ECONOMIC DEVELOPMENT & INSTITUTIONS

EDI WORKING PAPER SERIES WP16/8.11

EDUCATION, INSTITUTIONS AND ECONOMIC DEVELOPMENT

Tessa Bold and Jakob Svensson
IIES Stockholm University

August 2016

Oxford Policy Management · UKaid from the British people · Université de Namur · Paris School of Economics – École d'économie de Paris · ADE Analysis for Economic Decisions

Abstract

A large share of children in low income countries learn little and complete their primary education lacking even basic reading, writing and arithmetic skills. We review the experimental literature on teacher effort, knowledge, and skills – areas we argue are crucial to raise the quality of education in low income countries – and show that little is still known about effective ways to address the huge shortcomings observed in these three dimensions of teacher quality. We further argue that attention to local and national implementation constraints, several them of institutional character, are urgently called for to complement the growing experimental literature and identify core questions and issues for future research.

Acknowledgements: We are grateful for comments from participants at the EDI workshops in Namur and Paris.

About Economic Development & Institutions
Institutions matter for growth and inclusive development. But despite increasing awareness of the importance of institutions on economic outcomes, there is little evidence on how positive institutional change can be achieved. The Economic Development and Institutions – EDI – research programme aims to fill this knowledge gap by working with some of the finest economic thinkers and social scientists across the globe.

The programme was launched in 2015 and will run until 2021. It is made up of four parallel research activities: path-finding papers, institutional diagnostic, coordinated randomised control trials, and case studies. The programme is funded with UK aid from the UK government. For more information see http://edi.opml.co.uk.

1. Introduction

An educated work force is necessary for long-run economic growth and a high standard of living. Increasing the human capital of the poor may also be one of the most effective ways to reduce absolute poverty and an increase economic mobility.

Over the last 25 years, school enrollment, at all levels, have increased universally and most children in developing countries now complete primary school. For example, the primary enrollment rate in South Asia and Sub-Saharan Africa was 51 and 61, respectively in 1970 and is in both cases over 100 in 2010 (see Table 1). However, enrolling in school does not guarantee that children acquire the human capital they are expected to do. In fact, a growing literature has documented that a large share of children in low income countries learn little and complete their primary education lacking even basic reading, writing and arithmetic skills. For example, across Grade 6 students in 15 Southern and Eastern African countries tested as part of the SAQMEC assessment in 2007, less than 50% got beyond the level of "reading for meaning" and less than 40% got beyond "basic numeracy". Among 6th grade students tested in 10 Francophone countries in Western and Central Africa in 2015 as part of the PASEC assessment, less than 45% surpass the "sufficient" competency level in reading or mathematics (Hungi et al. 2010). Test result from (rural) India in 2013 show, for example, that 38% of children in third grade in public schools could not read simple words, and less than 27% could master double digit subtraction (ASER, 2013). Thus, in many low income countries, the increase in school enrollment has not resulted in a general increase in human capital.

In response to these poor learning outcomes, there has in the last 5-10 years been a push, both by economists and education researchers, to identify ways to improve the quality of education. To date, there are more than 150 high quality impact studies from developing countries on various ways to raise test scores and at least eight systematic reviews, based on the experimental and quasi-experimental research, synthesizing and drawing conclusions about the best ways to improve education quality. However, as is clear from these reviews, there is no clear consensus among researchers about what are the most effective ways to increase the quality of primary education.

The starting point in this review is different. While the systematic reviews are important, there is also clear limitations of what one can learn from them – given variation in context, variation in the details of the design, and the fact that most interventions, by design, include various components. Maybe even more problematic, given the decentralized way research is organized,

and rightly so, and the interdependence of various factors determining learning outcomes, and the lack of more systematic inside-the-box information and stylized facts about primary schooling in low income countries, it is difficult to assess whether the interventions we have the most evidence from really are the interventions one ought to focus on. For these reason, this reviews starts of from recent work by Bold et al. (2016), which use data from nationally representative surveys from seven Sub-Saharan African countries, representing close to 40% of the region's total population, to assess what primary school teachers in Sub-Saharan Africa know and do. This assessment provides an empirical context through which the growing experimental and quasi-experimental literature on education in low-income countries can be interpreted and understood. It also points to important gaps in our knowledge, with implications for future research and policy.

Question about government implementation and scaling up are also largely missing from the existing reviews. That is, the reviews draw heavily from studies that largely study demand and supply factors at the individual or school level, and sometimes under partly controlled circumstances with well-functioning NGOs and/or research team playing an important role in implementing, monitoring and supervising the intervention. While these studies are crucial for identifying potentially scalable cost-effective interventions, and for estimating key mechanisms at play, when moving from impact estimates to policy guidance it is crucial to also take into account various implementation constraints. We focus on two such implementation constraints here.

First, many interventions implemented at the school level hinge on the existence of a third party that can monitor the program. But in many low-income countries, those (public) agencies/actors assigned to monitor education programs are weak and possibly corrupt. We label these type of constraints "local rent-seeking constraint" and discuss both attempts to study them and ways to relax them. We also briefly discuss the role of voice versus exit. In settings in which public institutions do not work well, an argument could be made to facilitate low-income families' exit options; i.e., enroll their children in private schools.

Second, national scale up may induce political economy reactions from groups whose rents are threatened by the new policy. We label these as national rent-seeking constraints and again discuss attempts to study them and speculate on ways to relax them. Importantly, we argue and show that an experimental approach can also be used to study both of these implementation constraints.

Overall, our focus here is on primary education in low income countries, with a specific focus on Sub-Saharan Africa.

We proceed by first providing a brief review of the recent reviews of the experimental and quasi-experimental literature on the quality of education. We then briefly discuss recent evidence of student learning. Thereafter we summarize the main findings in Bold et al (2016) on what teachers know and do; i.e. we provide an empirical context through which the literature on improving learning outcomes could be interpreted. The stylized facts we present, we argue, suggest that teacher effort, knowledge, and skills are crucial binding constraints in low-income countries, and especially so in Sub-Saharan Africa, and we therefore briefly discuss the available evidence on attempts to address them. We thereafter turn to the issues of scaling up and the politics of education and discuss two crucial implementation constraints (local and national rent seeking constraints). We end by outlining research questions and areas that need more workn.

2. A short review of the reviews

There is no shortage of reviews about the best ways to raise student test scores. In fact, over the last few years at number of such reviews have been published. Kremer, Brannen, and Glennerster (2013) – a narrative review of a set of randomized controlled trials (RCTs) – argue that interventions that match teaching to student learning levels, contract teachers, and interventions improving access to schooling are the most effective. Krishnaratne, White, and Carpenter (2013) – a meta-analysis of 69 RCT and quasi-experimental studies – argue that the most compelling evidence of what works is computer-assisted learning tools, a result closely in line with McEwan's (2014) meta-analysis of 77 RCT studies. Murnane and Ganimian (2014), on the other hand, in a narrative review drawing on 115 RCT and quasi-experimental studies, conclude that the strongest evidence (unconditionally) of impact comes from studies providing information about school quality and returns to schooling. Conn (2014) – a meta-analysis based on 56 studies conducted in Sub-Saharan Africa – finds that pedagogical interventions (changes in instructional techniques) have the highest effect size on achievement outcomes. Glewwe and Muralidharan (2015) – employing a voting counting approach based on 118 studies (of which 80 RCTs) – conclude that

teaching at the right level (remedial programs), and teacher performance and accountability interventions are the most promising.[1]

Overall, the reviews show that there is by now a substantial number of high quality, micro based, studies on the effects of various demand and supply interventions. Several of these interventions have been implemented in different contexts, and with slight variations in the design, but still yielded similar impact estimates. Still, based on the reviews, no clear consensus exists, based on the micro estimates only, about the most effective ways to increase student learning.

3. The quality of primary education – with a focus on Sub Saharan Africa

What do we really know about quality (learning) in primary education in Sub-Saharan Africa? In the last decade, the International Association for the Evaluation of Education Achievement (IEA) and OECD testing programs have expanded dramatically, with more than 100 participating countries in at least one of these assessment in 2012 (Hanusheck and Woessmann, 2015). Low income countries, and countries in Sub-Saharan Africa in particularly, are however still largely not included in these international assessments. For example, only one Sub-Saharan African country (Botswana) participated in the last IEA mathematics tests at the primary level and only three countries (Botswana, Ghana, and South Africa) participated at secondary level.[2] As illustrated in Table 2, test results at the secondary level for these three countries suggest *average* test scores below the lowest *5th percentile* score in the US. But even those low scores are likely biased upward since in all three countries a significant share of students performed worse than chance (based on multiple choice items), implying that the tests are unable to provide a reliable measurement of average student achievement (TIMSS, 2011).

There exists a number of regional testing programs in Africa, including SAQMEC (tests students in 15 Southern and Eastern African countries), PASEC (tests students in Francophone countries in Western and Central Africa), and Uwezo (tests children in Kenya, Tanzania and Uganda). The results from these assessments also paint a bleak picture. For example, of the grade

[1] Evans and Popova (2015) provide a review of these reviews to assess which findings are consistent, and which divergent, across these studies.

[2] The Trends in International Mathematics and Science Study (TIMMS) by the IEA test grade 4 (primary) and grade 8 (secondary) students on common elements of primary and secondary school curricula. In Botswana, grade 6 students were assessed as it was deemed too difficult for fourth grade students to take the TIMSS fourth grade assessment. Similarly, in Botswana and South Africa, ninth grade students were assessed with the TIMSS eight grade assessment. No Sub-Saharan African country participate in the testing program run by OECD (Programme for International Student Assessment, PISA).

6 students tested as part of the SAQMEC assessment in 2007, less than 50% got beyond the level of "reading for meaning" and less than 40% got beyond "basic numeracy". Among 6th grade students tested in 10 Francophone countries in Western and Central Africa in 2015 as part of the PASEC assessment, less than 45% surpass the "sufficient" competency level in reading or mathematics (Hungi et al. 2010). Bold et al. (2016) present evidence from the Service Delivery Indicator program, which to date have collected data from nationally representative surveys of primary school from seven Sub-Saharan African countries, representing close to 40% of the region's total population. Fourth grade students in sampled schools were assessed in basic reading, writing, and arithmetic skills, the "three Rs", based on a review of primary curriculum materials from 13 African countries.

After three completed years of primary public schooling, almost half of the students assessed could not read a simple word.[3] 70 percent of the students could not read all the words in a basic sentence. And only one in seven students could read a simple paragraph and infer meaning from it.

In mathematics, two out of ten students could not recognize numbers and five out of ten students could not order numbers between 0-100?. 20 percent could not do single-digit addition, and three in four could not do single-digit subtraction. Just over half of the students assessed managed double digit addition, and less than a third mastered double digit subtraction – tasks that according to the official curriculum students should master at the end of grade 3.

While a significant share of children in primary schools in Sub-Saharan Africa do not learn much in school, Sub-Saharan Africa is not a specific case. Test result from (rural) India in 2013 show, for example, that 38% of children in third grade in public schools could not read simple words, and less than 27% could master double digit subtraction (ASER, 2013).

While the reasons behind this policy failure is multifaceted, evidence from recent research on education in developing countries, showing that traditional educational inputs have little impact on test scores but increasing teacher effort and pedagogy do, suggest that teacher behavior and teacher ability play a crucial role. We turn to this next, again drawing on the work by Bold et al (2016).

4. What Do Teachers Know and Do?

[3] Depending on country, the language was English, French, or Portuguese.

Bold et al (2016), using data collected through direct observations, unannounced visits, and tests, provide quantitative answers to three questions: How much time do teachers actually spend teaching? Do teachers have the relevant knowledge to teach both basic and higher-order language and mathematics skills? Do teachers have the pedagogical knowledge and skills to transfer what they know to students in an effective way?

On the first question, Bold et al (2016), show that students receive about two and a half hours of teaching a day—or less than half the scheduled time – largely because teachers, even when in school, do not teach. Absence from school (about 2 out of 10 teachers on average) turns out to be about as common as being present in the school but absent from class, implying that only 6 out of 10 teachers – on a regular day – are actually teaching. Absenteeism rates also appear to be remarkably stable over time in countries. Chaudhury et al (2006) estimated a school absence rate of 27% in Uganda in 2002-03. Bold et al (2016) find that 30% of teachers are absent from school, using data collected in 2013.

For teachers to be effective, they must have the knowledge necessary for good teaching. In particular, they must know the subject they teach (subject content knowledge), how to translate this knowledge into meaningful teaching (general and pedagogy content knowledge) and how children learn (knowledge of the context of learning). Subject content knowledge was assessed by given synthetic pieces of children's work in the language of instruction (English/French/Portuguese) and in Mathematics, based (mainly) on the curriculum the teacher was teaching, and Bold et al (2016) present result for two outcome measures: whether teachers master their students' curriculum in Language and whether they have minimum subject (in Language and mathematics) content knowledge.[4]

Bold et al. (2016), show, even using the very low bar of mastering their students' curriculum, that 40% of primary school teachers are not as knowledgeable as their students should be.[5] Looking at the score for "minimum knowledge", only one in ten language teachers in Sub-Saharan Africa passes the bar.[6] Moreover, subject knowledge among language teachers is also uniformly low

[4] In contrast to other approaches where teachers sit exams, this method of assessment aimed to assess teachers in a way that was consistent with their normal activities—namely marking student work – and to value them as professionals and not undermine their self-esteem. See Bold et al (2016) for details.
[5] There is wide variation across countries in this score. See Bold et al (2016) for details.
[6] Having knowledge equivalent to the fourth grade curriculum is of course not sufficient to teach language in lower primary, because language teaching is monolithic. Bold et al (2016) define minimum subject content knowledge for Language teachers if he or she can confidently correct children's work in such aspects of literacy as reading comprehension, vocabulary and formal correctness (grammar, spelling, syntax and punctuation), all of which are

across the seven countries. To illustrate what these findings mean, it is informative to disaggregate the scores by task. Bold et al (2016) show, for example, that one in five teachers could not spell a simple word (like "traffic" for example).

In mathematics, minimum subject content knowledge was defined as being able to accurately correct children's work in such aspects of numeracy as manipulating numbers using basic operations and solving simple math story problems. This requirement amounts to scoring correctly on 80% of the questions on the lower primary portion of the mathematics test. Bold et al (2016) show that less than two thirds of the mathematics teachers have minimum knowledge according to this definition. Again, the disaggregated scores give an informative sense of teachers' knowledge. 23 percent of mathematics teacher could not subtract double digits and 30% could not multiply double digits. When it comes to understanding and solving a simple math story problem, about half of the teachers managed to do so. Bold et al (2016) further show that when looking at more advanced tasks, such as interpreting information in a Venn diagram and/or a graph, only a minority of mathematics teachers, and in some countries almost none, could master them.

Knowing one's subject and curriculum is a necessary, but not sufficient, condition for good teaching. Teachers must also know how to translate their subject knowledge into effective pedagogy. Bold et al (2016) provide information on two scores – teachers' knowledge of pedagogy and the context of learning – and in both cases the results are depressing. Only 10% of teachers reached the threshold for having minimum general pedagogy knowledge and essentially no teachers were knowledgeable about the context of learning.[7]

Combining their measures of teacher effort, knowledge, and skills, Bold et al (2016) estimate that very few primary public school students in the seven countries they surveyed attend schools where teachers spend a significant share of the scheduled time teaching a subject they master to teach.

A growing literature have documenting the large impact for students, both in the short and longer run, of being taught by high quality teachers. For example, Chetty, Friedman, Rockoff (2014) estimate, using data for more than one million US children, that replacing a teacher whose

competencies a teacher in lower primary would routinely be required to have. To this end, the language test contained (in addition to the spelling and grammar exercises) Cloze passages to assess vocabulary and reading comprehension, and a letter written to a friend describing their school, which the teacher had to mark and correct. Minimum knowledge in language measured as marking at least 80% of the items on the language test correctly.

[7] For details see Bold et al (2016). They also examine to what extent teachers apply what they know in the classroom.

measured quality is in the bottom 5% with an average teacher would increase the present value of students' lifetime income by approximately $250,000 per classroom. Similarly, the results presented in Bau and Das (2016), using data from the province of Punjab, Pakistan, strongly suggest that teacher quality matters as much or more for student in low income countries. Thus, teacher quality appears to be a crucial factor in accounting for student learning and the Service Delivery Indicator data shows that teacher quality is strikingly low in many low income countries.

We next turn to the growing experimental literature in education to review what we know about effective ways to raise teacher effort, knowledge, and skills.

5. Teacher effort, knowledge, and skills: What's the evidence?

5.1. Teacher effort

On teacher effort, knowledge, and skills, economists have, by far, so far, focused on interventions aimed at incentivizing teachers to exert higher effort. Following Glewwe and Muralidharan (2015), we can group these interventions into four sets: (a) Performance pay; i.e., linking payments to attendance or test score; (b) Community monitoring; i.e., bottom-up approaches intended to increase monitoring by parents and the local community; (c) Strengthening existing institutions, such as school committees or PTAs that to play a monitoring and disciplining role; (d) Contract teachers; i.e., linking tenure to performance.

Performance pay

The results from studies assessing performance pay is overall positive. Duflo, Hanna, and Ryan (2012) assess and intervention in which teacher attendance in informal schools in Rajasthan, India, was monitored using cameras and teachers were paid as a function of the number of valid days of attendance. They document a large reduction in teacher absence rates (from 42% to 21%), and a smaller, but significant increase in student test scores.

Muralidharan and Sundararaman (2011) evaluate a bonus payment program to teachers (in India) based on the average improvement of their students' test scores. They do not observe any reduction in teacher absence from school (and do not measure absenteeism from classrooms conditional on being in school). However, they too document a significant increase in student test scores, including in subjects that were not "incentivized".

Glewwe, Ilias, and Kremer (2010), on the other hand, who also evaluate a bonus payment program (in Kenya), but where bonuses were paid to schools based on best performance, show that students scored higher on the exams linked to teacher incentives but not on other exams. That is, consistent with Holmström and Milgrom's (1991) multi-tasking model, teacher appeared to have reallocated effort way from uncompensated, but potentially important, activities toward compensated activities ("teaching to the test").

Community monitoring

The results from studies evaluating community monitoring as a way to strengthen local accountability relationships are mixed. Banerjee et al. (2010) study different approaches to enhance community involvement in primary schools in Uttar Pradesh, India. They find no impact on community participation, teacher effort, or learning outcomes.

Pandey, Goyal, and Sundararaman (2009) evaluate an information campaign to improve parental participation in village education committees (in India). The find positive impacts on various process measures but in general mixed evidence on learning outcomes.

Strengthening local institutions

A related set of interventions focus on strengthening already existing local institutions, with the objective of getting these locally grounded institutions to become more active as monitors of local service provision.

Pradhan et al. (2014) assess a series of interventions with the focus to enhance community participation in school management (in Indonesia). In one of their intervention arms, which involved democratic elections of school committee members and linking the school committee to the village council through facilitated meetings, they document an increase in language test scores, but no effects in mathematics.

Lassibille et al. (2010) and Glewwe and Maiga (2011) evaluate the AGEMAD program in Madagascar – a large scale program aimed at strengthening school management at the district, sub-district, school and teacher levels. Neither study found an impact on student test score.

On the other hand, Duflo, Dupas, and Kremer (2015) study of a contract teacher program in which, in one treatment arm, school-based management (SBM) training were provided to the PTA committee and, importantly, the decision whether to renew the contract teachers' contracts were delegated to the PTA, showed better performance of contract teachers and significant improvements students' test scores.

Contract teachers

Another set of studies focus on providing incentives to teachers (specifically newly hired teachers) by making tenure decision conditional on performance.

Duflo, Dupas, and Kremer (2015) design their experiment (in Kenya) in a way to separate out the direct effect of incentives from the effect of class size reduction; i.e. students can benefit of two effects from hiring a new contract teacher with a contract where tenure is conditional on performance. First, provided the incentives bind, students taught by the contract teachers may benefit from a teacher exerting higher effort. Second, as the number of teachers teaching increase in response to the hire, students may benefit from a lower student-teacher ratio. The authors find that holding class size constant, students taught by contract teachers scored higher on tests than those taught by civil-service teachers, even though the contract teachers are paid much lower salaries.

Bold et al. (2013) replicate the Duflo et al (2015) study across different parts of Kenya. They find positive and significant effects of the contract teacher program, across diverse baseline conditions, when the contract teacher program was administered by an international NGO.

Muralidharan and Sundararaman (2013) assess the impact of an extra contract teacher in government-run schools (in India). They show that contract teachers were less likely to be absent from school than civil-service teachers and students in schools with an extra contract teacher performed significantly better in both math and language tests, respectively.

Sum-up: Teacher effort

There is strong evidence that teacher effort can be raised, leading to improved learning outcomes, by linking payments or tenure to performance. Softer incentives, through strengthening and encouraging existing local institutions or the community more generally to become more active as monitors of teachers, appear to be less effective. However, while incentivizing effort results in higher effort exerted by the teacher, and increased test score of students, it is important to stress that incentivizing effort alone will likely not be enough to substantially increase student learning outcomes, given teachers' limited knowledge and skills as briefly presented above. That is, there is an obvious upper bar on the effect of exerting more effort to teach when one cannot master the subject one teaches. This is in fact indirectly evident from the studies themselves: Given the large gaps in student knowledge, the estimated effect sizes are relatively small.

5.2. Teacher knowledge

While there by now are a numerous studies on ways to elicit more effort of teachers, the evidence on cost effective ways to increase content knowledge of teachers, and the impact thereof, are largely missing. In fact, there is no high quality study on the impact on student learning of raising the content knowledge of teachers and research on this topic is urgently called for.

5.3 Teacher skills

A smaller number of studies have evaluated intervention that at least broadly could be interpreted as dealing with primary school pedagogy. One can divide these studies into three groups:[8] (i) Remedial education programs, where new instructors are trained to provide basic education to children lagging behind; (ii) Tracking of students, where students with similar observable skills are grouped together in order to make it easier for teachers to more effectively match the difficulty level they teach to the level of their students; (iii) Literary programs, in which new or existing teachers are trained to provide a pre-specified literary programs.

Before reviewing the evidence on these three types of interventions it is important to note that all of these interventions involve more changes at the school or student level than changing the pedagogy. Of course, in itself this should not be viewed as a shortcoming – on the contrary there are very strong reasons to believe there are important complementarities between school inputs, teacher effort, knowledge, and skills, and too little attention by researcher has been put on on quantifying and understanding these complementarities – but it does raise the question of exactly what combination of changes are driving the findings.

Remedial education programs

A smaller set of randomized trials have assessed the impact of remedial education programs, all finding positive, and typically large, effects on student learning. Banerjee et al. (2007) evaluate a remedial education program in the Indian cities of Mumbai and Vadodara, where an informal teacher taught children lagging behind two hours per day. They show that the program resulted in improved students' test scores in both math and English and argue that the effect was driven by the fact that students being pulled out of the classroom (to receive remedial education) were taught at a level corresponding to their current proficiency.

[8] There are also a number of recent evaluations of so-called a computer-assisted learning programs. We do not review them here as the programs differ too much in design and implementation details to be categorize as a family of similar interventions.

Banerjee et al (2010) assess a remedial instruction program, also in India, where in reading camps were run by trained youth volunteers. The program resulted in a significant and relatively large increase in literacy.

Lakshminarayana et al. (2013) evaluate a remedial education program in a set of villages in villages in Andhra Pradesh in which volunteers provided two hours of remedial instruction per day after normal school hours. They too document large and significant increases on language test scores.

Tracking

While a number of studies have evaluated different versions of remedial education programs, much less is known about the impact of tracking students. Duflo, Dupas, and Kremer (2011) is an exception. They assess an intervention in Kenya in which students where tracked and streamed based on initial test score. The intervention resulted in increased student test scores that remained (at least) one year after the program ended.

Literary programs

The impact of various literacy programs have been assessed in several studies, with somewhat varying results. Abeberese, Kumler and Linden (2014) evaluate the short run impact of a reading program for grade 4 students in Philippines where teachers were trained and provided with age appropriate reading materials – materials that were intended to be incorporated into their teaching. The program had significant but relatively small effects on reading scores after four months.

Lucas, McEwan, Ngware, and Oketch (2014) assess a reading instruction intervention implemented in a set of schools in both Kenya and Uganda. The program involved several components, including training of teachers, regular support to teachers, and provision of new material. The evaluation showed that the reading intervention resulted in significant positive effects in Uganda, but smaller and less precise effects in Kenya.

Piper and Korda (2011) evaluate a literacy instruction program in Liberia. As in Lucas et al (2014), the intervention involved several components. Specifically, in the full treatment group, reading levels were assessed; teachers were trained on how to continually assess student performance; teachers were provided frequent school-based pedagogic support, resource materials, and books; and, in addition, parents and communities were informed of student performance. After 18 months, and starting from a low level, the program dramatically accelerated children's learning.

Sum-up: Teacher skills (Pedagogy)

There is strong evidence that very basic pedagogical interventions, including providing simple teaching at the right level; i.e., remedial education programs, and providing teachers with hands-on, focused training on how to teach scripted lessons effectively, improve student outcomes starting from very low levels.

However, it is important to note that these programs, just like some of the successful computer-assisted learning programs, are to some extent programs intended to "automate" the teacher (to follow rather simple tasks). Given the low level of knowledge and skills of teachers, as discussed in section 4, such interventions may at least in the short run be an effective method to raise learning outcomes, especially at early grades and for lower-order learning. But it is also clear that there are obvious limits to what these kinds of program can achieve. More generally, there is still scant evidence on how to effectively improve general pedagogy at primary school in low income countries. For example, we know little about how to effectively teach teachers how to teach "higher-order" tasks. We know little of how to improve pedagogy along the dimensions uncovered in the SDI, including having clearly structured lessons; having an appropriate mix of basic and higher-order tasks; being able to write lesson plans; and being able to assess students' progress and give feedback.

6. From experimental studies to scalable policies

In the best case scenario, with high level of governmental capacity and commitment, the evidence reviewed above could have a direct bearing on policy design and policy reform in the education sector. Such reforms would likely involve reform of the curriculum (including defining learning standards in core subjects for every grade level and curricula well-aligned with the learning standards); introduce process for assessing student mastery of the standards; develop teacher training programs that attract talented students and prepare them to teach the curriculum effectively; and put in place effective incentives schemes that ensures high effort and continued upgrading of knowledge and skills.

However, in most low income settings, the problems identified at the school level – low efforts, capacity, skills – characterize the whole service delivery chain, making the transition from the identification of cost effective interventions to evidence based policy making a complicated one.

Assembling information on the cost-effectiveness of various potential reforms/interventions provides a crucial first step in a strategy to develop an evidence based policy program, but in general, and especially in countries characterized by weak institutions and a capacity-constrained public sector, that evidence only is not enough to guide policy. Thus, while many of these studies differ starkly from so-called proof-of-concept studies in medicine and public health, where the participants' behavior is partly controlled, they still provide estimates of effectiveness that may significantly depart from what can be achieved when going to scale and/or when moving to government implementation.

When identifying, choosing, and designing programs to be scaled-up, and especially when the scaled-up version involves government implementation, it is important to have an understanding of the whole delivery chain; from the institutional constraints that affect central government policy decisions, through the incentive constraints that influence different layers of government agencies and officials implementing a given policy, to the actions and incentives of teachers and students (and parents).

One way to conceptualize the problem is to think of implementation design as identifying the most cost effective interventions subject to two rent-seeking constraints. First, local actors – be it teachers, local public officials or supervisors – may try to extract rent from the program which, if they are successful, lowers the program's impact. Second, large scale implementation may provoke political economy reactions because rents of powerful insiders are threaten. Such an endogenous policy response – a "seesaw effect" using the terminology of Acemoglu (2010) – may counteract and even reverse the objectives of reform. We can think of the first implementation constraint as a "local rent-seeking constraints" and the second as a "national rent-seeking constraints". We turn to a discussion of the research on them next.

6.1. Local rent-seeking constraints

Social service delivery in many developed countries grew out of a system of locally-governed institutions. Importantly, these locally-governed institutions, some of them nowadays formalized into the local institutional framework that is intended to ensure high quality and adequate services, still play an important role. In many low income countries, on the other hand, the adoption of centralized state control over service delivery at independence, coupled with a weak central state, have resulted in a system characterized by centralized, but weak, state control and often low-capacity locally-governed institutions for education provision. The education system is

characterized by extensive leakage and inefficiencies, where parents have little influence on how schools are managed. Any intervention intended to be sustainably operated in such an institutional environment needs to carefully be designed to take these institutional constraints into account.

The research on how community participation and civil society organizations could be strengthened and the impact thereof take its starting point in this institutional failure to deliver high quality services. In short, as beneficiaries (parents) ought to have intrinsic motivations to ensure that children learn in school, local control is viewed as a way to overcome the "monitoring the monitor problem". The focus on civil society and community participation is also important since historically, civil society has played an important role in the advancement of institutional/political reforms in many countries. Moreover, community participation interventions, primarily meant to (informally) influence norms and collective actions and thereby local decision making, can be viewed as an attempt to enhance common interests, albeit starting at a small scale, and the strength of common interest in society, together with the structure of political institutions, has recently been highlighted in the macro-political literature as key underlying determinants for policy decisions in sectors like education and health (see for instance Besley and Persson, 2011).

In policy discussions, and at an abstract level, service delivery is often discussed as a principal-agent problem where the government is the principal that sets policy and delegates implementation and monitoring to various lower level public actors. Implicit in this approach is that lower level public actors will mechanically execute what they are ask to do (and do so in a way that benefit the intended beneficiaries). This, however, contrast starkly with the way in which policy implementation is carried out in practice. In practice, policy implementation and monitoring are delegated to agents that themselves are not monitored, resulting in a broken accountability link (the who's monitoring the monitors problem).

The micro-political research on service delivery, again at an abstract level, basically turn the incentive problem up-side-down, viewing individual citizens as the principals and service delivery actors as agents. As the principal's ability to control the agents typically are constrained by the local institutions in place, the core objective of this research program is to identify ways to relax such constraints.

The canonical contract problem requires the principal to write sufficiently detailed contracts ex ante and the ability to measure and verify performance ex ante, and such type of contracts may simply be unfeasible in many settings. Thus a complementary approach, again at an abstract level,

is the incomplete contract approach, where the issue is how to optimally assign decision rights to different actors, including individual citizens or communities, to influence bargaining strengths and thus influence incentives.

We next turn to discuss two incentive contract approaches to deal with local rent-seeking constraints: Local control and private provision

Local control

There is a growing literature on local accountability (local control, beneficiary control, community monitoring) with what appears, at first sight, at best, mixed findings. However, a closer investigation of the interventions that have been designed and evaluated suggests that a pattern is emerging. Specifically, interventions that yield users decision rights (formal or informal) or assist in the process of agreeing on contract specifying actions, and, importantly, address underlying informational asymmetries by either provision of relevant and actionable information or provision of tools for beneficiaries to acquire such information, do seem to results in improved service delivery performance and improved outcomes. On the contrary, interventions focusing on enhancing participation where service delivery problems are vented and discussed, and/or provide information that is not directly informative or actionable, do not appear to result in systematic changes in service delivery performance.

Specifically, Björkman and Svensson (2009) and Björkman, de Walque and Svensson (2015) analyze the impact of various ways to strengthen local accountability in primary health care in Uganda. Björkman and Svensson (2009) design an intervention (*information & participation intervention*) which combined traditional tools to enhance community participation with the dissemination of report card information on staff performance. The intervention can, again at a somewhat abstract level, be viewed as the facilitation of a contract problem; i.e., the community's (the principal's) problem of eliciting effort from the service provider (the agent). In fact, a core component of the intervention was the design of a joint contract specifying actions that need to be taken and by whom and when. In such a contract the final outcome depends on the principal's ability to identity tasks that both maps into better health service and can be influenced by the health workers' actions. It also depends on the principal's ability to observe actions and the compensation scheme.[9] When applying the framework to community monitoring one must also take into account

[9] The compensation scheme; i.e., the ability to reward and punish the agent, was not explicitly addressed in these type of program. Instead the community had to rely on informal forms of social recognition and/or social

the fact that there are multiple principals (the users) that may disagree about both objectives and the mapping between actions and outcomes. The participatory components of the intervention was intended to facilitate the process of reaching an agreement. Providing information on performance was also intended to enhance that process and but it was primarily intended to assist the principal to identify and possibly over time observe actionable tasks.

Björkman and Svensson (2009) show that a year after the intervention begun, treatment communities were more involved in monitoring the provider and the health workers appear to exert higher effort to serve the community. They further document large increases in utilization and improved health outcomes, including reduced child mortality, that compare favorably to some of the more successful community-based intervention trials reported in the medical literature.

Björkman, de Walque and Svensson (2015) assess the longer-run impact of the community driven local accountability studied in Björkman and Svensson (2009). They also examine the role of information in these kinds of community driven accountability programs. To that end an additional intervention (*participation intervention*), which replicated the participatory components of the original intervention but did not provide communities with baseline information on performance, was designed, evaluated and compared to the *information & participation intervention*.

Björkman, de Walque and Svensson (2015) present two key findings. First, the large treatment effects on staff performance and health outcomes identified in the short run (in the *information & participation intervention*) remain more than four years after the initial intervention despite minimal follow-up. Second, the impact of the cheaper *participation intervention* differed markedly from the impact of the *information & participation intervention* – both in the short and longer run. In fact, without information, the process of stimulating participation and engagement had little impact on the staff's behavior, health outcomes or the quality of health care.

To investigate why the provision of information appeared to have played such a key role, the authors use data from the implementation phases of the two interventions. A core component of both experiments was the agreement of a joint action plan, or contract, outlining the community's and the providers' agreement on what needs be done, and by whom, in order to improve health care delivery. They show that while the process of reaching an agreement looks similar on some

opprobrium. Within reasonable bounds, it is possible that non-financial rewards (social recognition) may be just as effective at eliciting effort as financial incentives.

observable measures in the two treatment groups – for example similar number of community members participated in the community meetings and, on average, the two groups identified the same number of actions to be addressed – the types of issues identified differed significantly. Specifically, and as illustrated in Figure 1, in the *participation* group the health provider and the community identified issues that primarily required third-party actions; e.g., more financial and in-kind support from upper-level authorities and NGOs. In the *information & participation* group, by contrast, the participants focused almost exclusively on local problems, which either the health workers or the users could address themselves, including absenteeism, opening hours, waiting time, and patient-clinician interactions. These results are consistent with the hypothesis that lack of information on performance makes it more difficult to identify and challenge (mis)behavior by the provider, and hence constrains the community's ability to hold providers to account. That is, with access to information, users are better able to distinguish between the actions of health workers and factors beyond their control and, as a result, turn their focus to issues that they can manage and work on locally.

The study by Duflo, Dupas, and Kremer (2015), while different in several dimensions, reinforce the conclusion above.[10] Specifically, Duflo et al. (2015) evaluate a contract teachers program in Kenya where one of their treatment arms included School-Based Management (SBM) training of the PTA committee. Committee members were then given specific training on how to monitor and assess teachers' effort and performance and a set of parents were asked to perform attendance checks on the teachers on a regular basis. A formal sub-committee of parents was formed to evaluate the contract teacher and deliver a performance report at the end of the year. The PTA was also given influence over the decision whether to retain the contract teacher or not. While this intervention did not directly provide quantitative information on performance, it provided detailed training to the community on how to measure it throughout the year and how to combine the information in a performance report that could be shared with others. Thus, in short, the intervention assigned decision rights to parents and provided them with tools to collect

[10] Additional support for the conclusion that delegating formal (as in Duflo et al, 2015) or informal (as in Björkman and Svensson, 2009) decision rights to users and addressing underlying informational asymmetries are key in successfully influence service delivery performance is provided in Barr et al. (2014). They study different ways to provide SMC members with tools to monitor teacher performance. In line with the findings in Bjorkman and Svensson (2009), they show that the provision of tools for assembling performance information, combined with an explicit participatory component, resulted in reduction in teacher absenteeism and increased student test scores.

performance information. Duflo et al. (2015) show that SBM training for PTA committees with formal decision rights reduced teachers' absenteeism and increased student test scores.

More generally, the study also highlights two additional important results. First, in the presence of weak institutions, increases in resources, as was the case here, may be undermined by the behavioral response of existing providers. This in essence is the local rent-seeking constraints. Second, and importantly, the solution to this problem is not some general decentralization reform, as devolving control of various educational aspects to districts, absent complementary governance reforms, may allow local agents of the state to use devolution to capture rents.

Private provision: Voice vs. exit

Decision rights could be delegated to a group of beneficiaries as in Duflo et al (2015). But it could also be delegated to individuals. One example of that is to provide parents with a choice of which service provider, or school more specifically, to send their children to. If funding follow the student, parents' bargaining rights are then strengthen and in principle this might provide incentives for public schools to improve service provision. More generally, in settings in which public institutions do not work well, and reforms of the system are difficult to implement, and even if public funds do not follow students, a strong argument could be made to facilitate low-income families' exit options; for example by providing vouchers to low-income households which would allow them to enroll their children in private schools.

There has been a rapid growth of private schools in low income countries in recent years, although public spending on education in many low income countries has increased. Baum et al. (2014), for example, estimate that private schools account for approximately 20% of total primary school enrollment in low-income countries. There are several reason for this private sector expansion. For example, many has argued that the huge growth in enrollment, largely as a result of removals of user-fees in public schools, has further diluted the quality of public education, resulting increased demand, especially from better off households.

The main criticism of private schooling, and public support for it, is that it results in increased economic stratification and reduced pressure of public schools to supply quality education. Whether that is true or not is likely to depend how the actual subsidization of privately provided education is organized. And even independent of that, if the idea is to subsidize poor households' ability to send their children to private schools, the concern that it will increase economic stratification is likely muted.

Of course, private school per see does not imply that the quality is high. However, one could at least make a credible argument that private schools should be more accountable and responsive to parents. Moreover, the Service Delivery Indicators data we discuss in section 4, while providing evidence of substantial shortcomings with respect to effort, knowledge, and skills also among primary teachers in private schools, still shows that in all these dimensions, teachers in private schools outperform teachers in public schools.

In the end, whether scarce funds should be spend on facilitating poor households' exit option or not is primarily an empirical question. A question for which we to date have very little high quality evidence on.

In fact, there is only one randomized control study on the impact of vouchers in primary education. Specifically, Muralidharan and Sundararaman (2015) evaluate a voucher program to assess the *individual impacts* of school choice. The find no difference between the test scores of voucher beneficiaries (or more precisely their children) in language (native language of Andhra Pradesh) and mathematics, but private schools spend significantly less instructional time on local language and math and more on English and other subjects (Hindi).

Clearly, more research is needed here. The impact of vouchers programs are likely to be context specific and depend on the quality of accessible and affordable private schools. General equilibrium effects, and especially whether, and if so to what extent, students in public sector schools are adversely affected if vouchers for private schools are provided to a significant segment of the student population are issues we know little about. Overall, and connecting back to the proof-of-principle discussion we started out this section with, whether the voucher program is done as a small separate activity or as part of an overall reform of the education sector (where funds follow students) will obviously matter greatly for the impacts one would tend to observe.

6.2. National rent-seeking constraints

Large scale policy implementation may provoke political economy reactions because rents of powerful insiders are threaten. Such a "seesaw effect" may undermine the very objective of the reform and failure to anticipate it, and adjust policy accordingly, may severely impact what in the end will be implemented.

National rent-seeking constraints, and the seesaw effect in particular, is illustrated in the work of Bold, Kimenyi, Mwabu and Sandefur (2013). This paper takes one of the criticisms of the RCT

literature head on; i.e. the issue of external validity in policy analysis, and importantly, it does so without relaxing the emphasis on identification.

Bold et al. (2013) takes its starting point in one of the more extensively tested school-level interventions to raise student learning in primary schools, as noted above – the provision of contract teachers; i.e. teachers hired on short term contracts. Compared to a civil servant teacher who is given immediate tenure and faces few pecuniary incentives to exert high effort as salaries and promotions are largely determined by seniority, a contract teachers are thought to improve learning through two channels – a selection effect (as better performing contract teachers may have a better chance to transfer into the better paid civil servant position) and an incentive effect (as poorly performing contract teachers may be fired or at least not re-hired).

One of two main questions addressed in Bold et al (2013) is whether successful NGO pilots can be replicated by governments. To answer that question they set up a randomized controlled trial to compare the impact of the contract teacher program run by the Kenyan government with one under the coordination of an international NGO. The experimental evaluation was implemented as part of the pilot phase of a nationwide program which intended to employ 18,000 contract teachers in 2010. The pilot was designed to test the Ministry of Education's ability to implement a fairly close variant of the NGO project described above (Duflo, Dupas and Kremer, 2015) and to replicate the results across diverse conditions, spanning urban slums in Nairobi and nomadic communities in the remote Northeastern province.

As part of the government's contract teacher pilot, schools were chosen from across all eight Kenyan provinces and randomly assigned to three arm: a control group, a group receiving a contract teacher as part of the government program, and a group receiving a contract teacher under the coordination of the local affiliate of an international NGO, World Vision Kenya. The timing, salary levels, recruitment procedures and all other experimental protocols were held constant across the government and NGO arms of the evaluation.

The results of the evaluation were striking. Similar to earlier studies the authors found a positive effect on student learning overall. However, the effect is entirely due to the increase in test scores in the randomly assigned sub-group of treatment schools where the program was administered by the international NGO. They find a significant, positive and relatively large effect on test scores in schools randomly assigned to NGO implementation, and zero effect in schools receiving contract teachers from the Ministry of Education.

Two crucial insights emerge from this work. First, the study confirms the hypothesis that interventions, and especially interventions which are likely to threaten the rents of insiders (in this case unionized teachers), cannot be adequately described without attention to the broader institutional setting. Second, the work illustrates how one through a RCT-based approach can identify important bureaucratic and political constraints facing, in this case, the Kenyan Ministry of Education's ability to scale up interventions that have been showed to work. That is, an experimental approach can also be used to study and assess constraints to scaling up.

So why did not the program work when implemented by the government? The evidence strongly suggest that the stark contrast in success between government implementation and NGO implementation can be traced back to the "see saw effect" identified by Acemoglu (2008) and the "implementer bias" as highlighted by Heckman (1991). Specifically, the government's ambitious plan to employ 18,000 contract teachers nationwide posed a significant threat to the Kenyan National Union of Teachers. In response, the teachers' union waged an intense political and legal battle against the contract teacher program, including a lawsuit which delayed implementation by over a year, street protests in central Nairobi, and the threat of a national strike. The political battle eventually altered the program in two key respects, and expectation that that would happen likely affected the teachers' performance in the government implementation group in the trial.

First, primary responsibility for employing contract teachers (both during the pilot and the full scale-up) was shifted away from its natural home within the government. Second, after a prolonged political battle, the Ministry of Education acquiesced to union demands to absorb the contract teachers into civil service employment at the end of their contracts, thus effectively removing the very incentives the contract teacher reform was intended to create.

7. Experimental approach to scaling-up

For several reasons, the recent experimental research in education have focused attention on demand and supply factors at the school and individual level. Research on effective ways to take these finding, and the implementation constraints we have highlighted above, to formalize effective and scalable policies, are, with few exceptions (see below) largely missing. An interesting exception is recent work by Banerjee et al (2016).

Bold et al (2013) show how an experimental approach can be used to assess and identify political constraints that have bearing on the government's ability to scale-up interventions shown

to be effective. Banerjee et al (2016) go one important step further. Specifically they study the design of a scalable remedial education program in India.

The starting point for their work is a set of proof-of-principle studies in India, reviewed in section 5, showing that learning camps that are held outside of normal school hours where trained instructors use learning-appropriate remedial materials was highly effective in raising test scores (Banerjee et al., 2007, 2010).

Based on these findings, two large experiments where launched in government schools in two districts in India. In this scaled-up version of the program, regular teachers were trained, new materials were provided, and volunteer teachers who could assist with remedial education were made available to schools. Unlike the proof-of-principle studies guiding the scaled-up version, and similar to Bold et al (2013), the authors found no effect in the government operated program, since teachers de facto did not change their behavior.

The research team did not stop there, however. Instead they used the lessons from the failure to identify key constraints, work with the bureaucracy and developed a variant of the proof-of-principle intervention, making the revised intervention an integral part of the curriculum. This new version was implemented (as RCTs) in two districts showing promising effects and the positive findings of the revised intervention in the end led to successful state wide roll-out, with 23,000 schools, 60,000 teachers, and 1.4 million children.

8. Research questions or areas that need more work

Our review of the literature and the evidence highlights important avenues for future research. The most general, but also the most important given the evidence, is that research on increasing the quality of primary education should be high on the agenda. It is hard to think of an area with larger potential impact on both short run well-being and longer run economic performance than improving the quality of education for those that cannot afford to buy high quality education in private markets. High quality public education may also be one of the most effectives ways to raise welfare of the poorer segment of the population and enhance economic mobility.

In policy discussion about education in low income countries, however, there has recently been a shift towards secondary and tertiary education. While access to and quality of higher education is clearly crucial, and has become more important over time, the notion that the large progress in primary school enrolment in developing regions should imply a shift in focus towards

higher education, at the expense of primary education, is a worrisome one, given that a large share of children in low income countries learn little and complete their primary education lacking even basic reading, writing and arithmetic skills.

The second slightly less general recommendation for future research, again based on the evidence, is that more focus needs to be put on teachers' effort, knowledge, and skills. It is hard to think of any policy reform in education that would have dramatic effects on learning outcomes, assuming teachers should still play an important role, if teachers only teach half of the scheduled time and a large share of them cannot even master their students' curriculum. There is also growing quasi-experimental research, so far primarily from the US, documenting the great values for students (in both the short and long run) of having high quality teachers (Chetty, Friedman, Rockoff, 2014). More specifically, while there is a growing evidence on promising interventions linking payments or tenure to performance, and some evidence on interventions that could be viewed as automating teachers to perform rather simple tasks, we know little about cost effective ways to teach teachers how to teach "higher-order" tasks or how to improve general pedagogy. There are also no high quality studies on how to cost effectively raise teachers' content knowledge or more general skills to teach.

Our review has also highlighted a general question that still remains under-researched. Namely how do we go from evidence from what can largely be viewed as proof-of-principle studies to scaled-up programs? We argue that this calls for research along two lines: one dealing with a what we labeled a local rent-seeking constraints – and here we argue for identifying and experimenting with interventions inspired by the incomplete contract approach; i.e., by assigning decision rights to the users and overcoming informational constraints. There is growing evidence here on various way to achieve this but research that explicitly deals with how interventions will work without relying on either monitoring and support from researchers or monitoring and support from other public actors that in many cases lack the incentive to effectively provide monitoring and supervision, ought to be high on the agenda.

We also have highlighted the role of information. Devolving authority to the actual users; i.e. strengthening beneficiary control, we argue, is in principle an effective way to deal with the monitoring-the monitors problem and various local rent-seeking constraints. However, lacking information to properly act on, devolution of authority may not be enough to constraint rent-seeking behavior by various public actors, including the providers themselves. Providing users with

actionable information, or ways to collect such information, appear to be crucial in order to mitigate the capture of rents. Here, though, more research is called for. The cost of collecting, assembling, and dissemination information may be high, so an especially promising line of research is to experiment with advances in ICT to bring both the cost down and increase users understanding of the information and willingness to act on it.

The macro political economy literature also highlights the crucial relationship between information dissemination and accountability (of politicians). Thus, systematically collecting "informative performance data" – the Service Delivery Indicators (SDI) is one example of this – and properly disseminating it can be an important complementary strategy to boost the incentives for effective reforms in education, both at the local and national level. And, as argued above, such information can be used by users and local communities to help identify problems and search for solutions.

More generally, experimenting with interventions that explicitly are designed taking implementation constraints – such as a weak public sector – into account (if/when going to scale) even though those constraints may not be particularly important when running the project on a small scale, is important. An example of that is Bold and Svensson (2016). Their design calls for evaluating an incentive program in primary education in which financial rewards, conditional on test score performance, is paid out in the form of a chance to win; i.e., taking part of a lottery.[11] From an implementation point of view, lottery incentives, compared to an incentive program that offered each school a certain payment conditional on performance, have one major advantage. Specifically, under certain conditions, while the chance to win offers incentives for all, only the winner's performance needs to be assessed and only the winner needs to be paid.

We have also argued and provided examples of how RCTs can be used to measure and identify constraints to government scale-up and how a research program relying on a series of RCTs each building on each other can be used to identify scalable proof-of-principle interventions. Both approaches, and especially the latter, are promising ways forward if the aim is to identify evidence-based policies for educational reforms in low income countries.

Finally, when public institutions do not work well, an argument could be made to facilitate low-income families' exit options; i.e. to subsidize the enrollment of their children in private

[11] For an evaluation of a lottery incentive program for HIV prevention, see Björkman, Corno, de Walque, and Svensson (2015).

schools. The encouragement and support of privately provided education is often criticized for leading to increased economic stratification and reduced pressure of public schools to supply quality education. However, increased economic stratification is likely not a problem if the idea is to subsidize private education of the poor. There is to date little high quality evidence of voucher programs, and to our knowledge none from Sub Saharan Africa. Again, this seems to us an important area for future research.

References

Abeberese,A.B., T.J. Kumler and L.L. Linden, (2014), Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines" *Journal of Human Resources*, 49(3):611-633

Bau, N., and J. Das, (2016), The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers, Working Paper, The World Bank.

Banerjee, A., R. Banerji, E. Duflo, R. Glennerster and S. Khemani, (2010), Pitfalls of participatory programs: evidence from a randomized evaluation in education in India" *American Economic Journal: Economic Policy* 2(1):1-30.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, S. Mukherji, M. Shotland, M. Walton, (2016), Teaching at the Right Level: Evidence from 3 Randomized Evaluations in India, Working Paper, J-PAL.

Banerjee, A., S. Cole, E. Duflo, and L. Linden, (2007), Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 1235-1264.

Björkman, M., D. de Walque, and J. Svensson, (2015), Experimental Evidence on the Long-Run Impact of Community Based Monitoring, *American Economic Journal: Applied* (forthcoming)

Björkman, M. and J. Svensson, (2009), Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda, *Quarterly Journal of Economics* 124:2.

Bold, T., D. Filmer, G. Martin, E. Molina, B. Stacy, C. Rockmore, J. Svensson, and W. Wane, (2016), Getting Inside the Black-box of Schooling: What Do Teachers Know and Do? A report card from African Primary Schools, Working Paper, The World Bank.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a and J. Sandefur, (2013), Interventions and Institutions: Experimental Evidence on Scaling Up Education Reforms in Kenya, Working Paper, IIES, Stockholm University.

Chetty, R., J.N. Friedman, J.E. Rockoff, (2014), Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood, *American Economic Review*, 104 (9): 2593-2632.

Glewwe, P., N. Ilias, and M. Kremer, (2010), Teacher Incentives, *American Economic Journal: Applied Economics*, 205-227.

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

Duflo, E., P. Dupas, and M. Kremer, (2015), School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools, Journal of Public Economics, 123: 92–110

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: 1739-1774.

Duflo, E., R. Hanna, and S.P. Ryan, (2012), Incentives Work: Getting Teachers to Come to School, *American Economic Review*, 102(4): 1241–78.

Evans, D., and A. Popova, (2015), What Really Works to Improve Learning in Developing Countries: An Analysis of Divergent Findings in Systematic Reviews, Policy Research Working Paper 7203. The World Bank. Washington, DC.

Glewwe, P. and E.W.H Maiga, (2011), The impacts of school management reforms in Madagascar: do the impacts vary by teacher type?, *Journal of Development Effectiveness* 3 (4): 435-469.

Glewwe, P., and K. Muralidharan, (2015), Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, RISE-WP-15/001.

Hanushek, E.A. and L. Woessmann, (2015), *The Knowledge Capital of Nations: Education and the Economics of Growth*, MIT Press.

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

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., Cappelle, F.V. (2010). SACMEQIII Project Result: Pupil Achievement Levels in Reading and Mathematics. Working Document Number 1. Paris: SACMEQ.

Kremer, M., C. Brannen and R. Glennerster, (2013), The Challenge of Education and Learning in the Developing World, *Science* 340:297-300.

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. New Delhi.

Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D., and Mann, V, (2013), The Support to Rural India's Public Education System (STRIPES): A Cluster Randomised Controlled Trial of Supplementary Teaching, Learning Material and Material Support. *PloS one*, *8*(7), e65775

Lassibille, G., J-P. Tan, C. Jesse, and T. Nguyen, (2010), Managing for results in primary education in Madagascar: Evaluating the impact of selected workflow interventions, *World Bank Economic Review* 24(2):303-329.

Lucas, A.M., P.J. McEwan, M. Ngware and M. Oketch, (2014), Improving early grade literacy in East Africa: Experimental evidence from Kenya and Uganda, *Journal of Policy Analysis and Management*, 33(4): 950-976.

McEwan, P., (2015), Improving Learning in Primary Schools in Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*.

Muralidharan, K. and V. Sundararaman, (2011), Teacher Performance Pay: Experimental Evidence from India, *Journal of Political Economy*, 119 (1): 39-77

Muralidharan, K. and V. Sundararaman, (2013), Contract Teachers: Experimental evidence from India. JPAL Working Paper.

Murnane, R., and A. Ganimian, (2014), Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations". NBER Working Paper no. 20284.

Pandey, P., S. Goyal and V. Sundararaman, (2009), Community participation in public schools: impact of information campaigns in three Indian states, *Education Economics*, 17(3): 355-375.

Piper, B. and M. Korda, (2011), EGRA Plus: Liberia: Program Evaluation Report, Report prepared for USAID/Liberia.

Pradhan, M., D. Suryadarma, A. Beatty, M. Wong, A. Gaduh, A. Alisjahbana, and R.P. Artha, (2014), Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia, *American Economic Journal: Applied Economics* 6(2): 105-26.

Table 1: Primary enrollment rates

	Primary enrollment 1970	Primary enrollment 2010	Primary completion rate 2010
OECD	104	103	99
East Asia &Pacific	95	112	99
East Europe & Central Asia	105	103	98
Latin America & Caribbean	102	107	98
Middle East & North Africa	77	105	98
South Asia	51	110	92
Sub-Saharan Africa	61	109	68

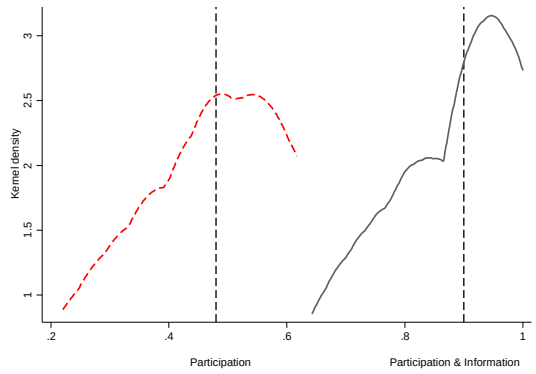
Source: World Bank Development Indicators

Table 2: International test scores, mathemathics

Country	Region	Mean Score	Grade assessed	Lowest 5th percentile	Percentage of students performing below chance
Ghana	Lower middle income	331	8th		33
South Africa	Lower middle income	352	9th		32
Botswana	Upper middle income	397	9th		19
USA	High income, OECD	509	8th	308	3

Source: TIMMS 2011

Figure 1: Share of location actions in action plan



Source: Björkman, de Walque, Svensson (2016)