meta-analysis does not include health-centric interventions and the evidence for South Asia is scarce or even nonexistent for some of the subcategories defined by McEwan (2014), for subcategories where sufficient evidence for South Asia exists his results are generally very much in line with ours. For instance, our meta-analysis confirms that making use of computers or instructional technology is a promising way for improving learning outcomes. Our meta-analysis is also cautious about the impacts on learning of interventions that rely solely on monetary grants or the dissemination of information. One area where conditions in South Asia might be different than in other parts of the developing world is the effectiveness of reforming school management or supervision. While evidence on this topic is relatively scarce for South Asia, our literature review and meta-analysis indicates a larger impact on learning outcomes of these types of interventions than does McEwan's (2014) meta-analysis.

Recognizing the importance of region-specific contexts and conditions, Conn (2014) aims to identify education-related interventions that are effective in improving students' learning outcomes specifically in Sub-Saharan Africa. Similarly as in South Asia, over the last decade much progress has been made in improving school access in large parts of Sub-Saharan Africa but learning levels generally remain very low. Against this backdrop, Conn's (2014) systematic review and meta-analysis covers all experimental and quasi-experimental evidence on education-related interventions in Sub-Saharan Africa that has been published in peer-reviewed journals, academic working papers or reports



published through academic institution or research organization and fulfills a number of additional criteria. In particular, to be included in her systematic review and meta-analysis, impact evaluations must come from the fields of economics, education or public health and need to compare a treatment to a business-as-usual control group. The resulting systematic review and meta-analysis contains 66 separate experiments, 83 treatment arms, and 420 effect size estimates and uses considerations about what channels could be considered binding constraints to learning to divide interventions into the following five groups and twelve subgroups: (i) quality of instruction (comprising interventions related to class size and composition, instructional time, pedagogy, and the provision of school supplies); (ii) student or community financial limitations (including the abolishment of school fees, cash transfers, and improvements in school infrastructure); (iii) school or system accountability (such interventions provide information to improve accountability, or promote school-based management or decentralization); (iv) student cognitive processing ability (treatments in this category encompass the provision mid-day meals; and interventions addressing students' health); (v) and student or teacher motivation (comprising programs targeting students' incentives; or teachers' incentives).

Using meta-regressions, Conn (2014) finds that for Sub-Saharan Africa by far the most promising subgroup of interventions is the one focusing on reforming pedagogical methods. More specifically, her meta-analysis demonstrates that interventions that employ adaptive instruction and teacher coaching techniques are particularly effective. In contrast, interventions that provide health treatments or mid-day meals have low or negligible impacts on learning outcomes. Generally speaking, Conn's (2014) findings for sub-Saharan Africa are consistent with the conclusions of our systematic review and meta-analysis for South Asia. These also documents large impacts on student learning of innovative changes in the curriculum. Besides, they also demonstrate the lack of significant impacts on student learning of mid-day meal programs (which are however found to exhibit relatively large enrollment and attendance effects). At the same, while pedagogical innovations are among the most promising type of education-related interventions that have been subject to rigorous impact evaluations in South Asia, they do not stand out as much in South Asia as they apparently do in Sub-Saharan Africa. In South Asia, a number of other types of interventions – e.g. those related to teachers' contracts or incentives, or the construction of new schools – appear to be about equally effective.

#### **4. Conclusions**

Promoting quality education and learning for all is a vital objective in South Asia both for individuals and for countries as a whole, particularly at this point in the region's development trajectory. Recognizing this importance and considering the second Millennium Development Goal of providing universal primary education for all, South Asian countries have made genuine, impressive progress to improve education access for their people. South Asia's net primary enrollment rate stood at 75 percent in 2000 and had risen to 89 percent by 2010. Concurrently, the number of out-of-school children between the ages of eight and 14 years fell from 35 million in 1999 to 13 million in 2010. Progress has been made throughout the region: Sri Lanka and the Maldives have consistently enrolled almost all their children in primary schools; Bhutan and India have recently increased enrollment rates

steadily to about 90 percent of children aged six to 14 years; and in Pakistan the primary net enrollment rate jumped from 58 percent to 74 percent between 2000 and 2011 (though that is still lower than the regional average). In addition, there has been significant movement towards gender parity as well as some success in drawing the more marginalized into school.

While there has been progress to improve the access to education in the region, access for all remains elusive, particularly in getting the most disadvantaged and marginalized into school. Still, the key challenging policy agenda now is on enhancing the quality of education and making progress towards improving learning outcomes – the ultimate goal of any educational system and a stronger driver of economic growth than years of schooling. A recent report on challenges, opportunities, and policy priorities in school education in South Asia, established that mean student achievements in mathematics, reading, and language are very low throughout the region, except maybe for Sri Lanka. Mean student achievements in arithmetic tend to be particularly low but mean student achievement in reading and language is generally low as well. Not surprisingly, competency in English is usually even more depressed than in native languages. Moreover, within countries mean levels tend to be low but variances high. Thus, a small proportion of students can meet international benchmarks while the rest perform very poorly (Dundar et al., 2014).

At the same time, across the region, there is growing recognition of the need to improve the quality of education so that people can enjoy broader life opportunities, become more productive and lift themselves out of poverty or out of risk of falling into poverty. Furthermore, there are continued efforts to get the most disadvantaged and hard to reach young people into school. However, while awareness of these challenges is a vital precondition to address them, a large and unresolved policy question is how to improve education quality and get all South Asian children into school. This paper has argued that an answer to this policy question might be found with the help of a different important trend that has been affecting our capacity to address development challenges: development researchers have increasingly used rigorous impact evaluations to analyze innovations that seek to improve development outcomes both in education and other fields. Using innovative analytic techniques that fall under the broad category of experimental and quasi-experimental designs, impact evaluations have made it possible to make a lot of progress on analyzing whether innovations really make the differences they intend to make. In particular in the education sector, South Asia has been at the forefront of this movement and by now a sufficient body of evidence exists to draw stringent conclusions with regard to the impacts of a large number and variety of policy innovations.

The review and meta-analysis of the evidence from rigorous impact evaluations can serve as a basis for recommendations for future impact evaluations. While such recommendations could potentially address a wide range of issues, three topics stand out. These are the impact evaluations' reporting, their focus on mechanisms in addition to impacts and the adequate time horizon between an intervention and the measurement of its impact.

Learning from impact evaluations in general and the compilation of reviews and meta-analyses in particular would benefit greatly from clearer and more consistent reporting of impact evaluations' design, methodology and results. At times, the precise intervention is not clearly described, reasons

behind the choice of the exact sample are not explicitly discussed and/or results are presented in a way that makes it hard to compare them with those of other studies. Often, very simple and straightforward fixes would make it much easier to place a specific result in the wider literature. For example, it would help enormously if exact sample sizes and baseline summary statistics were always reported (if average class size is not known and an intervention's effect on enrollment is only indicated by the additional number of students induced to enroll in treatment classes, it is impossible to make a meaningful comparison between the intervention's impact to that of other policies). Going one step further, if the reporting standards set forth in the CONSORT Statement – an evidence-based set of recommendations for reporting the approach, implementation and results of RCTs, cf. Moher et al. (2010) – were more widely adopted, this would further facilitate the critical appraisal and interpretation of the findings from impact evaluations.

Besides, a large proportion of impact evaluations are very much focused on establishing whether there is a causal relationship between a certain intervention and one or more education-related outcome variables. But while the attribution of a causal impact to a specific intervention is the defining feature of any impact evaluation, much more could be learned if more attention was devoted to the mechanisms that purportedly relate an intervention and the outcomes measure(s). At present, a clear results chain connecting inputs, activities, outputs, outcomes and impacts is only rarely explicitly mapped. Statistical tests of a result chain's individual links are even scarcer. Sometimes it is uncertain whether an intervention was even properly implemented. Especially in those cases where an impact evaluation fails to find any statistically significant impacts of an intervention, a further emphasis on implementation, outputs and intermediate outcomes could shed light on the reasons behind an intervention's apparent failure and could offer guidance for how to design more promising interventions in the future.<sup>12</sup>

Finally, most impact evaluations focus on relatively short time horizons. The median duration between an intervention and the measurement of its impact in the impact evaluations underlying this paper's systematic review and meta-analysis is twelve months. The reasons behind the usually very short time horizons are clear: Every further round of follow-up surveys and measurements is very costly, political relevance often demands the timely reporting of impacts and some impact evaluation designs – e.g. an RCT of an intervention that is implemented in stages over a whole country or region – even preclude the measurement of long-term impacts. Still, many promising interventions in the education sector might only become effective over the longer term. Others, in contrast, might only be short-term fixes that quickly lose their impact. In a third group of interventions, important differences between seemingly equally effective treatment arms might only become apparent over the longer term – as was the case with the impacts on learning of group- and individual-based bonus payments to teachers in Andhra Pradesh, India (Muralidharan and Sundararaman, 2011, and Muralidharan 2012). All this means that the designers of impact evaluations should make every effort to capture interventions' effects over a period spanning at least two to four years.

---

<sup>12</sup> See IEG (2012) for a how-to guide for designing a results framework.

Among the policy-related lessons that emerge from a rigorous meta-analysis of the available evidence is that demand-side intervention targeting either individual households or whole communities can be very effective in raising school enrollment and attendance. However, interventions targeting teachers or schools and thus the supply-side of the education sector are generally much more adept at improving learning outcomes. What is more, both interventions that provide different actors with resources and those that incentivize them to change their behaviors show moderate but statistically significant impacts on student learning. Thus, there is no basis for an outright dismissal of primarily input-oriented interventions. Instead, a mix of input- and incentive-oriented interventions tailored to the specific conditions on the ground might appear most promising for fostering learning outcomes in South Asia.

Slightly less rigorous but somewhat more tangible lessons can be learned from a systematic review of the existing evidence on education-related interventions in South Asia. Concerning teachers, such a review shows that it appears highly dubious whether there is a relationship between standard teacher variables (like a teacher's formal education and experience) and learning outcomes. In fact, contract teachers appear to be at least as effective as regular teachers even though they usually make much lower salaries, have less formal qualifications and enjoy only reduced job security. The hiring of other non-traditional types of teachers or even the engagement of volunteer teachers recruited from local communities appears promising as well. Positive and cost-effective impacts on students' learning outcomes might also be achieved by monitoring teachers' performance and linking part of their pay to performance. Bonus payments based on individual teachers' performance appear particularly promising.

With regard to schools, there are rigorous impact evaluations that show how the provision of block grants to schools or investments in schools' physical infrastructure can increase enrollment levels. However, such interventions tend to be relatively costly but to have no or only very small impacts on learning outcomes. In contrast, positive learning effects can be achieved at very moderate costs by expanding and updating the curriculum, by providing remedying education and by making better use of information technology if this is done with due consideration of local conditions and as a complement rather than a substitute of regular lessons. Placing a school in a village – either for the first time or in addition to existing schools – can have positive effects on learning and enrollment as well. Indeed, school building programs rank among the most effective (if not necessarily cost-effective) education interventions in South Asia.

There also exist a fair amount of impact evaluations that analyze interventions directly targeting the demand side (either households or families, or whole communities). In this context, strong evidence exists with regard to the impacts of interventions that provide households with free tuition, cash transfers or goods meant to reduce the cost of schooling. These kinds of interventions are very effective in increasing school enrollment and attendance but their impact on learning outcomes tends to be limited at best. Similarly, only very modest positive learning effects seem to be achieved by interventions that focus specifically on mothers. In contrast, interventions that introduce incentives to learn, increase school choice and promote non-state education providers appear very promising

both with regard to their impacts on enrollment and learning and their cost-effectiveness. However, the evidence on these kinds of interventions' impacts on education outcomes is still rather scarce.

Finally, interventions that target not individual households but aim to involve whole communities in holding schools accountable appear moderately promising. However, limiting an intervention to the provision of information might reduce costs but not be sufficient for eliciting significant improvements in the quality of education. Instead, comprehensive interventions combining different aspects of citizen's engagement and demand side incentives appear more promising as at least some of the interventions in this group that have been analyzed by rigorous impact evaluations have demonstrated moderate to strong positive effects on learning and other education outcome variables.

These findings from rigorous impact evaluations provide guidance on policy options that might be worth pursuing in order to improve access or quality of education. However, even if the review and meta-analysis shows that a specific group of interventions has positive effects on education outcome variables, these interventions should not be seen as one-size-fits-all solutions that work under each and every circumstance. Instead, it is important to keep a few caveats in mind. In particular, evidence on many important topics is still relatively scarce – e.g. on incentives to learn, increases in school choice and the promotion of non-state education – and even for groups of interventions where more evidence is available, this evidence does not always point in the same direction. While these are exactly the reasons for focusing on South Asia as a distinct region of the world and for doing a rigorous meta-analysis – which as discussed above is basically a statistical tool to combine results from two or more individual interventions – the current state of knowledge allows us to perform meaningful meta-analyses only on rather high level of aggregation. This means the meta-analysis allows us to draw firm conclusions related more to rather abstract concepts (education supply and demand, or teachers, schools, households and communities) than to specific interventions (such as for instance the engagement of volunteer teachers recruited from local communities or incentives to learn for children or their parents).

Besides, impact evaluations tend to analyze very specific interventions under tightly controlled conditions and frequently for a relatively small or pilot sample. This has led to criticisms about whether results from impact evaluations are generalizable. While this kind of critique does not only apply to impact evaluations but to empirical research more generally, the issue of external validity certainly is critical and every planned policy intervention that is being justified through results from relevant impact evaluations will have to answer questions such as: Are the conditions on the ground similar to the ones in the setting where the impact evaluation(s) was/were conducted? Will costs or effects of the intervention change when it is scaled up; i.e., are there positive or negative returns to scale? Will there be general equilibrium effects; i.e. will the scaling up of the policy lead to spillovers or changes in behavior that did not happen when it was piloted on a much smaller scale? Only a thorough deliberation of these kinds of questions will assure that a scaled-up version of a successful pilot program or even the replication of an intervention in a different context will have similarly positive impacts on education outcome variables.

Another issue that has to be considered when one uses a review of existing impact evaluations on education-related interventions to draw policy conclusions relates to the political economy. Even if rigorous evidence exists that convincingly shows a certain type of interventions has positive impacts on specific education-related outcome variables, it might still not be worth putting a lot of effort in promoting this type of interventions if it is clear at the outset that politically influential actors will strongly oppose it. At the very least, one should be conscious about political realities on the ground and address or incorporate these into any policy recommendations. For instance, while rigorous and robust evidence exist on the benefits of supplementing traditional teacher contracts by more flexible and performance-based contracts, such reforms tend to face strong opposition, in particular from politically well-connected teachers unions. Recognizing these political economy realities, Pritchett and Murgai (2007) and Muralidharan (2013) suggest integrating contract and regular teachers into a career ladder with bonuses, pay raises, and promotion to regular civil-service rank based on performance over time.

These considerations illustrate that evidence from impact evaluations should not be blindly trusted. Rather, it should be seen as one – important – building block of informed decision making. At the same time, one should not overlook the huge potential offered by the relatively plentiful and high-quality education-related impact evaluations that have been conducted in South Asia. During the last ten years, these impact evaluations have created more knowledge on the costs and effects of a diverse set of often innovative interventions than was ever available before. Moreover, South Asia is well ahead of other regions in terms of the number of education-related impact evaluations and the average quality of these impact evaluations is quite impressive with most of them paying close attention to common methodological pitfalls and biases. Today, the region enjoys a unique opportunity to benefit from the knowledge that has been created over the last ten years and to harness this knowledge for evidence-based policy making. If used appropriately, the rigorous impact evaluations of education-related interventions can help South Asian economies in mastering the formidable policy challenges they face in terms of strengthening their education systems. In particular, they can provide a sound basis for reforms that improve the quality of education and provide learning that reaches even the most disadvantaged groups.

## References

- Abeberese, A., T. Kumler and L. Linden. 2014. "Improving reading skills by encouraging children to read: A randomized evaluation of the Sa Aklat Siskat reading program in the Philippines." *The Journal of Human Resources* 49: 611–633.
- Afridi, F. 2010. "The Impact of School Meals on School Participation: Evidence from Rural India." Discussion Paper 10-02. Indian Statistical Institute, Delhi.
- Andrabi, T., J. Das and A.I. Khwaja 2013. "Report Cards: The Impact of Providing Schools and Child Test Scores on Educational Markets" World Bank.
- Andrabi, T., J. Das, A.I. Khwaja, T. Vishwanath and T. Zajonc. 2007. "Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate." World Bank.
- Angrist, J., E. Bettinger, E. Bloom, E. King and M. Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92(5): 1535-1558
- Aslam, M. and G. Kingdon. 2011. "Evaluating Public Per-Student Subsidies to Low-Cost Private Schools - Regression-Discontinuity Evidence from Pakistan." *Economics of Education Review* 30: 559–574
- Atherton, P. and G. Kingdon. 2010. "The Relative Effectiveness and Costs of Contract and Regular Teacher in India." CSAE Working Paper no 2010\_15.
- Aturupane, H., P. Glewwe, T. Keeleghan, R. Ravina, U. Sonnadara, and S. Wisniewski. 2013. "An Impact Evaluation of Sri Lanka's Policies to Improve the Performance of Schools and Primary School Students through its School Improvement and School Report Card Programs", World Bank.
- Baker, J. 2000. "Evaluating the Impact of Development Projects on Poverty – A Handbook for Practitioners." *Directions in Development*. The World Bank.
- Banerjee, A., R. Banerji, E. Duflo, R. Glennerster and S. Khemani, 2010a. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy* 2(1).
- Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston and N. Singh, 2012. "Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police." NBER Working Paper 17912.
- Banerjee, A. Cole, D., Duflo, E. and Linden, L. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122, no. 3: 1235–1264.

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

Banerjee, A., E. Duflo, R. Glennerster and D. Kothari. 2010b. "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives." *British Medical Journal* 2010; 340:c2220.

Banerjee, R. Berry, J. Shotland, M. 2013. "The Impact of Mother Literacy and Participation Programs on Child Learning: Evidence from a Randomized Evaluation in India." J-PAL.

Barrera-Osorio, F. and L. Linden. 2009. "The Use and Misuse of Computers in Education – Evidence from a Randomized Experiment in Colombia", World Bank Policy Research Paper 5465.

Barrera-Osorio, F. and D. Raju. 2010. "Short-run Learning Dynamics under a Test-based Accountability System", World Bank Policy Research Paper 4836.

Barrera-Osorio, F. and D. Raju. 2011. "Evaluating Public Per-Student Subsidies to Low-Cost Private Schools – Regression-Discontinuity Evidence from Pakistan." World Bank Policy Research Paper 5638.

Barrera-Osorio, F. and D. Raju. 2013. "The impacts of differential changes in benefit levels on female enrollment: Evidence from a gender-targeted CCT program in Pakistan." Unpublished manuscript.

Barrera-Osorio, Blakeslee, Hoover, Linden, Raju and Ryan. 2013. "Leveraging the Private Sector to Improve Primary School Enrolment: Evidence from a Randomized Controlled Trial in Pakistan." Unpublished manuscript.

Behrman, J., S. Parker and P. Todd. 2011. "Medium-term impacts of the Oportunidades conditional cash transfer program on rural youth in Mexico." In: S. Klasen and F. Nowak-Lehmann, eds. *Poverty, Inequality, and Policy in Latin America*, Cambridge, MA: MIT Press, 219-270.

Benjamin, L., A. Tarozzi, A. Pfaff, S. Balasubramanya, K. Matin Ahmed and A. van Geen. 2013. "Impact of a Randomized Controlled Trial in Arsenic Risk Communication on Household Water-source Choices in Bangladesh." *Journal of Environmental Economics and Management* 65(2): 225-240.

Berry, J. 2014. "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India." Cornell University.

Berry, J. and L. Linden. 2009. "Bridge Classes and Peer Networks among Out-of-school Children in India." MIT.

Bobonis, G., E. Miguel and C. Puri-Sharma 2006. "Anemia and School Participation." *Journal of Human Resources* 41, 4: 692-721.



- Borkum E, He F, and Linden LL. 2012. “The Effects of School Libraries on Language Skills: Evidence from a Random-ized Controlled Trial in India.” NBER Working Paper 18183.
- Burde, D, and Linden LL. 2013. “Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools.” *American Economic Journal: Applied Economics* 5, no. 3: 27–40. doi:10.1257/app.5.3.27.
- Cabezas, V., J. Cuesta and F. Gallego. 2011. “Effects of short-term tutoring on cognitive and non-cognitive skills: evidence from a randomized evaluation in Chile.” Unpublished manuscript.
- Carillo, P., M. Onofa and J. Ponce. 2010. “Information technology and student achievement: evidence from a randomized experiment in Ecuador.” Working Paper No. 223. Inter-American Development Bank.
- Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. Halsey Rogers. 2006. “Missing in Action: Teacher and Health Worker Absence in Developing Countries.” *Journal of Economic Perspectives* 20: 91-116.
- Chaudhury, N. and D. Parajuli. 2010a. “Conditional cash transfers and female schooling: the impact of the female school stipend programme on public school enrolments in Punjab, Pakistan.” *Applied Economics* 42: 3565–3583.
- Chaudhury, N. and D. Parajuli. 2010b. “Giving it Back: Evaluating the Impact of Devolution of School Management to Communities in Nepal.” Unpublished manuscript.
- Conn, K. 2014. “Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations. Dissertation. Columbia University.
- Cristia, J., P. Ibararán, S. Cueto, A. Santiago and E. Severín. 2012. “Technology and child development: evidence from the One Laptop per Child program.” Working Paper No. 304. Inter-American Development Bank
- Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman. 2013. “School Inputs, Household Substitution, and Test Scores”, *American Economic Journal: Applied Economics* 5:2.
- Dhaliwal, I., E. Duflo, R. Glennerster and C. Tulloch. 2012. “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education”. J-PAL.
- Dundar, H., T. Beteille, M. Riboud and A. Deolalikar. 2014. “Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities.” World Bank.
- Duflo, E., P. Dupas and M. Kremer. 2011. “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.” *American Economic Review* 101(5): 1739-74.

- Duflo, E., R. Hanna and S. Ryan. 2012. “Incentives Work: Getting Teachers to Come to School”. *American Economic Review* 102(4): 1241-78.
- Egger, M., G. Davey Smith and A. N. Phillips. 1997. “Meta-Analysis: Principles and Procedures.” *British Medical Journal* 315:1533.
- Egger, M., G. Davey Smith, M. Schneider and C. Minder. 1997. “Bias in Meta-Analysis Detected by a Simple, Graphical Test.” *British Medical Journal* 315:629–634.
- Evans, D., M. Kremer and M. Ngatia. 2009. “The impact of distributing school uniforms on children’s education in Kenya.” Unpublished manuscript.
- George, C.M., A. van Geen, V. Slavkovich, A. Singha, D. Levy, T. Islam, K. Matin Ahmed, J. Moon-Howard, A. Tarozzi, X. Liu, P. Factor-Litvak and J. Graziano. 2012. “A Cluster-based Randomized Controlled Trial Promoting Community Participation in Arsenic Mitigation Efforts in Bangladesh.” *Environmental Health* 11:41.
- Gertler, P., H. Patrinos and M. Rodríguez-Oreggia. 2012. “Parental Empowerment in Mexico: Randomized Experiment of the 'Apoyos a la Gestion Escolar (AGE)' in Rural Primary Schools in Mexico.” Unpublished manuscript.
- Glewwe, P., E. Hanushek, S. Humpage and R. Ravina, 2011. “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010.” National Bureau of Economic Research Working Paper 17554.
- Glewwe, P., N. Ilias and M. Kremer. 2010. “Teacher Incentives.” *American Economic Journal: Applied Economics* 2: 205–227.
- Glewwe, P., A. Park and M. Zhao. 2012. “The impact of eyeglasses on the academic performance of primary school students: Evidence from a randomized trial in rural China.” Working Paper 12-2. Center for International Food and Agricultural Policy.
- Goyal, S. and P. Pandey. 2009. “Contract Teachers.” South Asia Human Development Sector Report No. 28. World Bank.
- Greenhalgh, T. and R. Peacock. 2005. “Effectiveness and Efficiency of Search Methods in Systematic Reviews of Complex Evidence: Audit of Primary Sources.” *British Medical Journal* 331: 1064-1065.
- Hanna, R. N. and L. L. Linden, 2012. “Discrimination in Grading.” *American Economic Journal: Economic Policy* 4. 4: 146-68.
- Hanushek, E. 2013. “Economic growth in developing countries: The role of human capital.” *Economics of Education Review* 37: 204-212.
- Hanushek, E. and L. Woessmann. 2012. “Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation.” *Journal of Economic Growth* 17: 267-321.

- Hasan, A. 2010. "Gender-targeted Conditional Cash Transfers – Enrollment, Spillover Effects and Instructional Quality." Policy Research Working Paper 5257. World Bank.
- He, F., L. Linden and M. MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Unpublished Manuscript.
- He, F., L. Linden and M. MacLeod. 2009. "A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial." Unpublished Manuscript.
- Hong, S.Y. and Sarr, L.R. 2009. "Long-term Impacts of the Free Tuition and Female Stipend Programs on Education Attainment, Age of Marriage, and Married Women's Labor Market Participation in Bangladesh." World Bank.
- IEG (Independent Evaluation Group). 2011. "Do Conditional Cash Transfers Lead to Medium-Term Impacts?: Evidence from a Female School Stipend Program in Pakistan." World Bank.
- IEG (Independent Evaluation Group). 2012. "Designing a Results Framework for Achieving Results: A How-To Guide." World Bank.
- Jalali, S. and C. Wohlin. 2012. "Systematic Literature Studies: Database Searches vs. Backward Snowballing." International Conference on Empirical Software Engineering and Measurement, September 19-20, 2012, Lund, Sweden.
- Jayaraman, J. and D. Simroth. 2011. "The Impact of School Lunches on Primary School Enrollment: Evidence from India's Midday Meal Scheme." CESifo Working Paper Series 3679, CESifo Group Munich.
- Jueni, P., F. Holenstein, J. A. C. Sterne, C. Bartlett and M. Egger. 2002. "Direction and Impact of Language Bias in Meta-Analysis of Controlled Trials: Empirical Study." *International Journal of Epidemiology* 31: 115–123.
- Kim, J., H. Alderman and P. Orazem. 1999. "Can Private School Subsidies Increase Enrollment for the Poor? The Quetta Urban Fellowship Program." *World Bank Economic Review* 13(3): 443-465.
- Kingdon, G. and F. Teal. 2010. "Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach." *Journal of Development Economics* 91: 278–288.
- Kremer, M., C. Brannen and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340, 297-299.
- Kremer, M., S. Moulin and R. Namunyu. 2003. "Decentralization: a cautionary tale". *Poverty Action Lab Paper No. 10*.
- Kremer, M., E. Miguel and R. Thornton, 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91, 3: 437-456.

- Krishnaratne, S., H. White and E. Carpenter. 2013. "Quality Education for all Children? What Works in Education in Developing Countries." International Initiative for Impact Evaluation (3ie) Working Paper 20.
- Lakshminarayana, R., A. Eble, P. Bhakta, C. Frost, P. Boone, D. Elbourne and V. Mann. 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, no. 7.
- Linden, L. L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." World Bank Infodev Working Paper (17).
- Linden, L. L. and G. K. Shastry. 2012. "Grain Inflation: Identifying Agent Discretion in Response to a Conditional School Nutrition Program." Journal of Development Economics 99, 1: 128-138.
- McEwan, P. 2014. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments". Wellesley College.
- Miguel, E. and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica 72, 1: 159-217.
- Mo, D., H. Zhang, L. Yi, R. Luo, S. Rozelle and C. Brinton. 2011. "School dropouts and conditional cash transfers: Evidence from a randomized controlled trial in rural China's junior high schools." Discussion Paper 283/2011. Katholieke Universiteit Leuven.
- Moher, D., S. Hopewell, K. Schulz, V. Montori, P. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, D. Altman. 2010. "CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials." British Medical Journal 2010; 340:c869.
- Muralidharan, K. 2012. "Long-term Effects of Teacher Performance Pay: Experimental Evidence from India." Society For Research On Educational Effectiveness.
- Muralidharan, K. 2013. "Priorities for Primary Education Policy in India's 12th Five-Year Plan." India Policy Forum 9: 1-46.
- Muralidharan, K., J. Das, A. Holla and A. Mohpal. 2014. "The Fiscal costs of Weak Governance: Evidence from Teacher Absence in India." NBER Working Paper 20299.
- Muralidharan, K., P. Niehaus and S. Sukhtankar. 2014. "Payments Infrastructure and the Performance of Public Programs: Evidence from Biometric Smartcards in India." NBER Working Paper 19999.
- Muralidharan, K. and N. Prakash. 2013. "Cycling to School: Increasing Secondary School Enrollment for Girls in India," NBER Working Papers 19305, National Bureau of Economic Research, Inc.
- Muralidharan, K., V. Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India", Economic Journal 120:546.

- Muralidharan, K., V. Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India", *Journal of Political Economy* 119:1.
- Muralidharan, K., V. Sundararaman. 2013a. "The Aggregate Effect of School Choice: Evidence from a two-stage experiment in India." Unpublished Manuscript.
- Muralidharan and Sundararaman. 2013b "Contract Teachers: Experimental Evidence from India". Unpublished Manuscript.
- Murnane, R. and A. Ganimian. 2014. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations." National Bureau of Economic Research Working Paper 20284.
- Pandey, P, S. Goyal and V. Sundararaman. 2008. "Public Participation, Teacher Accountability and School Outcomes: Findings from Baseline Surveys in Three Indian States", World Bank Policy Research Working Paper 4777.
- Pandey, P, S. Goyal and V. Sundararaman. 2011. "Does information Improve School Accountability? Results of a Large Randomized Trial." South Asia Human Development Sector Report No. 49. The World Bank.
- Parajuli, D., G. Acharya, N. Chaudhury and B.B. Thapa. 2012. "Impact of Social Fund on the Welfare of Rural Households – Evidence from the Nepal Poverty Alleviation Fund." Policy Research Working Paper 6042. World Bank.
- Petrosino, A., C. Morgan, T. A. Fronius, E. E. Tanner-Smith, R. F. Boruch. 2012. "Interventions in Developing Nations for Improving Primary and Secondary School Enrollment of Children: A Systematic Review." *Campbell Systematic Reviews* 2012:19.
- Pradhan, M., D. Suryadarma, A. Beatty, M. Wong, A. Alishjabana, A. Gaduh and R. P. Artha. 2011. "Improving Educational Quality through Enhancing Community Participation – Results from a Randomized Field Experiment in Indonesia." Policy Research Working Paper 5795. World Bank.
- Pritchett, L. and R. Murgai. 2007. "Teacher Compensation: Can Decentralization to Local Bodies Take India from Perfect Storm through Troubled Waters to Clear Sailing?," *India Policy Forum* 3: 123-168.
- Rao, G. 2014. "Familiarity Does *Not* Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools." University of California, Berkeley.
- Sankar, D. 2013. "Improving Early Childhood Development through Community Mobilization and Integrated Planning for Children." World Bank. SASHD Discussion Paper Series. Report No. 59.
- Sarr, L.R. Dang, H-A. Chaudhary, N. Parajuli, D and Asadullah, N. 2010. "Reaching Out-of-School Children (ROSC) Project Evaluation Report." World Bank.
- Sterne, J. and R. Harbord. 2004. "Funnel Plots in Meta-Analysis." *Stata Journal* 4: 127-141

World Bank. 2004. "World Development Report 2004: Making Services Work for Poor People." The World Bank.

World Bank. 2012. "Bangladesh Education Quality Note." Draft background paper prepared for the Bangladesh Education Sector Review. South Asia Human Development Unit. The World Bank.

## Appendices

### Appendix A: Background on Systematic Review

This appendix will provide methodological and other background information on the systematic review of the literature on education interventions in South Asia underlying this paper. More specifically, the appendix will describe the systematic search strategy underlying the review, explain the criteria that determine whether a specific study was included in the final sample of 29 studies documenting rigorously evaluated education interventions in South Asia, and present descriptive statistics on countries covered, methodologies used and other features of these 29 studies.

In the systematic search, a two-step approach was used to identify the universe of studies that document rigorously evaluated education interventions in South Asia. *First*, a search was performed in a number of databases to identify all studies that (i) explain education-related outcome variables like enrollment, attendance, math, native language, English or composite test scores, (ii) are potentially able to attribute differences or changes in this outcome variable to a clearly-defined education-related intervention, and (iii) use data for at least one South Asian country. Databases used for the search were the American Economic Association's Econlit database, the World Bank's Impact Evaluations in Education (IE<sup>2</sup>) database, the database of randomized evaluations from the Abdul Latif Jameel Poverty Action Lab (J-PAL) and an internal database of impact evaluations compiled by the office of the World Bank's Chief Economist for the South Asia region. *Then*, additional relevant articles were identified through the reference lists of the articles found in the four databases. This "snowballing" is enormously helpful for the identification of unpublished ("grey") or otherwise obscure literature and for evidence that does not neatly fit the inclusion criteria defined by the different databases. For instance, the American Economic Association's EconLit database only lists studies in the economics literature.<sup>13</sup>

The two-step search procedure ensured that a complete and unbiased picture of the available evidence could be collected. It resulted in a preliminary sample of 49 studies. In order to be included in the final set of studies, a number of additional conditions related largely to the quality of the methodology and the comparability of results had to be satisfied. The final sample of the systematic review included only the 29 studies that (i) relied on an RCT or a credible quasi-experimental design, (ii) reported impacts on school-aged children, and (iii) had a "business-as-usual" control group, (iv) reported enrollment effects in percentages or percentage points, (v) reported attendance effects in percentages, (vi) reported learning effects in standard deviations, (vii) listed not only marginal effects but also the associated standard errors, and (viii) did not answer the same research question using the same data as an earlier study already included in the systematic review. These additional criteria disqualified a number of studies from the final sample of the systematic review and the ensuing meta-analysis that nevertheless offer valuable insights into education access and quality in South Asia. For instance,

---

<sup>13</sup> Most systematic reviews use "Snowballing" in addition to or instead of relying exclusively on searching for a set of key words or strings in different databases. Conclusions and patterns identified through both approaches tend to be quite similar (Greenhalgh and Peacock, 2005; Jalali and Wohlin, 2012).

Aslam and Kingdom (2011) did not rely on an RCT or quasi-experimental design, Barrera-Osorio and Raju (2011) gave enrollment effects not in terms of percentages or percentage points, IEG (2011) did not report their estimates' standard errors, and Berry (2014) did not have a "business-as-usual" control group. All these studies are included in the more verbal discussion of section 3.

In order to provide an overview over the rigorous evidence on education-related interventions in South Asia, important features of the studies included in the systematic review and meta-analysis are visualized in Figure 5. The figure summarizes the number of studies or interventions by actor, country, methodology and extent of World Bank involvement. Its upper-left panel shows that a certain amount of rigorous evidence exist for all four actors identified above – teachers, schools, households and communities. Since some of the studies included in the final sample of the systematic review document impacts of more than one intervention and these might target different actors, the upper-left panel of Figure 5 lists not the number of studies but the number of interventions by actor. It demonstrates that eleven interventions targeted the community, nine interventions households, ten interventions teachers and 19 interventions schools or school administrations.<sup>14</sup>

The other three panels of Figure 5 are aggregated on the level of studies instead of actors. Among other things, they show that in geographic terms, the sample is much less balanced than with regard to the four actors. In fact, as the upper-right panel of Figure 5 demonstrates, the large majority of rigorous education-related impact evaluations focus on India.<sup>15</sup> Whereas 20 studies evaluate interventions' impacts for this country, only one study each uses data for Afghanistan, Bangladesh or Sri Lanka. While slightly more evidence is available for Nepal and Pakistan, Bhutan and the Maldives are not covered by the sampled studies at all.

Figure 5's lower-left panel demonstrates that 21 of the 29 studies included in the final sample of the systematic review are RCT. Two of these are RCTs with incomplete compliance where the randomized assignment is used as an instrumental variable. The other remaining studies rely on quasi-experimental designs such as difference-in-differences, regression discontinuity or a combination thereof.<sup>16</sup> Finally, the lower-right panel of Figure 5 categorizes the 29 studies according to the proportion of their authors that indicated being affiliated with the World Bank at the time of the study's publication. This is one measure of how prominent the World Bank is in commissioning and executing rigorous impact evaluations of education-related interventions in South Asia. As is evident from Figure 5, authors affiliated with the World Bank participated in 13 of the 29 sample studies, i.e. they were involved in

---

<sup>14</sup> One of the interventions evaluated by Sarr et al. (2010) combines grants to schools and education allowances to students and thus addresses different actors at the same time. It is not listed in Figure 5.

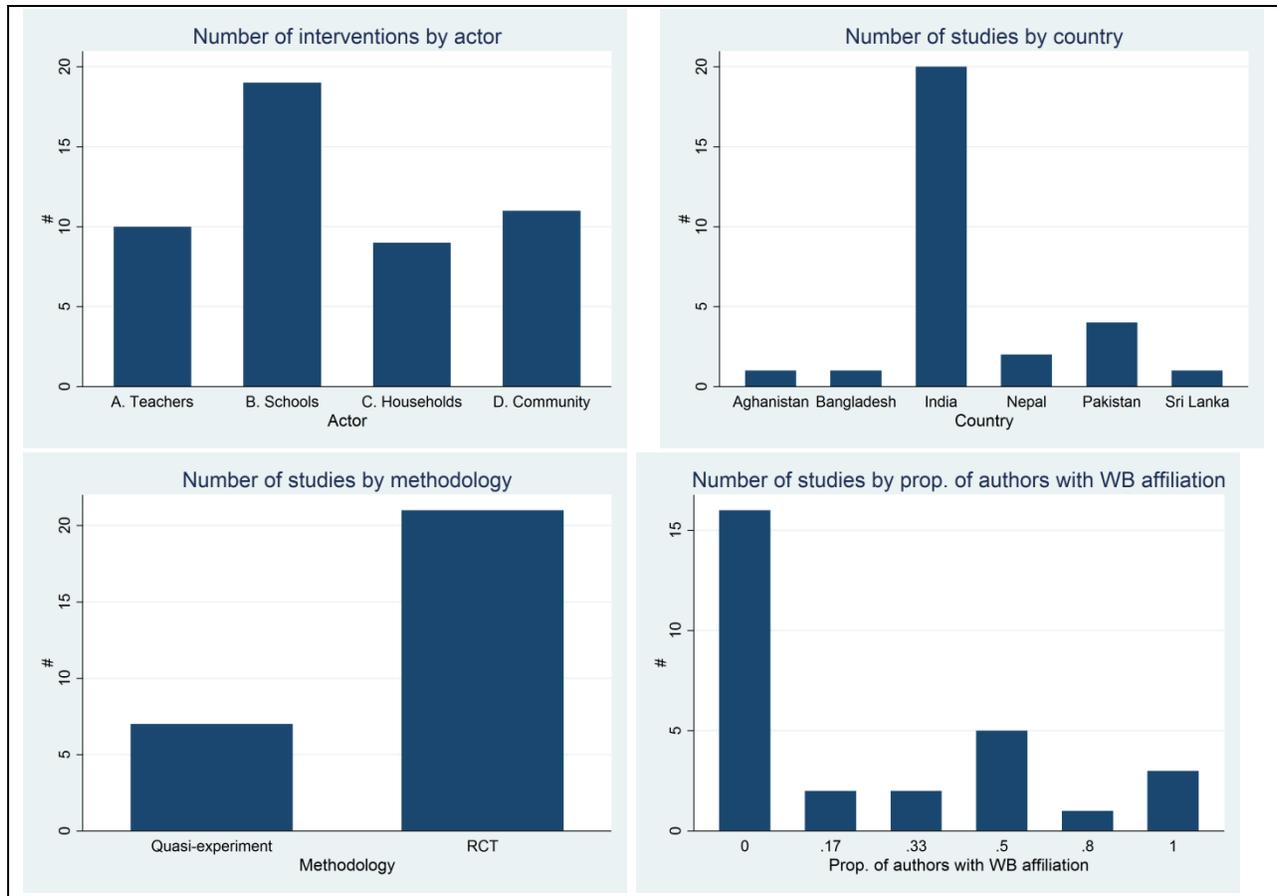
<sup>15</sup> No study contains data for more than one South Asian country. Das et al. (2013) evaluate evidence both from India and Zambia.

<sup>16</sup> Difference-in-differences methods compare a treatment and comparison group (first difference) before and after a program (second difference). Unlike in an RCT, assignment to treatment or comparison group is not random (Baker, 2000). Regression discontinuity designs are possible for programs that assign a threshold separating the treatment and control groups by comparing data points on either side of the threshold. A number of other techniques such as regressions controlling for fixed effects, instrumental variable regressions and propensity score matching estimators are sometimes included in a broader definition of quasi-experimental designs but not considered in our systematic review and the ensuing meta-analysis because they are arguably less rigorous than RCTs or quasi-experiments in the narrow sense of the word.



almost half of the studies identified by the systematic review. Three of these 13 studies were exclusively authored by World Bank staff. The remaining 16 studies did not have a World-Bank-affiliated author but even some of the studies categorized into this group were at least partly financed by the World Bank or benefitted from other forms of World Bank support. Banerjee et al. (2007), for instance, acknowledge financial support from the World Bank.

**Figure 5 – Features of the Studies Included in the Systematic Review and Meta-Analysis**



**Source:** World Bank staff calculations based on studies listed in Table A.

## Appendix B: Critical Assessment of Available Evidence

A critical evaluation of the available evidence can guide the interpretation of results in terms of potential for bias in studies included for the meta-evaluation. Studies were evaluated independently by two reviewers according to a five-point scale separately for RCTs and quasi-experiments; any differences were resolved until a consensus was reached. The Quality-scale evaluates studies according to five indicators, with each indicator awarded a one point for “Yes” answer and zero points for “No” answer. For RCT based studies the indicators include the following questions: (i) was the study

published in a peer-reviewed journal or completed within the last three years, the average lead-time for journal publications (“published or recent”); (ii) was randomization appropriate and no threats to internal validity present (“no endogeneity concerns”); (iii) was information provided on the number and reasons of drop out of study participants (“no systematic attrition”); (iv) was the study sufficiently powered with explicit power calculations reported or was the size of treatment/control clusters 50 or more (“arm size  $\geq 50$ ”); (v) was the study sufficiently representative of the state/region where the experiment was conducted (“no sample selectivity”). For Non-RCT studies, all indicators were the same except for indicator (iv) which was replaced by an indication of the robustness of methods: were the study results robust to at least two sets of statistical methods (“effects robust across methods”).

**Table D – Quality of Available Evidence by Quality Criterion and Method**

Criterion	Sample	
	RCTs only	Both RCTs and quasi-experiments
Published or recent	65 %	63 %
No endogeneity concerns	97 %	82 %
No systematic attrition	71 %	70 %
Arm size $\geq 50$	76 %	-
No sample selectivity	24 %	26 %
Effects robust across methods	-	50 %

**Notes:** See text for definitions of criteria. Arm-size criterion used for RCTs only, robustness criterion for quasi-experiments only.

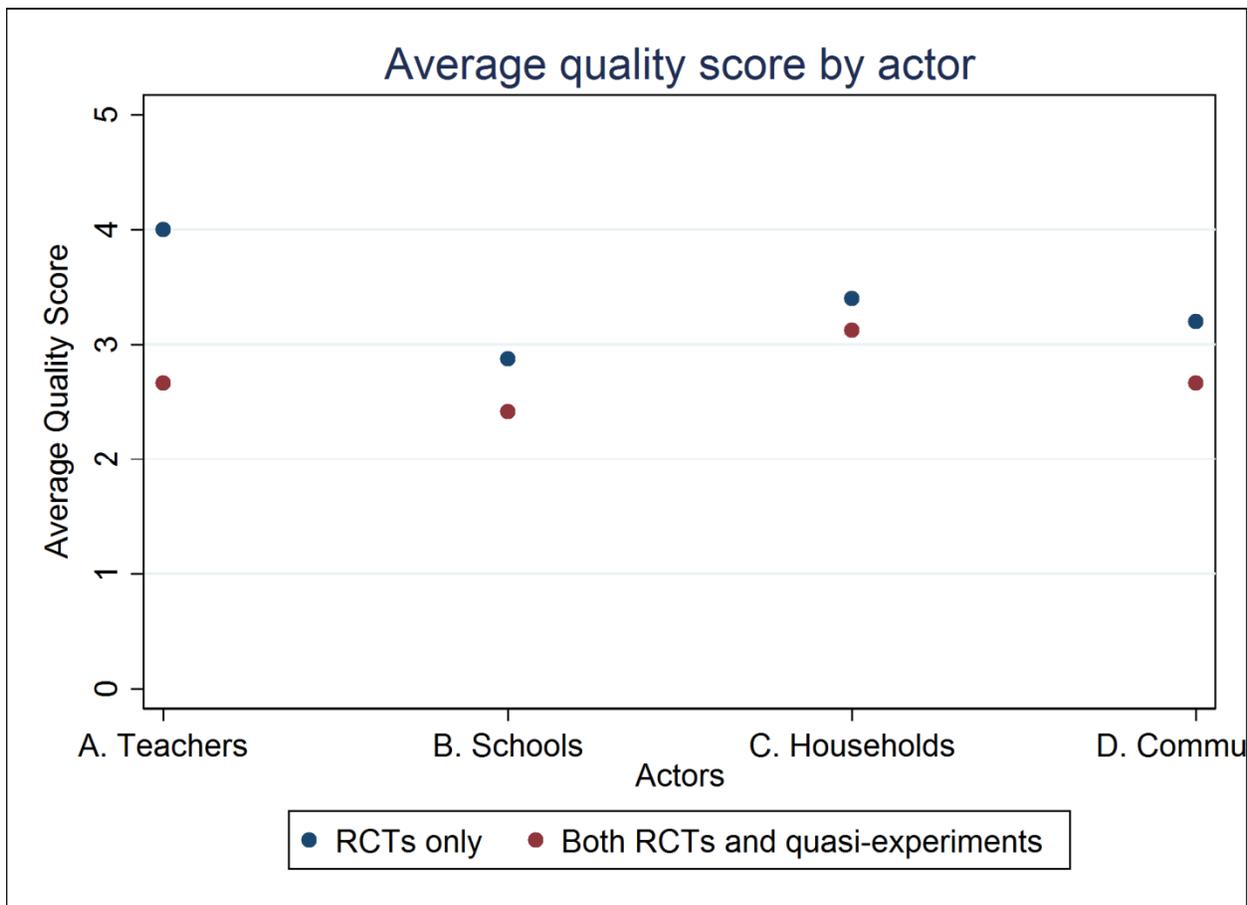
**Source:** World Bank staff calculations based on studies listed in Table A.

Table D summarizes the result of the critical assessment of the available evidence. It shows that almost all RCTs identified through the systematic review of section 3 appropriately address endogeneity concerns. Such concerns are more of an issue for quasi-experiments but all in all 82 percent of the studies underlying the meta-analysis of section 3 are of high quality according to this criterion. Quality is also generally high with respect to the sample attrition criterion; in only 30 percent of cases was the information provided on the number and reasons of study participants that had dropped out of the sample insufficient. For this criterion, no marked difference in average quality between RCTs and quasi-experiments exists. The same is true for the criterion that measures whether a study was published in a peer-reviewed journal or completed within the last three years. For both RCTs and quasi-experiments, this criterion is fulfilled for somewhat less than two thirds of the studies. Whereas sufficient arm size is not a critical issue for the large majority of RCTs either, the remaining criteria appear generally more problematic and are fulfilled by fewer studies. In particular, only half of the relevant impact evaluations that rely on quasi-experiments can show that results are robust to at least two sets of statistical methods and only about a quarter of both RCTs and quasi-experiments use samples that are representative at least on the level of the state or region of data collection.

In spite of these shortcomings, we would argue that the impact evaluations included in the systematic review and meta-analysis of section 3 are generally of high quality. This impression is confirmed by

Figure 6 which visualizes the average quality of the available evidence by the method that was used as well as by the four actors introduced in section 2. Irrespective of which method is used and what actor addressed, all average scores are at least around two and a half (out of five) and sometimes considerable higher on the quality scale. While this represents far from a perfect result, it still appears remarkable that the relatively young literature that uses rigorous methods to evaluate education-related interventions in South Asia has already attained such a respectable level of quality.

Figure 6 – Quality of Available Evidence by Actor and Method



Notes: See text for definitions of criteria.

Source: World Bank staff calculations based on studies listed in Table A.

Figure 6 also shows that RCT studies systematically rank higher than the quasi-experiment studies for all four actors on the five-item quality scale. Moreover, studies pertaining to teachers and households score higher than studies on schools and community. One possible reason for this could be a better theoretical understanding of how teachers and parents respond to resources and incentives, resulting in better designed experiments compared to studies on schools and communities where such

understanding is still limited. Due to the relatively few studies included for each actor, the findings on the quality should be interpreted as preliminary and not necessarily conclusive or generalizable.

A complementary way to assess the overall quality of the body of literature that evaluates the effectiveness of education-related interventions in South Asia is to rely on so-called funnel plots. Funnel plots are scatterplots where the treatment effects estimated from individual impact evaluations are plotted on the horizontal axis and the standard errors of these treatment effects are charted on the vertical axis. The underlying idea is that in the absence of biases, results from studies with a small sample size will tend to be imprecise and will scatter widely at the bottom of the graph. In contrast, larger studies evaluating the same or similar interventions will produce point estimates that will hardly differ from each other and will also be very precisely measured.

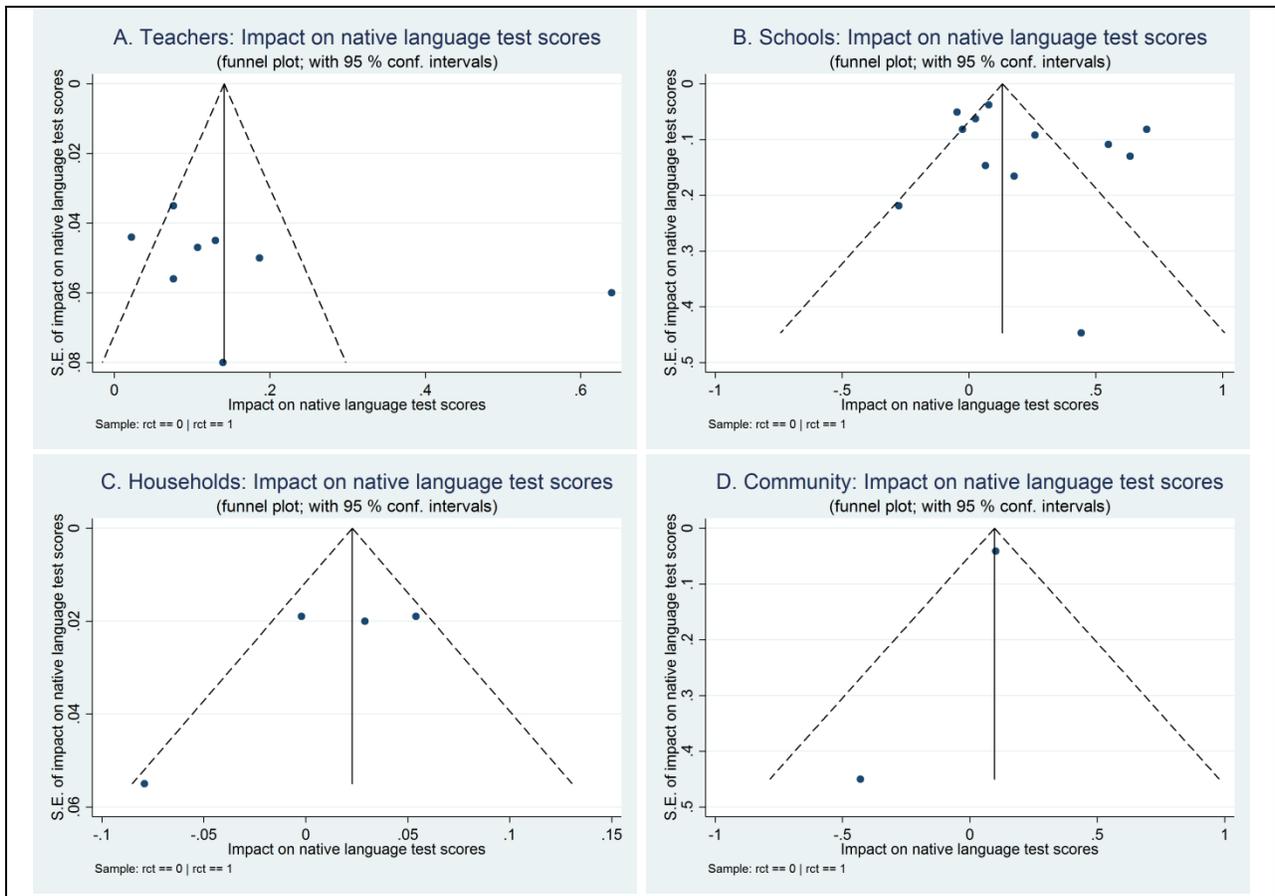
Funnel plots for education-related interventions' impacts on native language, mathematics and composite test scores are visualized in Figures 7 to 9. Following the conceptual framework of section 2 and the meta-analysis of section 3, interventions are grouped according to whether they target teachers, schools, households or communities. That is, the grouping follows the four actors that were identified in section 2 and also at the heart of the meta-analysis of section 3. Of course, it would once again be perfectly feasible to draw funnel plots not for the four actors but instead to compare biases for demand- or supply-focused interventions, programs targeting individuals or collectives and interventions aimed at providing resources or influencing actors' incentives.

As noted by Sterne and Harbord (2004), "(f)unnel plots were first proposed as a means of detecting a specific form of bias – publication bias." Publication bias exists if among all studies that test the effectiveness of different interventions, those that show significant positive effects are more likely to be published in scientific journals or similar outlets than those that find negative or statistically insignificant effects. This particular form of bias might for instance be caused by journal editors who are most interested in publishing results of startlingly effective interventions or by funding or implementing agencies that enjoy bragging about their achievements but would rather not admit the failure of some of the interventions they sponsor. Publication bias is a severe threat to the validity of any meta-analysis because it can lead to upwardly-biased estimates about the effectiveness of certain interventions. It can partly be mitigated by a thorough search strategy of the systematic review underlying a meta-analysis that incorporates not only studies in peer-reviewed scientific journals but also unpublished results or "grey" literature like working papers or preliminary reports. Such a search strategy is likely to help in identifying studies that rigorously document the impacts of various interventions but fail to be officially published because they do not demonstrate that a certain intervention works in improving outcomes. As mentioned in Appendix A, the search strategy that was the part of the systematic review underlying our meta-analysis follows exactly such guidelines and aims to be as comprehensive as possible.

Once a body of literature is identified through a systematic review, funnel plots can provide further evidence about whether publication bias is present. If this is the case, funnel plots will have asymmetrical appearances. More precisely, there will be a gap in the left bottom side of the graphs (the place where one would expect to see studies that find that interventions have low, statistically

insignificant or even negative effects). As again noted by Sterne and Harbord (2004), the interpretation of funnel plots is further facilitated by the inclusion of vertical lines for a group of intervention’s treatment effect as estimated with the help of a meta-analysis and of diagonal lines representing the (pseudo) confidence intervals around these effect. Such diagonal lines – by convention drawn here for 95 percent confidence intervals – show the expected distribution of studies in the absence of publication biases.

**Figure 7 – Funnel Plots for Impacts on Native Language Test Scores**



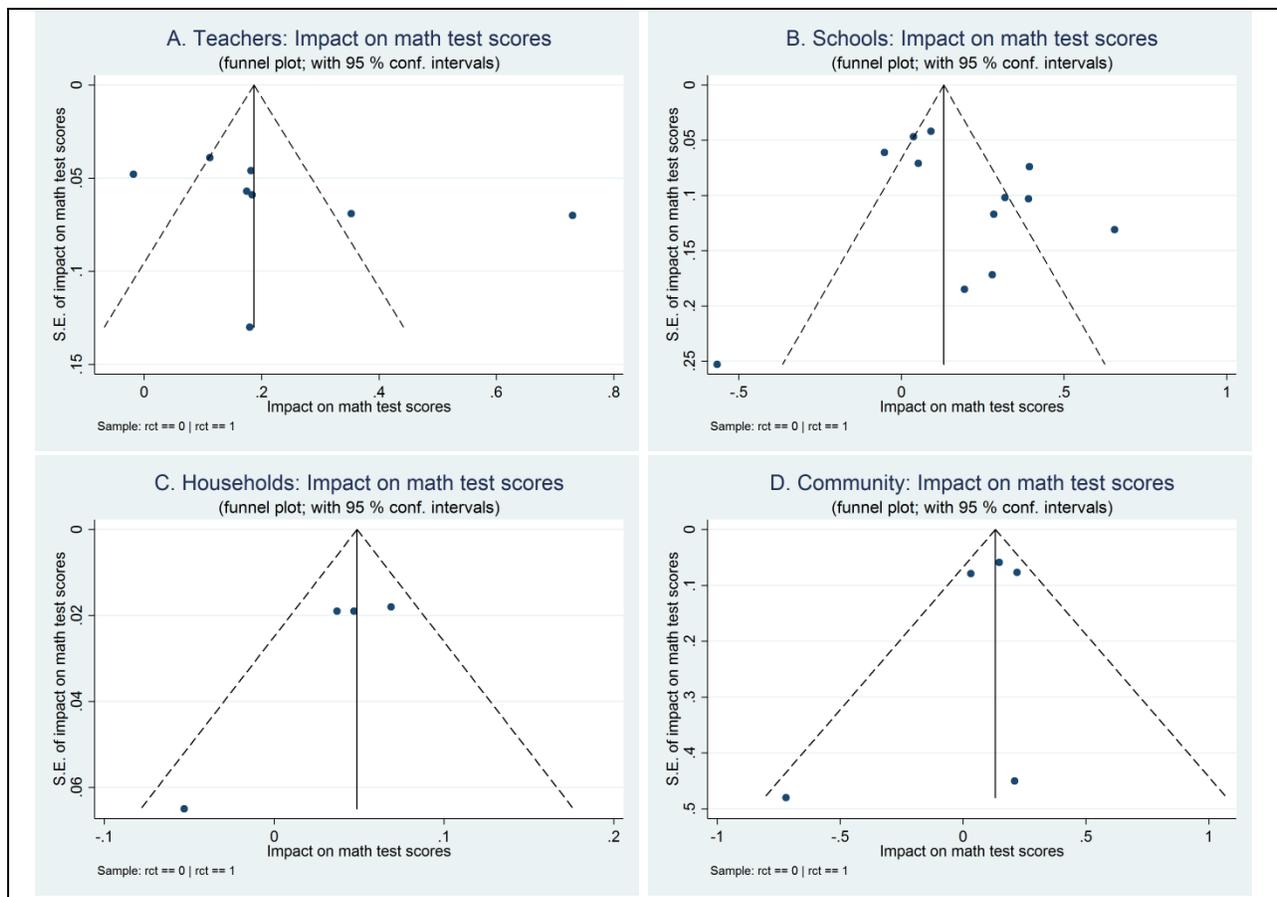
**Notes:** Vertical lines denote point estimates from meta-analyses and diagonal lines 95 percent pseudo confidence intervals around these estimates. Asymmetrical appearances and gaps in the left bottom side of graphs are indications of small sample effects.

**Source:** World Bank staff calculations based on studies listed in Table A.

It should be noted, however, that according to Egger et al. (1997) publication bias is not the only possible explanation for any small study effects such as asymmetries in funnel plots or data points lying outside the 95 percent confidence limits. First, distortions caused by the failure to include all relevant estimates and especially those from based on small samples in a meta-analysis (“selection bias”) might be due to factors other than publication bias. For instance, Jueni et al. (2002) document

that studies reporting the results of clinical trials that are published in languages other than English tend to include relatively small samples and to be overlooked by many meta-analyses. In addition to different forms of selection bias, asymmetries in funnel plots might also be caused by heterogeneous effects of individual interventions, and data and methodological irregularities (which might be most severe in small studies), artefacts or simply chance. Heterogeneous effects of individual interventions, in particular, might also be behind studies that fall far outside the 95 percent confidence intervals, as is arguably the case for one of the teacher-centric intervention in each of the three upper-left panels of Figures 7 to 9 (in fact, all three data points relate to the intervention providing remedial teaching by community volunteer and other support analyzed by Lakshminarayan et al., 2013) and some of the school-centric interventions in the figures' three upper-right panels.

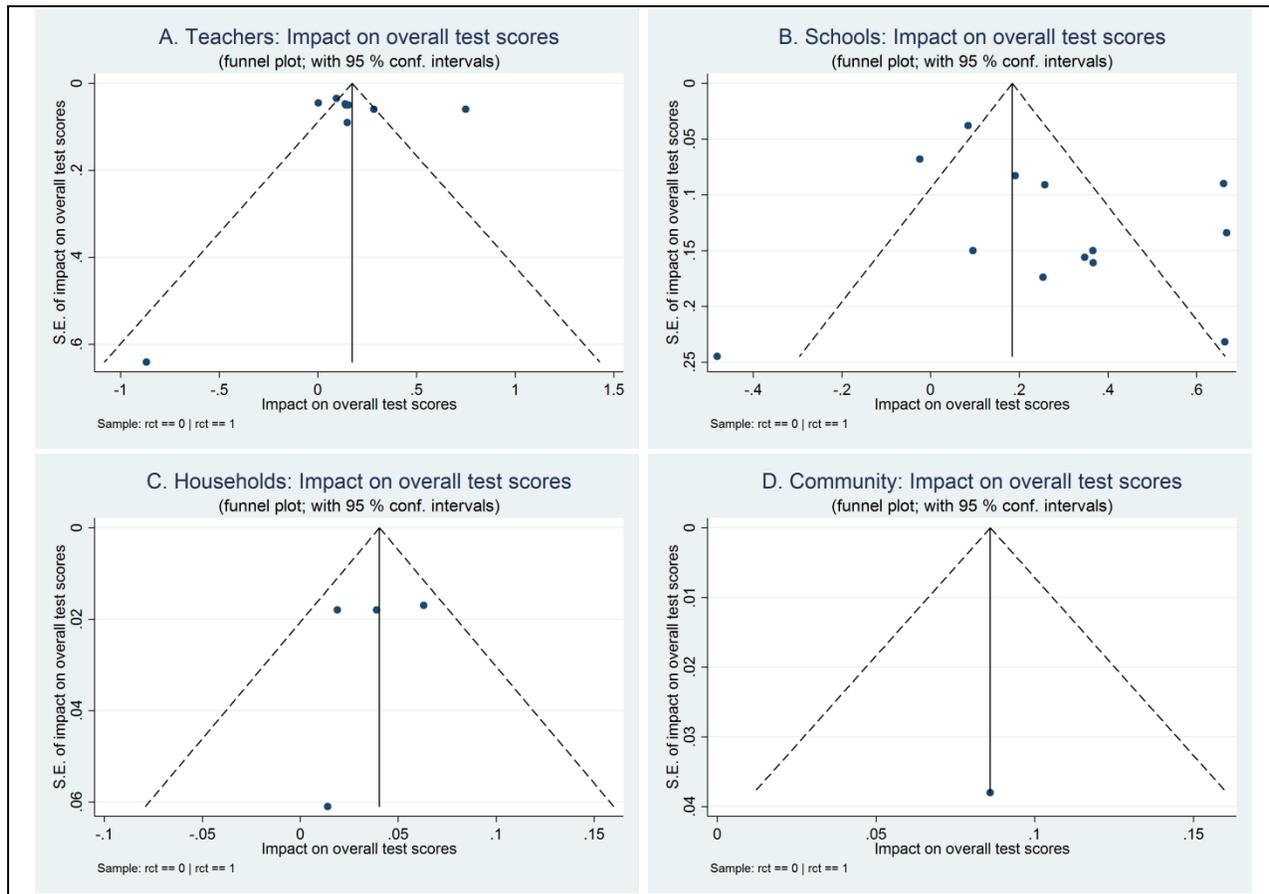
**Figure 8 – Funnel Plots for Impacts on Math Test Scores**



**Notes:** Vertical lines denote point estimates from meta-analyses and diagonal lines 95 percent pseudo confidence intervals around these estimates. Asymmetrical appearances and gaps in the left bottom side of graphs are indications of small sample effects.

**Source:** World Bank staff calculations based on studies listed in Table A.

**Figure 9 – Funnel Plots for Impacts on Composite Test Scores**



**Notes:** Vertical lines denote point estimates from meta-analyses and diagonal lines 95 percent pseudo confidence intervals around these estimates. Asymmetrical appearances and gaps in the left bottom side of graphs are indications of small sample effects.

**Source:** World Bank staff calculations based on studies listed in Table A.

Nevertheless, and irrespective of the actual reasons behind any asymmetric funnel plots, Figures 7 to 9 contain little evidence that publication bias is a grave worry in our context. For interventions targeting households or communities, the evidence for impacts on native language, mathematics or composite test scores is generally rather limited but funnel plots appear very symmetric. For teacher-centric programs somewhat more evidence is available but again the respective funnel plots do not indicate any publication biases. If anything, they are asymmetric in a way that suggests that fewer studies who find that such interventions have large impacts on learning outcomes are present in the meta-analysis than one would expect if no publication biases or other small study effects were present. Finally, the only group of interventions for which there is some indication of a publication bias is the one that focuses on schools. In particular with regard to interventions’ effects on mathematics and composite tests scores, quite a bit of evidence is available but the funnel plots do not really look

symmetrical. Instead, in both cases there appears something of a gap in the left bottom side of the graph. While these gaps are not very marked, they nevertheless caution against drawing very strong policy recommendations from the apparent effectiveness of school-centered interventions to improve mathematics and composite test scores in South Asia.

## **Appendix C: Details on Meta-Analysis**

This appendix will provide more details on the meta-analyses of section 3. In particular, it will elaborate on the methodology underlying these meta-analyses and list a table with detailed outputs. Moreover, it will show the findings of meta-analyses of education-related interventions in South Asia that were pooled over all four actors. Finally, it will elaborate on the mechanics underlying the meta-analyses' results and summarize the results of meta-analyses of education-related interventions in South Asia where interventions are grouped differently or alternative methodologies are used. In this context, it will explore how results change when only impact estimates derived from RCTs are considered and whether conclusions differ systematically for the subgroup of studies with at least one author with a World Bank affiliation at the time of publication.

As mentioned in section 3, in the context of a meta-analysis the selection of the underlying evidence, the grouping of individual interventions, the weighting of the evidence, and the selection of adequate outcome measures are of paramount importance. In addition to that, a meta-analysis of education interventions in South Asia posed a number of more practical challenges. Maybe most critically, some studies reported interventions' impacts on learning outcomes in a way that was not directly comparable to estimates from other studies. For instance, Lakshminarayana et al. (2013) reported estimated impacts on test scores in terms of changes on a percentage scale. Whenever possible, every effort was made to transform estimated impacts into standardized effects. The aim was to construct a set of evidence underlying the meta-analyses that was both comprehensive and unbiased. For instance, because baseline standard deviations were also reported by Lakshminarayana et al. (2013), their estimated impacts in terms of changes on a percentage scale could be transformed into impacts in terms of changes in standard deviations comparable to what was reported by other studies.

Moreover, many studies report results for different groups, time periods and specifications. This means a decision had to be made about which impact estimates to include in the meta-analysis. Here, this decision was made according to the general principle that each meta-analysis would include exactly one impact estimates for each distinct intervention. That is, two impact estimates were included in a meta-analysis for a study detailing two distinct interventions but only one impact estimate for one distinct intervention discussed by more than one study. For RCTs, distinct interventions were defined as those having their own treatment arm. Besides, whenever available, a meta-analysis would include pooled impact estimates for boys and girls. Only in cases where these were not available, it would include separate impact estimates for boys and girls. Similarly, whenever impact estimates for different time horizons were available, a meta-analysis would include the one closest to interventions' impacts after twelve months which was the median time span between intervention and measurement of



impacts. Finally, whenever impact estimates for different specifications were available, a meta-analysis would include the one for the author's (or authors') preferred specification and whenever separate impact estimates for an intervention's *intent to treat effect* and *treatment effect on the treated* were available, a meta-analysis would include the impact estimate for the *intent to treat effect*.

Together with the general principles outlined in section 3, these more practical considerations led to a set of impact estimates. These are summarized in Table E. Most of the conclusions that can be drawn from meta-analyses of the interventions listed in the table have already been outlined in section 3, but it might be worth mentioning three additional points. First, Table A (which contains all rigorously evaluated education interventions in South Asia identified by the systematic review underlying the meta-analyses) lists 29 distinct studies. In contrast, Table E only contains 22 studies. This discrepancy can be explained by the fact that whereas Table A contains the results of the systematic review, Table E only lists those studies that actually underlie the meta-analyses in section 3. As a consequence, seven studies recorded in Table A that contain estimated impacts on enrollment, attendance or other education-related variables but not on standardized test scores in mathematics, native language or composite are not listed in Table E.

Second, Table E lists the estimated effects of 36 distinct interventions. As mentioned above, the number of distinct interventions is higher than the number of distinct studies mostly because a large number of RCTs estimate impacts for more than one treatment arm. These different treatment arms might analyze the impacts of very different treatments (for instance the employment of a *balsakhi* and the introduction of computer-assisted learning; Banerjee et al., 2007), different implementations of basically the same intervention (like four distinct implementations of a literacy program; He, Linden and MacLeod, 2009) or even the same intervention's impacts in different samples (e.g. school-based management in grade three and five; Chaudhury and Parajuli, 2010b).<sup>17</sup> Since it often appears arbitrary whether different treatment arms are analyzed in one publication or in a larger number of separate studies, the meta-analyses treat them as entirely independent.

Third, in addition to the details on the meta-analyses of interventions' impacts on learning grouped by the four actors, Table E also contains the results of three meta-analyses of the learning effects of education interventions overall. These results are displayed in the last row of the table. They show that, overall, education interventions that have been the subject of rigorous impact evaluations increase native language test scores by 0.07 standard deviations, mathematics test scores by 0.09 standard deviations and composite test scores again by 0.09 standard deviations. While the policy implications of a meta-analysis of such a broad and diverse set of interventions might be debatable, one could certainly follow the interpretation of Krishnaratne, White and Carpenter (2013, p.42) that performed a similar exercise for education interventions from all over the world and concluded simply by saying that "(i)nterventions aimed at getting children into school work."

---

<sup>17</sup> At the same time, at least two interventions are discussed by more than one study. The short-run impacts of both individual and group bonuses for teachers in Andhra Pradesh, India, are analyzed in Muralidharan and Sundararaman (2011) whereas their medium- and long-term impacts are evaluated in Muralidharan (2012).

**Table E – Details on Meta-Analysis of Interventions’ Impacts on Native Language, Math and Composite Test Scores**

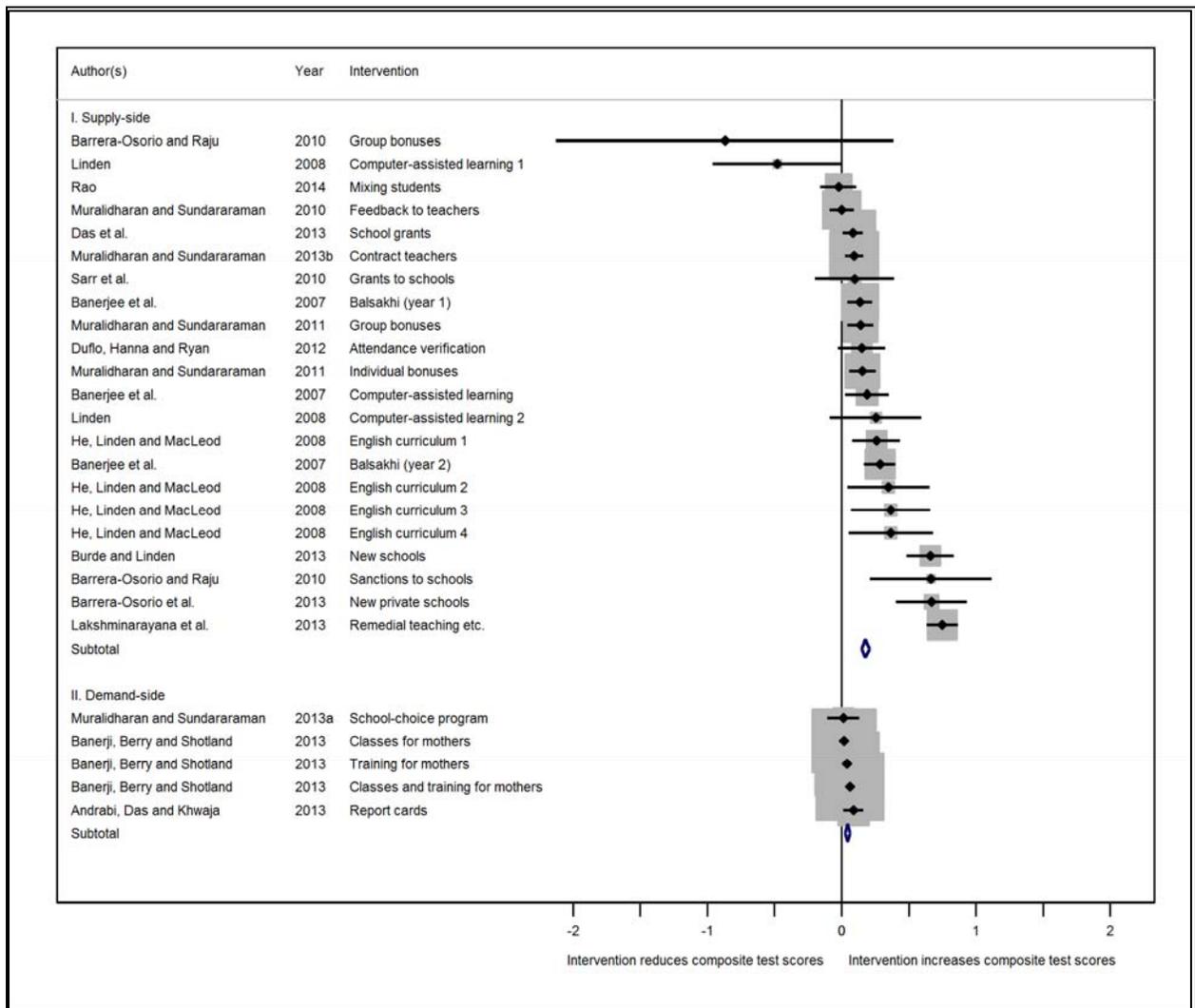
Study	Intervention	S / D	I / C	R / I	Impacts on Native Language Test Scores			Impacts on Math Test Scores			Impacts on Composite Test Scores					
					E.S.	95% C.I.	Wgt.	E.S.	95% C.I.	Wgt.	E.S.	95% C.I.	Wgt.			
<b>Andrabi, Das and Khwaja (2013)</b>	Report cards	D	C	I	0.10	0.02	0.18	4.13	0.15	0.03	0.26	2.02	0.09	0.01	0.16	4.49
<b>Aturupane, et al. (2013)</b>	School-based management	D	C	I					0.22	0.07	0.37	1.19				
	Report cards	D	C	I					0.03	-0.12	0.19	1.13				
<b>Banerjee et al. (2007)</b>	Balsakhi (year 1)	S	I	R	0.08	-0.03	0.19	2.21	0.18	0.09	0.27	3.33	0.14	0.05	0.23	2.93
	Balsakhi (year 2)	S	I	R	0.19	0.09	0.29	2.78	0.35	0.22	0.49	1.48	0.28	0.17	0.40	1.80
	Computer-assisted learning	S	C	R	-0.03	-0.19	0.14	1.03	0.39	0.25	0.54	1.29	0.19	0.03	0.35	0.94
<b>Banerji, Berry and Shotland (2013)</b>	Classes for mothers	D	I	R	-0.00	-0.04	0.04	19.24	0.04	-0.00	0.07	19.52	0.02	-0.02	0.05	20.00
	Training for mothers	D	I	R	0.03	-0.01	0.07	17.36	0.05	0.01	0.08	19.52	0.04	0.00	0.07	20.00
	Classes and training for mothers	D	I	R	0.05	0.02	0.09	19.24	0.07	0.03	0.10	21.75	0.06	0.03	0.10	22.43
<b>Barrera-Osorio and Raju (2010)</b>	Group bonuses	S	I	R									-0.87	-2.13	0.39	0.02
	Sanctions to schools	S	C	I									0.66	0.21	1.12	0.12
<b>Barrera-Osorio et al. (2013)</b>	New private schools	S	C	R	0.64	0.38	0.89	0.41	0.66	0.40	0.91	0.41	0.67	0.41	0.93	0.36
<b>Borkum, He and Linden (2013)</b>	Libraries	S	C	R	-0.05	-0.15	0.05	2.67	-0.05	-0.17	0.07	1.89				
<b>Burde and Linden (2013)</b>	New schools	S	C	R									0.66	0.49	0.84	0.80
<b>Chaudhury and Parajuli (2010b)</b>	School management (grade 3)	D	C	I	-0.43	-1.31	0.45	0.03	-0.72	-1.66	0.22	0.03				
	School management (grade 5)	D	C	I					0.21	-0.67	1.09	0.03				
<b>Das et al. (2013)</b>	School grants	S	C	R	0.08	0.00	0.15	4.81	0.09	0.01	0.17	3.99	0.09	0.01	0.16	4.49
<b>Duflo, Hanna and Ryan (2012)</b>	Attendance verification	S	C	R	0.14	-0.02	0.30	1.09	0.18	-0.08	0.44	0.42	0.15	-0.03	0.33	0.80
<b>He, Linden and MacLeod (2008)</b>	English curriculum 1	S	C	R					0.05	-0.09	0.19	1.40	0.26	0.08	0.44	0.78
	English curriculum 2	S	C	R					0.32	0.12	0.52	0.68	0.35	0.04	0.65	0.27
	English curriculum 3	S	C	R					0.28	0.06	0.51	0.51	0.37	0.07	0.66	0.29
	English curriculum 4	S	C	R					0.39	0.19	0.59	0.66	0.37	0.05	0.68	0.25
<b>He, Linden and MacLeod (2009)</b>	Literacy program 1	S	C	R	0.26	0.08	0.44	0.82								
	Literacy program 2	S	C	R	0.44	-0.43	1.32	0.03								
	Literacy program 3	S	C	R	0.55	0.34	0.76	0.58								
	Literacy program 4	S	C	R	0.70	0.54	0.86	1.03								
<b>Lakshminarayana et al. (2013)</b>	Remedial teaching etc.	S	I	R	0.64	0.52	0.76	1.93	0.73	0.59	0.87	1.44	0.75	0.63	0.87	1.80
<b>Linden (2008)</b>	Computer-assisted learning 1	S	C	R	-0.28	-0.71	0.15	0.14	-0.57	-1.06	-0.07	0.11	-0.48	-0.96	-0.00	0.11
	Computer-assisted learning 2	S	C	R	0.18	-0.15	0.50	0.25	0.28	-0.06	0.62	0.24	0.25	-0.09	0.60	0.21
<b>Muralidharan and Sundararaman (2010)</b>	Feedback to teachers	S	I	I	0.02	-0.06	0.11	3.59	-0.02	-0.11	0.08	3.06	0.00	-0.09	0.09	3.20
<b>Muralidharan and Sundararaman (2011) and Muralidharan (2012)</b>	Group bonuses	S	I	I	0.11	0.02	0.20	3.14	0.18	0.06	0.29	2.17	0.14	0.04	0.24	2.59
	Individual bonuses	S	I	I	0.13	0.04	0.22	3.43	0.18	0.07	0.30	2.02	0.16	0.06	0.25	2.59
<b>Muralidharan and Sundararaman (2013a)</b>	School-choice program	D	I	R	-0.08	-0.19	0.03	2.30	-0.05	-0.18	0.07	1.67	0.01	-0.11	0.13	1.74
<b>Muralidharan and Sundararaman (2013b)</b>	Contract teachers	S	I	R	0.08	0.01	0.15	5.67	0.11	0.04	0.19	4.63	0.09	0.03	0.16	5.29
<b>Rao (2014)</b>	mixing wealthy and poor students	S	C	I	0.03	-0.10	0.15	1.75	0.04	-0.06	0.13	3.19	-0.02	-0.16	0.11	1.40
<b>Sarr et al. (2010)</b>	Grants to schools	S	C	R	0.07	-0.22	0.35	0.32	0.19	-0.17	0.56	0.21	0.17	-0.20	0.39	0.29
<b>A. Teachers</b>					0.14	0.11	0.18	23.84	0.19	0.150	0.23	18.55	0.17	0.14	0.21	21.03
<b>B. School administrators</b>					0.13	0.09	0.18	13.86	0.13	0.09	0.17	14.58	0.18	0.14	0.23	10.31
<b>C. Households</b>					0.02	0.00	0.04	58.13	0.05	0.03	0.07	62.46	0.04	0.02	0.06	64.18
<b>D. Community</b>					0.10	0.02	0.18	4.17	0.13	0.05	0.21	4.41	0.09	0.01	0.16	4.49
<b>Overall</b>					0.07	0.05	0.09	100	0.09	0.08	0.11	100	0.09	0.07	0.10	100

**Notes:** “S/D” denotes supply vs. demand, “I/C” individual vs. collective and “R/I” resources vs. incentives.

**Source:** World Bank staff calculations based on studies listed in Table A.

Figures 10, 11 and 12 summarize the results of meta-analyses of education-related interventions in South Asia where interventions are not grouped by the four actors (teachers, schools, households and communities) but by the more fundamental categories introduced in the conceptual framework of section 2. That is, interventions are grouped by whether they (i) address education supply or demand, (ii) target individuals or a collective, or (iii) provide resources to actors or change their incentives. For conciseness, interventions' impacts on native-language and mathematics are omitted. Instead, the focus is on interventions' impacts on overall learning levels.

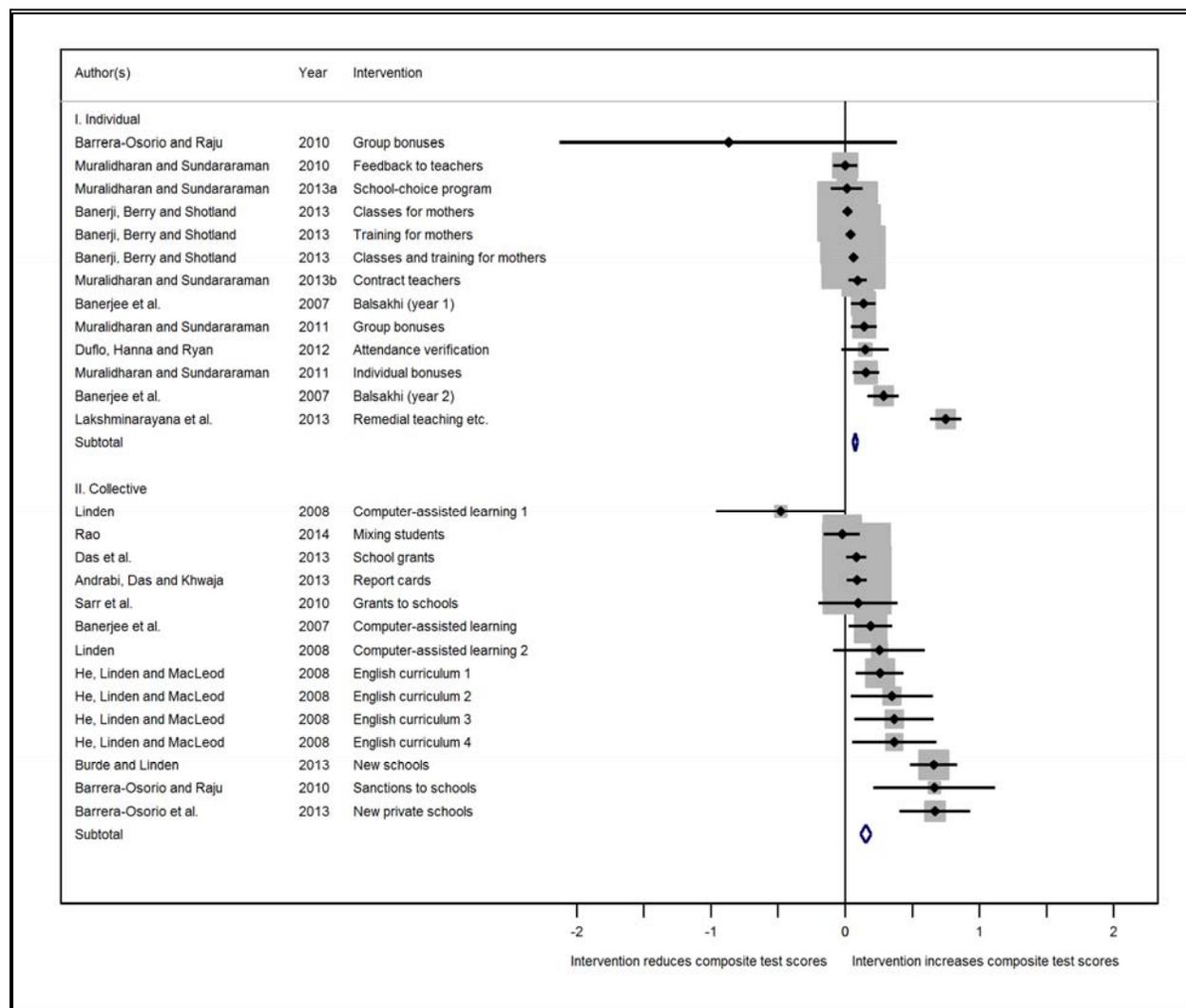
**Figure 10 – Meta-Analysis of Interventions' Impacts on Composite Test Scores (Supply vs. Demand)**



**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates' weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds' left and right edges 95 percent confidence intervals; the sample includes only RCTs.

**Source:** World Bank staff calculations based on studies listed in Table A.

Figure 11 – Meta-Analysis of Interventions’ Impacts on Composite Test Scores (Individual vs. Collective)



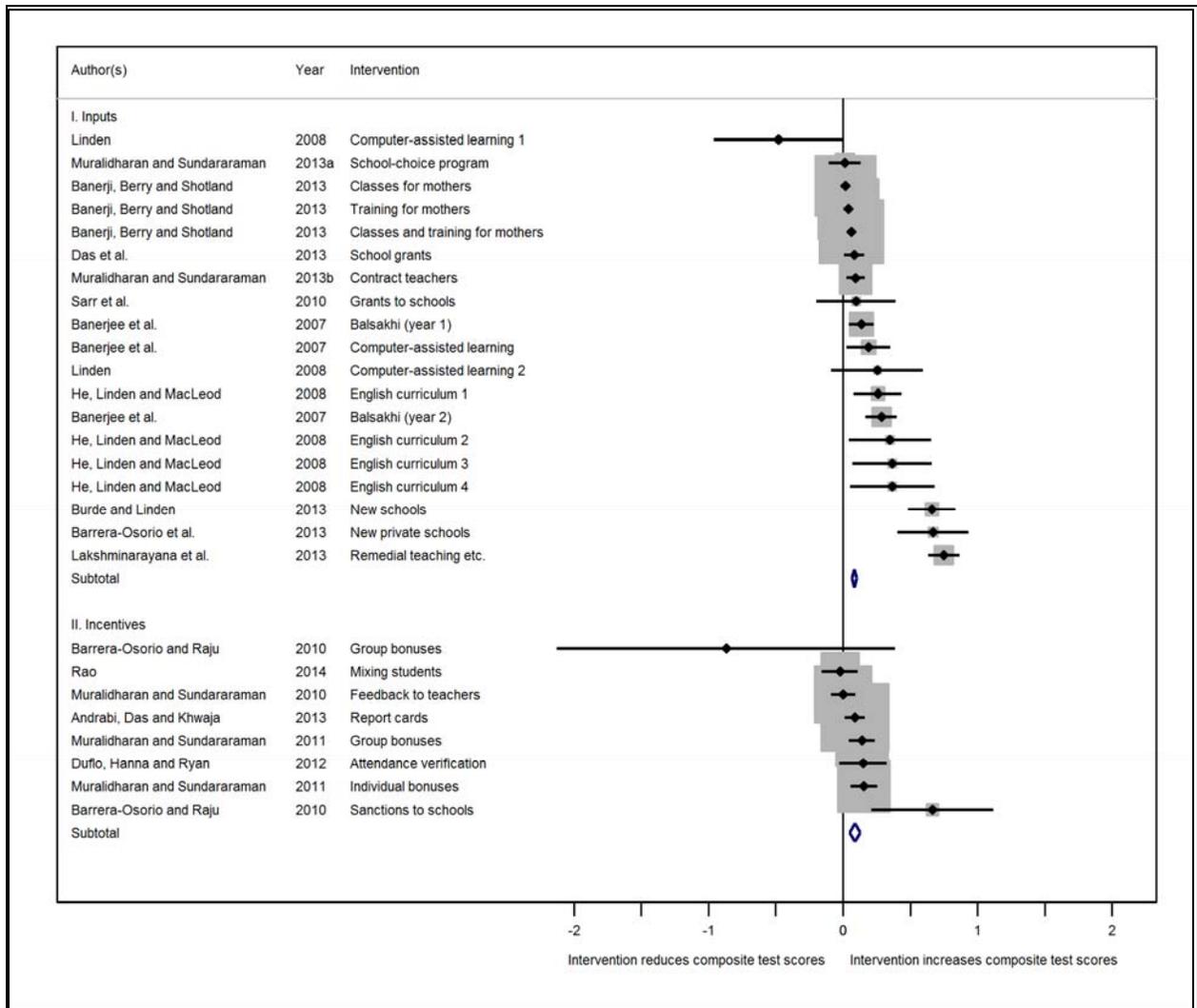
**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates’ weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds’ left and right edges 95 percent confidence intervals; the sample includes only RCTs.

**Source:** World Bank staff calculations based on studies listed in Table E.

As already mentioned in section 3, one of the main results of the meta-analyses is that supply-side interventions (i.e. those targeting teachers or schools) appear significantly more promising for inducing moderate to important improvements in learning outcomes than demand-side interventions (Figure 10). If interventions are categorized according to whether they target individuals or a collective, the result is that programs centered on collective bodies such as schools or communities are more effective in raising learning outcomes than those targeting individual teachers or households. Again, this was

also already mentioned above (Figure 11). Finally, if one categorizes interventions according to whether they emphasize the provision of additional resources or aim to alter individual or collective incentives, one finds positive and significant impacts on composite test scores for both groups of interventions but no significant differences between the groups (Figure 12).

**Figure 12 – Meta-Analysis of Interventions’ Impacts on Composite Test Scores (Resources vs. Incentives)**

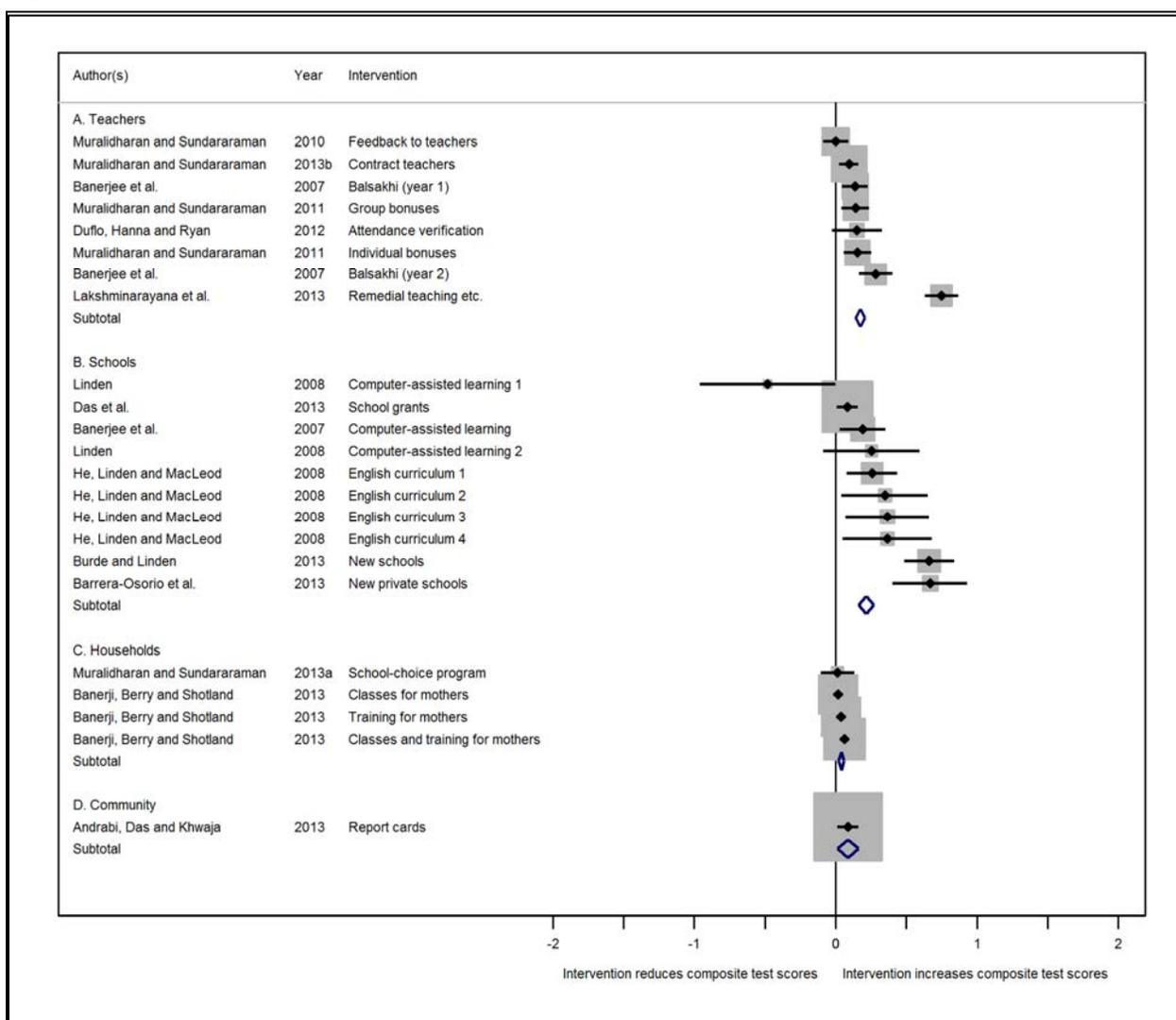


**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates’ weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds’ left and right edges 95 percent confidence intervals; the sample includes only RCTs.

**Source:** World Bank staff calculations based on studies listed in Table A.

In order to better understand the process behind the findings of the meta-analyses of education-related interventions in South Asia, Figures 13 and 14 explore how results change when only impact estimates derived from RCTs are considered or whether conclusions differ systematically between those studies where at least one author was affiliated with the World Bank at the time of publication and impact evaluations where this was not the case. Both figures focus on interventions' impacts on composite test scores as composite test scores are the most comprehensive outcome variable utilized in the meta-analyses of section 3.

**Figure 13 – Meta-Analysis of Interventions' Impacts on Composite Test Scores (RCTs only)**



**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates' weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds' left and right edges 95 percent confidence intervals; the sample includes only RCTs.

**Source:** World Bank staff calculations based on studies listed in Table A.

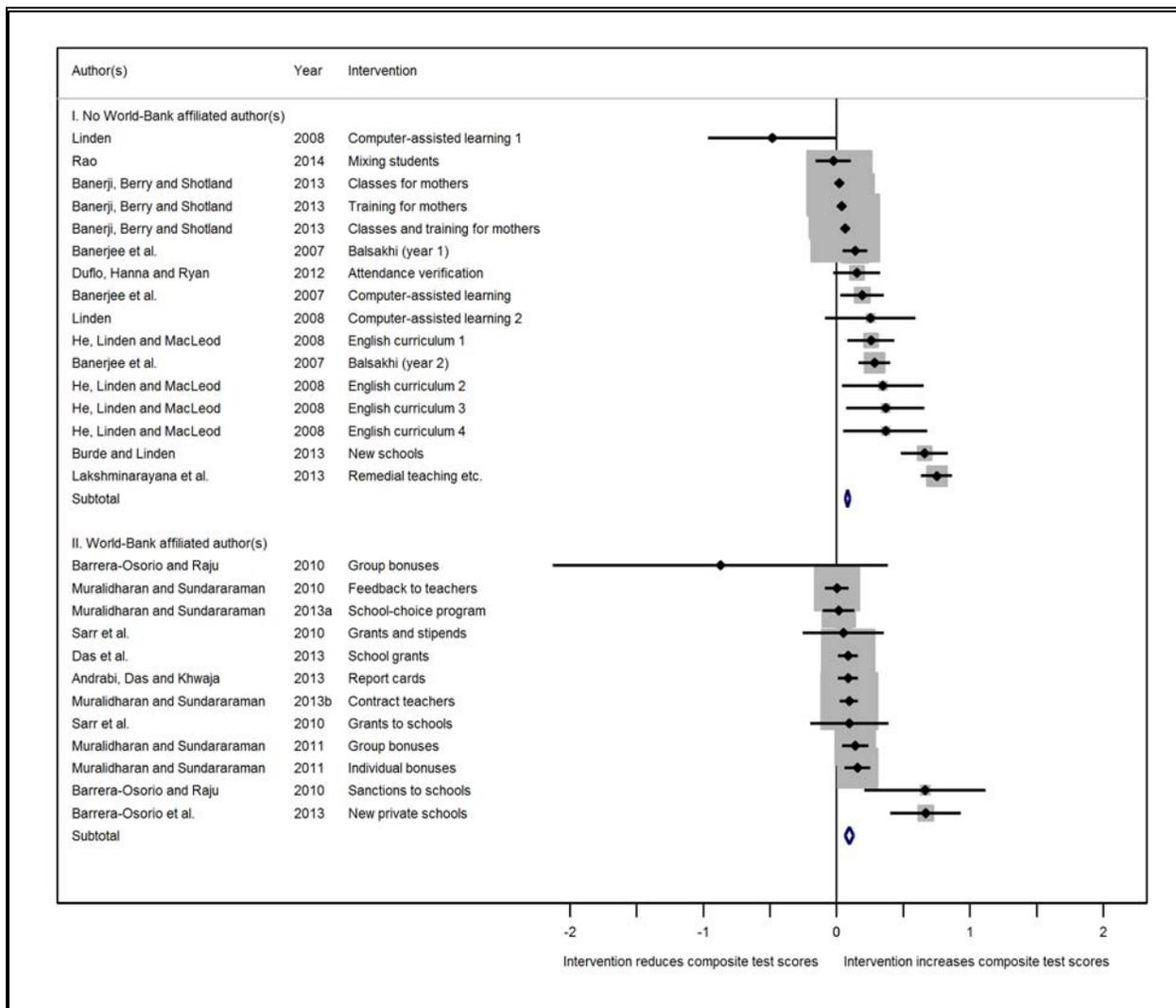
Once again, Figure 13 groups interventions according to the actor whose behavior they primarily aim to influence. But in contrast to the meta-analyses visualized in Figure 4 in section 3, it only uses evidence derived from RCTs and not from quasi-experiments. Since – as already outlined in the introduction – RCTs are generally considered to be even more rigorous than quasi-experiments, the objective is to determine whether this move to even more rigor changes any substantial results or conclusions. Of course, forgoing the evidence from quasi-experimental studies reduces the number of interventions that can be incorporated into the meta-analysis. However, the reduction in the number of interventions is relatively small: The meta-analysis visualized in Figure 13 draws on 23 interventions, i.e. only four less than the one of Figure 4. Eight of 23 interventions that utilize RCTs fall into the “teachers” category, ten belong to the “schools/school administrations” group, and four are centered on households. As was already the case for the broader set of evidence, a meta-analysis of community-focused interventions on composite learning outcomes is rendered futile because only one intervention falls into the “communities” category.

Figure 13 shows that the results from a meta-analysis that only uses evidence derived from RCTs are very similar to those derived for the full sample of rigorously evaluated education-related interventions in South Asia. As before, all three groups of interventions for which a meta-analysis of impacts on children’s average overall test scores is feasible have a statistically significant impact on this outcome variable. Moreover, at 0.17, 0.22 and 0.04 standard deviations, the effects of teacher-, school- and household-centric interventions are similar or identical to the ones identified for the sample that included both RCTs and quasi-experiments.

While before interventions were always grouped according to the actor they primarily target, this is no longer the case in Figure 14. Instead, interventions are categorized according to whether any of the author(s) of the corresponding study indicated being affiliated with the World Bank at the time of the study’s publication. The objective of this exercise is to determine whether impact evaluations with strong World Bank involvement are more or less likely to find positive impacts of education-related interventions. If interventions where at least one author was affiliated with the World Bank at the time of publication were more likely to find positive learning impacts, this could be a sign that the World Bank was more successful in identifying promising interventions than other institutions. Alternatively, it could also be interpreted as a worrying symptom of the internal World Bank political economy. Impact evaluations performed by World Bank staff are often tied to operational projects and, allegedly, there is pressure on staff to demonstrate that these operational projects are producing meaningful results on the ground.

As Figure 14 demonstrates, it is futile to speculate what could explain differing results between impact evaluations with and without a strong World Bank engagement. This is because overall impacts on composite test scores from both groups of interventions are extremely similar: The meta-analysis reveals that the twelve interventions that benefitted from the involvement of authors affiliated with the World Bank increased composite test scores by 0.10 standard deviations while the 16 interventions without direct involvement of World Bank staff raised these test scores by almost identical 0.08 standard deviations. The two impacts are not statistically significantly different.

Figure 14 – Meta-Analysis of Interventions’ Impacts on Composite Test Scores (by Affiliation)



**Notes:** Learning impacts measured in standard deviations; for individual interventions, solid diamonds indicate point estimates, black lines 95 percent confidence intervals and grey rectangles estimates’ weights in meta-analysis; for subtotals, centers of transparent diamonds indicate point estimates and the spreads between diamonds’ left and right edges 95 percent confidence intervals; the sample includes both RCTs and quasi-experiments.

**Source:** World Bank staff calculations based on studies listed in Table A.