EBA

02
2016

EDUCATION IN DEVELOPING COUNTRIES
WHAT POLICIES AND PROGRAMMES AFFECT LEARNING AND TIME IN SCHOOL?

Amy Damon, Paul Glewwe, Suzanne Wisniewski, Bixuan Sun

Education in developing countries – what policies and programmes affect learning and time in school?

Amy Damon
Macalester College

Paul Glewwe
University of Minnesota

Suzanne Wisniewski
University of St. Thomas

Bixuan Sun
University of Minnesota

Rapport 2016:02
till
Expertgruppen för biståndsanalys (EBA)

Amy Damon is an Associate Professor of Economics at Macalester College in St. Paul Minnesota. She received her doctorate in Agricultural and Natural Resource Economics at the University of Minnesota. Her research focuses on education in developing countries as well as household poverty reduction strategies and food security issues in developing countries.

Paul Glewwe is a Distinguished McKnight University Professor in the Department of Applied Economics at the University of Minnesota. Before coming the the University of Minnesota in 1999, he was a senior research economist at the World Bank. His research focuses on education in developing countries, and he also does research on child nutrition, poverty and inequality in developing countries, and on education in the United States.

Suzanne Wisniewski is an Associate Professor of Economics at the University of St. Thomas in St. Paul, MN. Her research is focused in areas of the economics of education, nutrition, health, migration and poverty in less developed countries. The largest body of her work was on Sri Lanka, in affiliation with World Bank, South Asian Human Development Sector. Other regions of research include Senegal, El Salvador and Peru.

Bixuan Sun is a current PhD student in applied economics at the University of Minnesota. Her research focuses on environmental economics and energy policies. She is also interested in program evaluation in developing countries

Acknowledgements

We would like to thank Stellan Arvidsson Hyving, Sonja Daltung, Eva Mineur, Jesper Sundewall, Jakob Svensson and Anna Tompsett for comments on earlier drafts of this report, and Jay Sayre and Eric Stephens for excellent research assistance.

ISBN 978-91-88143-12-9
Printed by Elanders Sverige AB
Stockholm 2016

Cover design by Julia Demchenko

Table of contens

Tables

Figures

Preface

One important challenge for development aid lies in the ability to (directly or indirectly) reinforce human capital in low- and middle-income countries, thereby positively affecting economic growth, and ultimately, to achieve poverty reduction. It is hardly possible to envisage long-term poverty reduction in the world's low- and middle-income countries that is not preceded by strengthened education systems and a more educated population. The links between education and economic growth, income distribution and poverty reduction are well established. On top of this, education is also a basic human right and a foundation for a more sustainable and inclusive society. The central and prominent role of education in global development has recently been confirmed by the Sustainable Development Goal 4: "Ensure inclusive and equitable quality education and promote lifelong learning". To increase the prospects of achieving the global goal of education for all, effective, good quality education policies, strategies and programmes must be in place.

The difficult part is to finding out what type of interventions is likely to work best in a given community or school. There are also many context-specific problems in the education sector that need to be addressed, such as low school attendance, ineffective pedagogy and unsatisfactory school performance in terms of test scores. Studies and research conclude that many children in low- and middle- income countries leave the school system without being able to read simple texts or perform simple mathematical exercises.

In development research, education is repeatedly cited as crucial from a variety of perspectives. At the same time, this sector has not been prioritised in Swedish development aid, despite substantial and alarming needs in low- and middle- income countries and despite the lack of funding for education systems. Donors and the research community on international education build up a considerable knowledge base, with hundreds of evaluations and impact studies with (potentially) important conclusions to draw on for effective future investment in the sector. However, the question remains how accessible and useful this knowledge base is, and also whether it is actually used by policy-makers and officials deciding on aid to education. This was the startingpoint for the Expert group for Aid Studies when it decided to commission two synthesis evaluations on aid to education.

In this report, Professor Paul Glewwe, Amy Damon, Suzanne Wisniewski and Bixuan Sun have made a comprehensive review and

analysis of recent research on education policies, programmes and interventions in developing countries. Their objective is to provide a tool to policy makers and donors who wish to know what research in the field has to say about effectiveness, for use when planning direct investments in the education sector. The studies included in the review are all based on quantitative analyses, where the causality between intervention and educational outcome has been established through a randomized control trial, regression discontinuity design or a difference-in-differences method. These strict inclusion criteria exclude many studies of educational outcomes, although these are, in part, analysed in the EBA report (2016:03) by Professor Joel Samoff, Jane Leer and Michelle Reddy, who take a completely different research- and methodological perspective.

The findings in the report are structured into 'what works', 'what often works', 'what seemsto be promising interventions but need more evidence' and 'what does not seem to work' in terms of achieving two broad educational goals: increasing time in school and improving learning outcomes. One of the key findings with respect to get children to increase their time in school is that the provision of conditional cash transfers and merit-based scholarships seem to stimulate improved learning outcomes. Professor Glewwe and his team point to the challenge of choosing the right measure to tackle an identified problem. In a given context, providing all children with deworming medicine may be more effective than supplying more pedagogical material to the school.

This report, together with the simultaneously published EBA report from Samoff et al. (2016:03), contains important lessons for future Swedish aid to education, but also conclusions of importance for aid effectiveness in general and for the work on evaluation of aid projects and programmes.

The work on this report has been conducted in dialogue with a reference group chaired by Professor Jakob Svensson of the EBA. The analysis and conclusions expressed in this report are solely those of the authors.

Stockholm, May 2016

Lars Heikensten

Sammanfattning

Denna rapport innehåller en övergripande sammanställning och analys av aktuell forskning om utbildningspolitik, utbildningsprogram och utbildningsinsatser i utvecklingsländer. Syftet med rapporten är att fungera som ett inspel till beslutsfattare och biståndsorgan som önskar använda sig av den mest rigorösa och tillförlitliga forskningen på området, för att bättre kunna rikta investeringar till utbildningssektorn. För att strukturera sammanställningen i rapporten på ett användarvänligt sätt har vi delat in den existerande forskningen i fyra kategorier som speglar utbildningssektorns största utmaningar: (1) låg efterfrågan på utbildning (2) otillräckliga medel till skolan (3) ineffektiv pedagogik och (4) dålig styrning av skolan. Under var och en av dessa fyra kategorier diskuteras hur effektiva olika insatser varit för att lösa specifika problem inom varje kategori. För var och en av dessa fyra stora utmaningar diskuteras hur effektivt insatserna fungerat för att uppfylla två övergripande målsättningar: (1) att öka elevernas tid i skolan (inklusive närvaron och antalet inskrivna elever samt minskade skolavhopp) och (2) att förbättra läranderesultaten (uppmätta med hjälp av provresultat). De viktigaste resultaten beskrivs kortfattat i denna sammanfattning. En sak som är unik med rapporten är att vi, där så är möjligt, ger en bakgrund till att enskilda insatser rönt framgång eller misslyckats i syfte att förse beslutsfattare med ett mer relevant underlag för bedömning av investeringar i samhällen som ställs inför särskilda utmaningar.

Samtliga 114 studier som ingår i denna analys uppfyller ett kvalitetsgränsvärde, som beskrivs i detalj i metodikavsnittet. Samtliga inkluderade studier etablerar på ett övertygande sätt ett orsakssamband mellan programinsatsen och utbildningsresultaten, antingen genom en randomiserad prövning eller genom två andra empiriska metoder som allmänt erkänns kunna fastställa orsakssamband dvs. *regression discontinuity design* (RDD) och *difference-in-differences*. Genom att bara inkludera studier i sammanställningen som klarat kvalitetskraven så innebär med nödvändighet att många andra kvantitativa (och kvalitativa) studier av utbildningsresultat utesluts.[1] Men för att effektivt kunna driva en riktad politik är det av avgörande betydelse att man med stor tillförlitlighet kan identifiera både ett programs förmåga att påverka läranderesultaten och denna inverkans potentiella storlek. För det andra

1 En granskning med ett komplimentärt metodologisk perspektiv på biståndsinsatser riktade till utbildning som mer fokuserar på kontext och komplexitet genomförs för närvarande på uppdrag av EBA av Professor Joel Samoff m.fl. (2016:03).

gör detta strikta kvalitetskriterium det möjligt för oss att presentera policyrekommendationer som grundar sig på den bästa tillgängliga forskningen i fråga om utbildningsinsatser.

När en studie bedömts ha klarat inkluderingskriteriet har den klassificerats under någon av de fyra största utmaningarna som anges ovan. När klassificeringen hade gjorts granskade vi antalet studier samt deras omfattning, inriktning och betydelse. Med hjälp av denna redovisningsmetod kunde vi föra en diskussion kring vilka typer av utbildningsinsatser som ”fungerar” för att hantera problem på de fyra utmaningsområdena. Insatserna klassificerades i en av fem effektivitetskategorier: (1) insatser som fungerar (2) insatser som ofta fungerar (3) lovande insatser där ytterligare belägg krävs (4) insatser som inte fungerar och (5) insatser där tillräcklig forskning saknas. Kriterierna för varje kategori redovisas i tabell 3.

Huvudsakliga resultat i fråga om att öka elevernas tid i skolan

Insatser som fungerar: De två mest effektiva insatserna när det handlar om att öka elevernas tid i skolan i samhällen där skolnärvaron och inskrivningsgraden var otillfredsställande handlade om att: (1) tillhandahålla villkorade utbetalningar (conditional cash transfers) (24 studier) och (2) bygga nya skolor där lokal tillgång till skolor saknades (6 studier). Både dessa insatser gjorde att det blev mindre ”kostsamt” för eleverna att gå i skolan. Båda insatserna har dock visat sig vara relativt dyra.

Insatser som ofta fungerar: Erbjudande av skolmåltider (5 studier) och privatskolor (4 studier) fungerar ofta när det handlar om att öka elevernas tid i skolan.

Lovande insatser som kräver ytterligare belägg: Det finns flera lovande insatser i fall där samhället haft problem med låg skolnärvaro och lågt antal inskrivna elever. De belägg som presenteras här grundas på insatser som behandlas i en eller två studier och som inbegriper tillhandahållande av information direkt till elever och föräldrar i form av studievägledning och information om utbildningsvinsterna, meritbaserade stipendier, öronmärkta utbetalningar, icke-villkorade utbetalningar, cyklar som elever kan använda för transport och matchande bidrag till utbildningssyften. Dessutom har både extralärare och extra undervisningsmaterial, måltider att ta med hem, stödgrupper och avmaskningsinsatser visat sig vara lovande för att öka elevernas tid i skolan. Både måltider att ta med hem och avmaskningsinsatser kan ses

som mekanismer som syftar till att öka efterfrågan i områden där dålig hälsa och hunger inverkar på skolnärvaron. Könsåtskilda skolor ger också lovande resultat. Vi rekommenderar att var och en av dessa ”lovande” insatser blir föremål för fler effektutvärderingar i syfte att bekräfta om, och under vilka omständigheter, insatserna ”fungerar”.

Insatser som inte fungerar: Rönen visar att övervakning av lärares prestationer (2 studier) och användning av skolbaserad ledning (5 studier), vilka båda innebär förändringar av skolans förvaltning, inte fungerar effektivt när det handlar om att öka elevernas tid i skolan.

Insatser där tillräcklig forskning saknas: Det finns en uppsjö av enskilda insatser (1 studie per insats) som inte visat sig ha någon effekt på elevernas tid i skolan. Vi nämner inte alla 17 insatserna här utan de framgår av texten och tabellerna nedan. Mångfalden bland dessa enskilda insatser visar dock på att beslutsfattare, givare och forskare har ett kreativt förhållningssätt till problematiken kring elevernas tid i skolan. Tyvärr saknas det ofta tillräckligt med belägg för de enskilda insatserna i denna kategori för att man ska förstå varför de *inte* haft effekt. Innan ytterligare finansiering söks för dessa insatser bör man på allvar granska de speciella omständigheter under vilka varje insats genomfördes i syfte att förstå hur de bäst kan anpassas för framtida bruk.

Huvudsakliga resultat i fråga om att förbättra läranderesultaten

Insatser som fungerar: I jämförelse med insatserna för att öka elevers tid i skolan fann vi här ett större utbud av effektiva insatser för att förbättra elevernas lärande. I fall där ett samhälle lyckats få barn att gå till skolan, behålla en relativt hög skolnärvaronivå och minska skolavhoppen förefaller åtminstone fyra insatser ha visat sig fungera ganska effektivt för att förbättra elevernas läranderesultat i skolan. Särskilt meritbaserade stipendier (4 studier), tillhandahållande av extraundervisning och stödundervisning (3 studier), ökad lärartäthet (3 studier) och byggnation av nya skolor (3 studier) har visat sig fungera effektivt när det handlar om att förbättra läranderesultat bland skolelever. Dessutom har dessa insatser visat sig vara effektiva i en mängd länder och sammanhang.

Insatser som ofta fungerar: Det finns även en lovande grupp insatser som fungerar när vissa förutsättningar föreligger där programmet genomförs. Villkorade utbetalningar har i vissa fall visat sig förbättra provresultat (8 studier), på samma sätt som skolmåltider (4 studier). Datorer och elektroniska spel har också förbättrat inlärningen i många fall (11 studier). Förändringar i skolans förvaltning genom införande av

skolbaserad ledning (7 studier), resultatbaserad lön för lärare (4 studier) och möjligheter att gå i privatskola (5 studier) är samtliga insatser som i vissa sammanhang har visat sig förbättra elevers lärande.

Lovande insatser som kräver ytterligare belägg: Många kreativa insatser (samtliga med 1-2 studier) har bedömts för att se huruvida de förbättrar elevernas lärande och ganska många olika program har visat sig vara lovande. Dock har insatserna i denna kategori antingen uppvisat blandade resultat eller inte åtföljts av flera studier. Exempelvis har alla insatser som inneburit att man förlängt skoldagens längd, erbjudit skolmåltider eller undervisningsmaterial med nivåindelning (i kombination med partnerskap mellan föräldrar och lärare), järntillskott, skolgång vid elitskola i offentlig regi, ett paket av infrastruktur, skolmaterial och undervisning samt inhyrda lärare uppvisat blandade resultat. Å andra sidan har samtliga insatser där man erbjudit glasögon, måltider att ta med hem och oväntade generella bidrag till skolan uppvisat positiva resultat. Dessa resultat grundas dock på bara en utvärdering av en enskild insats.

Insatser som inte fungerar: Det är inte troligt att övervakad lärarnärvaro, utan beaktande av incitamentsrelaterade betalningar, kommer att förbättra lärarnas närvaro och därmed elevernas lärande.

Insatser där tillräcklig forskning saknas: Vad gäller de fjorton insatser som sammanfattas i tabellerna 8-11 vet vi för lite om deras effektivitet när det handlar om att förbättra provresultat. Det krävs mer forskning, särskilt när det gäller de mer innovativa insatserna i denna grupp, som t.ex. öronmärkta utbetalningar som genom att minska villkorandet av utbetalningen eventuellt sänker de övervakningskostnader som är förenade med villkorade utbetalningar, eller pedagogik som kräver mycket inläsning och åtföljs av läromedel, då dessa uppvisat både en del positiva resultat och obetydliga resultat.

Övergripande rekommendationer

Mot bakgrund av den omfattande granskning som presenteras i detta dokument föreslår vi en möjlig övergripande investeringsstrategi som grundas på dessa resultat. Det här dokumentet innehåller för det första en rad investeringsrelaterade prioriteringar, i form av insatser som bevisats fungera när det handlar om att förbättra lärandet, inskrivningsgraden, skolnärvaron eller för att minska skolavhoppen. Vi vill verkligen rekommendera betydande investeringar i dessa beprövade insatser. Resultaten i dokumentet samt de granskade studierna ger en detaljerad beskrivning av dessa prioriterade insatser.

Förståelsen av utbildningsinsatserna och programverksamheten har även resulterat i en andra prioritering i fråga om utbildningsrelaterade investeringar, som handlar om att öka kunskapen om vad som ”fungerar”. Det råder inte brist på kreativa insatser, vilket vi sammanfattar här, och fler är på väg. Sida och EBA har möjlighet att ta en ledande roll, genom att bl.a. öka förståelsen för dessa kreativa insatser genom att investera mer i koncepttestprogram (proof of concept) som både är kreativa och genomförs på ett sätt som är möjligt att utvärdera. Även om vi gjort framsteg när det gäller att förstå hur effektiva vanliga insatser är så har vi fortfarande mycket kvar att lära när det gäller att förbättra läranderesultaten för många barn runt om i världen.

Excecutive Summary

This report provides a comprehensive summary and analysis of the recent research on education policies, programs and interventions in developing countries. The objective of this report is to provide a tool to policy makers and aid agencies who wish to use the most rigorous and reliable research in the field to direct investment in the education sector. To organize this review in a user-friendly manner, we classify existing research into four categories that correspond to the major challenges faced by the education sector: (1) low demand for education; (2) inadequate school inputs; (3) ineffective teaching pedagogy; and (4) low-quality school governance. Within each of these four categories we discuss the effectiveness of various interventions at addressing specific problems within each category. Under each of these four major challenges, we discuss the effectiveness of interventions at achieving two broad goals: (1) increasing time in school (including attendance, enrollment, and reduced drop-out rates) and (2) improving learning outcomes as measured by test scores. The major findings are briefly summarized in this executive summary. One unique aspect of this report is that, wherever possible, we provide context for the success or failure of specific interventions in order to provide policy makers with more relevant information as they consider investments across diverse communities that face specific challenges.

All 114 studies included in this analysis meet a target quality threshold described in detail in the methodology section. The studies included all have convincingly established causality between the programmatic intervention and the educational outcome either through a randomized control trial, or through two other empirical methods that are generally recognized as capable of establishing a causal relationship: regression discontinuity design and difference-in-differences. Using this threshold necessarily excludes many quantitative (and qualitative) studies that examine educational outcomes.[2] However, to direct policy effectively it is critical to identify with a high degree of confidence both the ability of a program to have an impact and the potential size of that impact on educational outcomes. Further, this strict inclusion criterion allows us to make policy recommendations based on the best research available on education interventions.

[2] A review with complementary methodological perspectives focusing on context and complexity commissioned by EBA is being conducted by Professor Joel Samoff with team (2016:03).

Once a study was deemed pass to the quality threshold it was then categorized as addressing one of the four main challenge areas listed above. Once categorized, we counted the number of studies and the size, direction, and significance of study estimates. Using this accounting method, we are able to discuss what types of interventions "work" for addressing problems within the four challenge areas. Interventions were categorized into one of five effectiveness categories: (1) interventions that work; (2) interventions that often work; (3) promising interventions that need more evidence; (4) interventions that do not work; and (5) interventions with insufficient research. Inclusion criteria in each category are presented in Table 3.

Main Findings – Increasing Time in School

Interventions that work: The two most effective interventions to increase time in school in communities where attendance and enrollment are suboptimal are: (1) provision of conditional cash transfers (24 studies) and (2) building new schools where local access is lacking (6 studies). Both of these interventions reduce the "cost" of attending school for students. However, both have been shown to be relatively expensive interventions.

Interventions that often work: The provision of school meals (5 studies) and private schools (4 studies) often work to increase time in school.

Promising interventions that need more evidence: There are several interventions that are promising when communities face problems of low student attendance and enrollment. The evidence presented here is based on interventions with 1 to 2 studies. These include providing information directly to students and parents in the form of school counseling and information on the returns to education; merit-based scholarships; labeled cash transfers; unconditional cash transfers; bicycles for student transportation; and matching remittances for educational purposes. Further, extra teachers and teaching materials, take-home rations, support circles and deworming interventions all are promising interventions to increase time in school. Both take-home rations and deworming interventions may be viewed as mechanisms for increasing demand in regions where poor health and hunger are a constraint on attendance. Single sex schools are also promising. We recommend that more impact evaluations on each of these "promising" interventions be conducted to confirm whether, and under what circumstances, they "work".

Interventions that do not work: The evidence suggests that monitoring teacher performance (2 studies) and implementing school-based management (5 studies), both of which are changes in school governance, are ineffective at increasing time in school.

Interventions with insufficient research: There are a multitude of unique interventions (1 study per intervention) that show no impact on time in school. We do not list all 17 interventions here; they are presented in the text and tables below. However, the diversity of these unique interventions points to the creative problem solving nature of policy-makers, donors, and researchers in addressing time in school problems. Unfortunately there is often too little evidence on any one of the interventions in this category to understand why there was *no* impact. Before pursuing additional funding for these interventions, one should seriously consider the unique circumstances under which each intervention was attempted to learn how best to adapt it for future use.

Main Findings – Improving Learning Outcomes

Interventions that work: Compared with interventions that increase time in school, we found a wider variety of effective interventions to increase student learning. Thus once a community can get children to school, keep attendance rates relatively high, and reduce dropping out, it seems that there are at least four interventions that have proved to be quite effective at improving students' learning outcomes once they get to school. Specifically, merit-based scholarships (4 studies), providing supplemental or remedial instruction (3 studies), decreasing pupil-teacher ratios (3 studies), and building new schools (3 studies) all have been proven to be effective at improving learning outcomes for students in school. Moreover, these interventions have been shown to be effective across a multitude of countries and contexts.

Interventions that often work: Also promising are a group of interventions that work given that certain preconditions exist where the program is being implemented. Conditional cash transfers have in some contexts been shown to increase test scores (8 studies), as has the provision of school-based meals (4 studies). Computers and electronic games have also improved learning in many cases (11 studies). Changes in school governance that implement school-based management (7 studies), provide teacher performance pay (4 studies), and provide opportunities to attend a private school (5 studies) have all been shown, in some contexts, to increase student learning.

Promising interventions that need more evidence: Many creative interventions (all with 1 or 2 studies) have been assessed to see whether they improve student learning, and a relatively wide variety of programs have shown promise. However, interventions in this category either lack multiple studies or have shown mixed results. For example, increasing the number of hours in a school day, providing school feeding or multi-level teaching materials with parent-teacher partnerships, providing iron supplementation, attending an elite public school, provision of a package of infrastructure, materials, and training, and contract teachers, all have shown mixed results. On the other hand, provision of eyeglasses, take-home rations, and unanticipated school block grants all have shown positive results but these findings are all based on just one evaluation of a single intervention.

Interventions that do not work: Monitoring of teaching attendance, without specific attention to incentive pay, is unlikely to improve teacher attendance, and thus is unlikely to increase student learning.

Interventions with insufficient research: There are 14 interventions summarized in Tables 8-11 that we know too little about in terms of their effectiveness at improving test scores. More research is needed, especially on the more innovative of the interventions in this group such as labeled cash transfers which reduce the conditionality of the transfer thereby possibly reducing the monitoring costs associated with CCTs, or reading-intensive pedagogy with accompanying reading materials, which has shown some positive results along with insignificant results.

Overall Recommendations

Given the extensive review provided in this document, we suggest a possible overall investment strategy based on the results found here. First, this document provides a set of investment priorities in terms of proven interventions that work to increase learning or improve enrollment, attendance, or drop-out rates. We highly recommend that significant investments be made in these proven interventions. The findings in this document and the studies reviewed provide a detailed account of these priority interventions.

However, u**n**derstanding the landscape of education interventions and programming has also brought about a second priority for education investment spending, which is to increase the knowledge base of "what

works". There is no shortage of creative interventions that we have summarized here, and even more that are in the pipeline. Sida and EBA have an opportunity to be leaders in widening the scope of understanding around these creative interventions by investing in more "proof of concept" programs that are both creative as well as implemented in a way that in amenable to evaluation. While we have made headway in understanding the effectiveness of popular interventions, we still have much to learn to improve the education outcomes for many children around the world.

I. Introduction

Economists and other researchers have accumulated a large amount of evidence that education increases workers' productivity and thus increases their incomes, which in turn leads to decreases in poverty. There are also many non-monetary benefits of education, such as improved health status and reduced crime (Lochner, 2011). At the country level there is also a large amount of evidence that education increases the rate of economic growth (Hanushek and Woessmann, 2015). These analyses all highlight the value of improving a country's human capital as an important pathway toward poverty reduction, and thus they provide the motivation for developing countries to invest in the skills and human capital of their populations through expanding, and improving the quality of, their formal education systems. They do not, however, indicate which types of specific investments should be pursued to improve the quality of education in developing countries.

While there has been a remarkable decline in the number of out-of-school children worldwide, from nearly 100 million in 2000 to 58 million in 2012 (UIS & UNICEF, 2015), it is nearly certain that the developing world has not achieved the Millennium Development Goal (MDG) of universal primary education in 2015. UNESCO (2014) argues that the progress of getting out-of-school children into school that occurred in the early 2000s has dramatically slowed, and there has been little progress since 2007, and that this is concurrent with a stagnation in aid to education, which has not changed as a percentage of official development assistance since 2002 (UIS & UNICEF, 2015). Moreover, many children in developing countries who do attend school appear to learn little during their time in school (Glewwe and Kremer, 2006; Glewwe et al., 2013; Hanushek and Woessmann, 2015).

Given the current global aid environment, well-targeted education aid should be directed toward programs that have been shown by rigorous evaluations to be effective for achieving the MDG of universal primary education and for improving the quality and effectiveness of the education that students receive. There is abundant knowledge about what issues and problems exist in the education sector in developing countries and there is no shortage of proposed policies to address these problems. However, with limited resources policy-makers need information from high-quality policy impact evaluations to make effective future investments.

Fortunately, many empirical studies on education in developing countries have been conducted in the last 25 years in an effort to determine which education policies and programs "work", and this research has accelerated in the past 5-10 years in terms of both quantity and quality. This research has focused on two main questions. The first is: What education policies and programs increase student enrollment, attendance, and completed years of schooling? The second is: What education policies programs increase student learning?

This report provides a summary of rigorous high-quality evaluations of education interventions in developing countries, and reports their impacts on educational outcomes. Of course, this report is not the first to provide a summary of this literature, however it adds to the other reviews in three important ways.[3] First it provides the most current review of the literature to date. Second, it includes rich contextual detail to provide some evidence (where available) of the mechanisms behind a program's success, or lack thereof. Third, this report widens the scope of the review by Glewwe et al. (2013) in several ways. It includes a broad search of the "grey" literature, [4] presents the sizes of the estimated impacts on (standardized) test scores, provides more context-specific details about findings, presents information on the costs of the studies, and finally uses the findings to make explicit policy recommendations.

All of the studies included in this analysis meet a quality threshold that is described in detail in the methodology section below. The studies included all have convincingly established causality of the programmatic intervention on the educational outcome either through a randomized controlled trial, or through two other empirical methods that are generally recognized as able to establish a causal relationship: regression discontinuity design and difference–in-differences. Using this threshold necessarily excludes many quantitative (and qualitative) studies that examine educational outcomes. [5] However, to direct policy effectively it is critical to identify with a high degree of confidence both the ability of a program to have an impact, and the potential size of that impact, on educational outcomes. Further, this strict inclusion criterion allows us to make policy recommendations based on the best research

3 These reviews are discussed in detail in Section II below.
4 The "grey literature" consists of papers and reports that attempt the estimate the impact of an educational program that have not been published in an academic outlet. These papers primarily take the form of papers, evaluations and reports written by international development agencies.
5 A review of qualitative studies is being conducted by Professor Joel Samoff with tema and therefore outside the scope of this report.

available on education interventions. Using studies that provide unreliable, or noisy, estimates risks directing policies toward unproductive or ineffective programs, and the consequent misallocation of resources may pose a large opportunity cost for both the donor agencies and the aid recipients.

We acknowledge that there are many challenges in synthesizing the evidence to reach a definitive view of what works, and what does not. These challenges include variations in context, variations in duration of the evaluations and the outcomes studied, and perhaps more importantly variations in the details of the interventions that have been evaluated. Therefore our objective is not only to synthesize a body of high quality and rigorous evaluations in order to identify "what works", but also to discuss and interpret these results and to discuss the reasons why some interventions appear to be effective and others do not, with the ultimate goal of drawing implications for both research and policy.

That being said, it is important to note that not all education policies, programs and interventions lend themselves easily to rigorous evaluation using the methods outlined below. In an effort to provide as broad of a policy perspective as possible, we have included many studies of "unique interventions" that provide evidence of innovative single interventions as well as interventions that are provided in a "package" to a school or community. Still, some institutional interventions or governance interventions, such as changes in how teachers are trained, are largely excluded from our analysis because they are difficult to evaluate rigorously.

All of the included studies identify and evaluate the effect of an education intervention on one or more educational outcomes. Following Glewwe and Muralidharan (forthcoming), the interventions evaluated by these studies are categorized into four broad types, and their impacts are assessed for two broad types of educational outcomes. [6] The four broad types of interventions are: 1. Interventions designed to increase households' demand for (interest in) sending their children to school; 2. Interventions that increase school inputs; 3. Interventions that attempt to improve teaching pedagogy; and 4. Interventions that attempt to improve school governance. The educational outcomes can be classified into two

6 Both this report and Glewwe and Muralidharan (forthcoming) are based on the same literature search, which was done by a combined team of academic economists and their students. Glewwe and Muralidharan is written for academic economists, while this report is written for international aid agencies. Consequently there are considerable differences between these two documents in both content and presentation.

broad categories: 1. Time in school variables (e.g. enrollment rates, on-time enrollment, daily attendance and, most importantly, years of schooling attained); and 2. Learning outcomes, which are typically measured by test scores. Where appropriate, we standardize the results in order to compare the outcomes of similar interventions implemented in different countries or regions. Examining the consistency or inconsistency of results across contexts helps to further improve the strength of (or appropriately qualify) our policy recommendations.

The most important objective of this synthesis of education program evaluations is to determine which education aid programs "work". In doing so, this study relies on a broad set of recent rigorous evaluations that have been conducted on programs that attempt to improve the education outcomes discussed above. While all evaluations have both strengths and weaknesses, the main objective of this synthesis is to identify the most rigorous studies on education in developing countries and to synthesize these results to obtain a broad understanding of the effectiveness of the many education interventions that have been implemented.

Given the difficulty of implementing high quality education interventions, and the complexity of conducting a rigorous and sound evaluation of such interventions, the number of "high quality" education evaluations is relatively limited. While there has been an increase in the number of high quality evaluations in recent years, there are also many evaluations of education programs that have serious methodological flaws that make it impossible to draw causal inferences from their results. This review is limited to only those evaluations whose quality is sufficiently high to provide credible estimates of program impacts. Table 1 (in the next section) outlines the four steps used to select papers. The result is that our analysis includes 114 high quality papers, of which 75 are randomized controlled trials.

II. Review of Complementary Work

The broad scope of education research in recent years provides ample room for broad summaries, similar to the one produced in this report, that attempt to systematically review or meta-analyze this entire body of work. This report is not the first to do so. In this section we review other summary papers and reports that also review research in the field of education policy and programs in developing countries.

In our search for other reviews on education research in developing countries we found a large number, 27, of varying breadth and quality. Our primary purpose in searching for these reviews was to double-check that we had captured all of the relevant studies to the fullest extent. This was a fruitful exercise and we identified 16 relevant published or working papers which we had missed on our first pass of the literature (which is described in Section III). These studies have been incorporated into our review of the literature in Sections V and VI.

The second reason for checking these reviews was to ensure that our review was as broad as possible. The character of each of the meta-analyses and systematic reviews that we found vary significantly. We found seven review papers (Conn, 2014; McEwan, 2015; Evans and Popova, 2015; Glewwe et al., 2013; Kremer, Brannen, and Glennerster, 2013; Murnane and Ganimian, 2014; Krishnaratne, White, & Carpenter, 2013) that attempted to summarize the entire body of work concerning education in developing countries, similar to the report presented here, while the other 20 reviews examined specific types of education interventions and programs in developing countries (for example, Baird et. al. (2013) for conditional cash transfers, and Morgan et. al. (2013) for school vouchers).

Kremer, Brannen, and Glennerster (2013) provide a review of education interventions in developing countries, yet their review includes only randomized evaluations. They conclude that school enrollment for poor families depends heavily on costs of enrollment and thus programs that reduce either the explicit costs (merit scholarships) or implicit costs (conditional cash transfers) help to increase enrollment. In addition, they find that providing information to families on the extent to which additional years of schooling leads to increased earnings, and child health interventions (such as providing deworming medicine), both provide cost effective ways to increase school enrollment. They also identify several interventions that increase student learning, such as matching pedagogical methods to students' learning levels and improving teacher

accountability. In contrast, they find that "traditional" interventions, such as providing textbooks, hiring teachers, and providing grants that schools can choose to use in a variety of ways all do little to change learning outcomes as measured by test scores.

McEwan (2015) also reviews only randomized controlled trials, and he focuses on interventions that attempt to improve test scores. He finds that deworming children has no effect on student learning outcomes (Kremer et al. focused on time in school), and the same is true of monetary grants. He finds that learning outcomes are most responsive to pedagogical interventions such as computer-based learning and instructional technology, teacher training, teacher performance incentives, and peer learning, as well as school inputs such as smaller class sizes and instructional materials. Conn (2014) corroborates McEwan's findings that pedagogical interventions have a significant impact on learning outcomes, and she states that they have a larger impact than any other of the 11 interventions that she analyzed. After pedagogical interventions, Conn finds that teaching methods that are aligned with students' learning styles, as well as teacher training or coaching, are the next most effective means of improving learning outcomes. Krishnaratne, White, and Carpenter (2013) find that conditional cash transfers (CCTs) and health interventions increase attendance and that instructional materials (chalkboards, flip charts, computers or teachers) for math increase math scores. They also conclude that providing teacher with pedagogical resources is promising for improving both attendance and test scores, yet they conclude that more studies are needed in this area.

Finally, Evans and Popova (2015) provide the most recent review of the education impact evaluation literature by providing a review of the reviews. They argue that reviews over the past five years have come to dramatically different conclusions based largely on their sample criteria. They conclude that future studies should include both a quantitative analysis of the studies in a given category along with a narrative review to explore the heterogeneity across studies and discuss the mechanisms for change behind the outcomes which each study identifies and measures.

In addition to these comprehensive studies, a myriad of studies (many supported by the International Initiative for Impact Evaluation (3IE) provide reviews that are specific to certain types of interventions. These include studies of conditional cash transfers (Baird, Ferreira, Özler, and Woolcock, 2013; Kabeer et al., 2013; Saavedra and Garcia, 2012), separate toilets for girls (Birdthistle, Dickson, Freeman, Javidi, 2011), social programs in Latin America (Bouillon and Tejerina, 2007), teacher

salaries (Carr et al., 2011), teacher attendance (Guerrero et al., 2012), sanitation in schools (Jasper et al., 2012), school feeding programs (Kristjansson et al., 2006), eliminating school fees (Morgan et al., 2012), school vouchers (Morgan et al., 2013), family and community support (Spier et al., 2014), deworming drugs (Taylor-Robinson et al., 2012), and economic transfers to women (Yoong et al., 2012).

This report builds on these studies by constructing a broad catalogue of rigorous, high quality studies and providing a narrative review along with a reporting of the estimates. In particular, as suggested by Evans and Popova (2015), we contribute to this body of research by discussing the heterogeneity across studies and considering, wherever available, the mechanism driving the relationship between the intervention and the outcome.

III. Methodology[7]

This section explains the methodology used for this review of the literature. The first subsection explains how high quality studies are defined, and how they were selected from the thousands of papers available. The second subsection briefly reviews the "grey literature" that is the studies and reports by international aid agencies that are not published in academic journals. Finally, the third subsection explains how interventions were classified into five broad types, ranging from interventions that almost always "work" to those that almost always "do not work".

Procedure for Selecting Studies

One of the main contributions of this report is its review of the evidence on the impacts of different types of education policies and programs on student learning and time in school. However, an important challenge for empirical research in this area is that of credible causal identification. We therefore limit our synthesis of the evidence to 114 high quality studies that were conducted from 1990 to 2014.[8] In this section, we discuss the criteria for selecting these studies from the hundreds, if not thousands, of relevant studies in the literature. To identify evaluations that have produced credible impacts of education programs or policies, the following four-step selection process was implemented. For convenience, the selection process is summarized in Table 1.

[7]This method for selecting studies was also used in the review of Glewwe and Muralidharan (forthcoming), which is written for a more academic audience.

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

Table 1. Steps Used to Select Papers Reviewed in this Report

Review Step	Procedures Used	No. of Papers
1	Search EconLit and ERIC databases.	13,437
	Review abstracts to eliminate duplicate papers and papers that did not estimate the impacts of school or teacher characteristics for a developing country.	1,017
2	Review full papers; eliminate papers based on lack of relevance or lack of quantitative analysis.	320
3	Exclude papers that are not "high quality" (RCT, RDD, DD)	114
4	Number of high quality studies that are RCT studies	75

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

To begin, a search was conducted on a wide variety of sources, after which evaluations were systematically eliminated that did not meet a series of criteria for relevance and quality. To be included in the search, studies had to have been published in (peer-reviewed) academic journals in the 25 years literature from 1990 to 2014, inclusive. Unpublished academic working papers written from 2010 to 2014 were also included. Academic working papers written before 2010 that had not been published by the end of 2014 were judged as likely to have some methodological flaws that have resulted in their not being published in peer-reviewed journals, so only academic working papers that were written from 2010 to 2014 were included. If a working paper version was available in this time period, and the published version also appeared by 2014, only the published version was considered since presumably the published version is likely to have some improvements that were not in the working paper version.

The very first task was to conduct a search for journal articles published between 1990 and 2014 using two search engines for the economics and education literatures, respectively: EconLit and the Education Resources Information Center (ERIC). All papers that list both "education" as a key word and any one of a list of 124 educational programs or policies as key-words (see Appendix 3 for this list) were included in this initial sweep of the literature. The search was also limited to papers that include the name of at least one developing country or the term "developing country" or "developing countries" in the abstract. Developing countries were defined using the International Monetary Fund's list of emerging and developing countries.

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

In addition to published papers, a search was conducted of several prominent working papers series: National Bureau of Economic Research (NBER) working papers; World Bank Policy Research working papers; the Institute for the Study of Labor (IZA); the Center for Economic and Policy Research (CEPR); the CESIfo Research Network; the Rural Education Action Project (REAP) at Stanford University; and Young Lives Working Papers. Papers listed as education papers on the Abdul Latif Jameel Poverty Action Lab's website were also searched. As mentioned above, working papers that appeared before 2010 were excluded based on the assumption that high quality working papers written before 2010 should have been published by 2014. As with the selection criteria for published papers, evaluations that do not focus on developing countries, or that do not estimate the impact of a program or policy on students' educational outcomes, were not included. The number of published papers and working papers that remained after reading their abstracts was 1,017.

Step 2: Read Entire Paper/Report to Verify Relevance

In the second step, all 1,017 of the evaluations that were not eliminated in the first step were reviewed to obtain further information about each study. During this step, additional papers were eliminated for lack of relevance that was not evident from reading the abstracts. Possible reasons for lack of relevance were: (1) The evaluation did not focus on a developing country (which was not always clear in the abstracts); (2) The paper did not evaluate any type of education policy or program; and (3) The paper did not include quantitative analysis of the impact of an education policy or program on students' educational outcomes. After this step was completed, 320 papers remained that were relevant for this review of the literature.

Step 3: Retaining Only High Quality Evaluations

In the third step, the evaluations that were not eliminated in the first two steps were reviewed for their quality. While regression analysis is commonly used to estimate the impact of a policy or program on an educational outcome, a very serious problem with regression analysis is

that some factors that have a causal impact on the education outcome variables of interest are unlikely to be available in the data, which can lead to bias in regression estimates of the impacts of education policies and programs, and thus to misleading results. This is the problem of *omitted variable bias*. Another problem with regression analysis is that regressions often included many school and teacher characteristics as control variables, and in many cases authors are not particularly interested in the coefficients associated with those variables, and so they should not be interpreted as estimates of the causal impacts of those variables.

Given these problems with regression estimation methods, all studies based on those methods alone are deemed not to be high quality studies. Since matching estimators invoke similar assumptions, in particular the assumption that conditioning on (other) observed variables implies that (observed and counterfactual) education outcomes are independent of the program participation variable, studies based on matching methods were also excluded. This leaves three types of studies that are considered to be high quality studies in this review. First, all evaluations based on a well-implemented randomized controlled trial (RCT) are included in the set of high quality studies, as these studies avoid, or at least minimize, many types of estimation problems. Second, estimates based on a difference in differences (DD) regression (which requires longitudinal data) are deemed to be high quality studies. Finally, evaluations based on regression discontinuity design (RDD) are also considered to be high quality studies. The set of papers that were retained after this third step contained 114 "high quality" studies.

Step 4: Identify Randomized Controlled Trial (RCT) Evaluations

The fourth and final step of the review set an even higher bar for the quality of an evaluation. Well implemented randomized controlled trial (RCT) studies arguably have the highest credibility, when implemented correctly. In particular, difference in differences (DD) studies must rely on the parallel trends assumption, which is difficult to verify, and regression discontinuity design (RDD) evaluations identify impacts only close to the "cutoff point", strictly speaking. Of the 114 high quality studies, about two thirds (75) were RCTs.

While RCT studies are likely to have the highest credibility, it is also worth noting that they do have some disadvantages. First, they can be expensive to implement, and this is especially true if randomization is

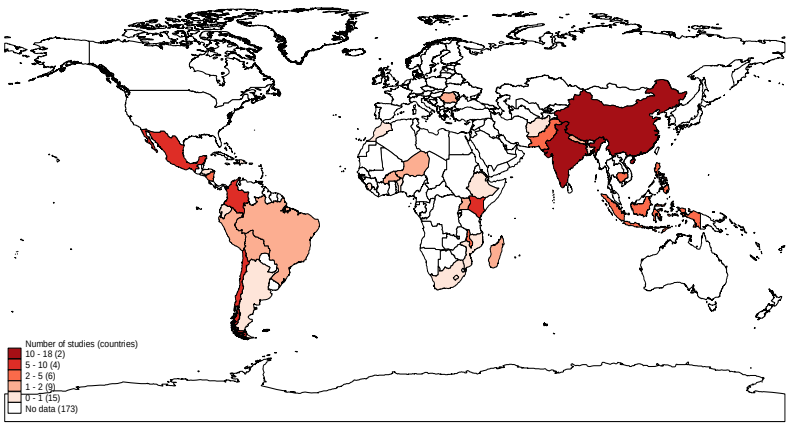
done at the school level instead of at the student level.[9] Second, since control schools typically need to be "treated" within 1-2 years of the start of the trial (in order to obtain their cooperation), it is often not possible to assess long-term program effects, although some control school students may never be treated because they "age out" of their schools before those schools are treated, so for this group of students it is possible, in principle, to measure long-term impacts. Third, it is worth noting that even the best RCT study has only internal validity, and so the generalizability of the findings to other circumstances cannot be known; yet this is a problem for most, and perhaps all, non-RCT studies as well. Fourth, many problems can arise when implementing RCTs, such as some students in the treatment schools refusing to "accept" the treatment (e.g. provision of medical care or meals in schools), some students in the control group managing to obtain the treatment, or differential rates of sample attrition between the treatment and control schools. Fifth, some types of education interventions may be hard to randomize, such as the introduction of a new curriculum at the same time in all schools, or changes to pedagogy in teacher training colleges.

This fifth disadvantage implies that these types of interventions are very unlikely to be "recommended" to policy makers using the methodology proposed in this report.[10] A final criticism of RCTs is that even when they give reliable estimate of the impact of a program, they shed little information on *why* the program is effective. In recent years RCTs have been conducted in ways that try to determine why a program is effective; at minimum, RCT evaluations should be combined with high-quality qualitative information or evaluation techniques to provide a more complete picture of the effects of a program.

The 114 high quality studies identified from this 4-step process are from 36 different developing countries. As can be seen from Figure 1, these countries are found across the developing world. The two countries with the most studies are China and India, which is not surprising since these countries have by far the largest populations in the world.

9 As discussed in Section 7, a rough estimate of the cost of an RCT for which randomization is done at the school level is $500,000 to $1,000,000 (not counting the cost of the intervention).
10 In principle, it may be possible to evaluate such policies using either regression discontinuity or difference in differences estimation, but in practice this is very unlikely.

Figure 1. Distribution of studies



Review of the Grey Literature

In addition to reviewing the academic studies we identified through the methodology just described in subsection III.A, we also reviewed the "grey" literature from several international organization and NGO sources using the same selection criteria. The purpose was to identify unpublished studies that may have employed the same types of high quality quantitative methods.

The "grey literature" consists of papers and reports that attempt to estimate the impact of an education program. These reports have not been published in any academic outlet, either as a working paper or in an academic journal. The grey literature consists of papers, evaluations and reports written by international aid agencies (both governmental and non-governmental) or by the consulting firms that those agencies hire. To find high quality evaluations in the grey literature, we searched the websites of, and/or directly contacted, the following organizations: Mathematica Policy Research, Save the Children Fund, Millennium Challenge Corporation, USAID, Sida, DFID, Norway Agency for Development Cooperation, Oxfam, UNICEF and the World Bank. The number of studies found through this search is outlined in Table 2.

Table 2. Summary of Grey Literature Search

Grey Literature Source	#of Education Related Studies or Reports	#of Quantitative Studies on Education	#of Additional High Quality Quantitative Evaluations
Mathematica Policy Research	67	5	1
Save the Children Fund	66	5	1
Millennium Challenge Corp (MCC)	51	2	0
USAID	0	0	0
Sida	75	0	0
DFID	0	0	0
Norway Agency for Devel. Cooperation	0	0	0
Oxfam	17	0	0
UNICEF	230	26	1

Overall, we found very few studies in the grey literature that were unpublished *and* met the quality standards laid out above. However, given the rising popularity of implementing randomized control trials for impact evaluations, this is likely to change in the coming years. Increasingly NGOs and bilateral aid agencies are interested in rigorously evaluating the impact of their programs in order to efficiently and effectively direct their limited resources. However, while the number of such evaluations is certainly increasing, rigorous impact evaluations of programs need to be more widespread given that opportunities to direct investments to more effective interventions are undoubtedly being missed. The more rigorous evaluations we have on diverse topics in the education sector, the clearer the answers will be for important policy concerns. More information about how the grey literature search was conducted is provided in Appendix 2.

Assessment of Intervention Effectiveness

Following Krishnaratne, White and Carpenter (2013), we created a rubric to assess the effectiveness of each type of education intervention. More specifically, we categorize each education intervention into one of five different effectiveness categories: (1) Interventions that work, which is defined as interventions for which there are three or more studies, and

most or all of the studies show positive and statistically significant impacts; (2) Interventions that often work; (3) Interventions that seem promising but for which more evidence is needed to determine their effectiveness; (4) Interventions that consistently do not work as they are intended; and (5) Interventions that are less promising but for which evidence is insufficient to draw any general conclusions, these are typically interventions with a single study with insignificant results. These categories are defined in Table 3.

Using the rubric discussed above, which is defined in Table 3 and summarized in Figure 2, we discuss the effectiveness of interventions to address common problems in the education sector. In doing so we focus on the two broadest types of goals for education interventions: *1. Increasing time in school; and 2. Increasing student learning as measured by test scores.*

Table 3. Rubric for Categorizing the Effectiveness of Studies.

What works	• Requires 3+ studies with positive and significant outcomes. May include a few non-significant or negative estimates, but most are significantly positive.
What often works	• Evidence that the intervention often works, but sometimes does not work. Before implementing, policy makers should carefully consider whether conditions in their countries are similar to those in the countries where the policy appears to have worked. • Requires 3+ studies • Relatively equal mix of positive and significant estimates with negative and/or nonsignificant estimates.
What is promising but needs more evidence	• Promising evidence that the intervention works, but needs additional evidence. • Requires: 1-2 studies. If 2 studies, 2+ positive and significant estimates are required; may include a limited number of non-significant estimates and even 1 negative estimate. • If 1 study, only positive and at least 1 significant estimate
What does not work	• No evidence that the intervention works • Requires 2+ studies • Most estimates are negative or insignificant
What we don't know and needs more research	• No evidence that the intervention works, in part due to too little evidence on the intervention effectiveness, needs further evaluation. • Requires: 1-2 studies

Figure 2. Five Categories of Effectiveness, as Determined by Findings and Number of Studies

	Number of Studies		
Findings:	1	2	3 or more
Mostly significantly positive	What is promising		What works
Sometimes significantly positive	What we don't know		What often works
Rarely significantly positive	What we don't know	What does not work	

IV. Common Challenges Faced by the Education Sector

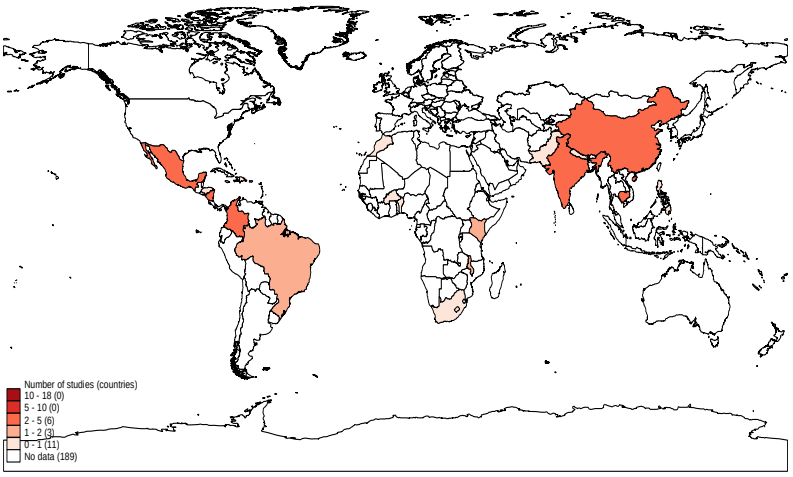
This section defines and discusses four broad challenge areas faced in the education sector. Sections V and VI then presents findings on the effectiveness of education interventions that are designed to address some of these challenges. As will be shown below, some types of interventions are more conducive to increasing time in school, while others may be more effective at increasing student learning.

A Challenge 1: Low Household Demand for Education.

Households' decisions on whether to send their children to school are based on comparing the (perceived) costs and benefits of doing so. In general, if they think that the benefits of sending a given child to school for another year exceed the costs, then they will enroll their child for another year. However, there are several reasons why households may not enroll their child in another year even if they think that the benefits exceed the costs. Possible reasons include: (1) Parents may not accurately perceive the returns, or costs, to education; and (2) Households may be credit constrained and thus cannot afford to send their children to school even when the benefits exceed the costs. Finally, even if parents can and do choose the level of education that maximizes household benefits, net of costs, they are unlikely to take into account the social benefits (benefits that accrue to the rest of society) from their children being educated, and so from the viewpoint of society as a whole they may under-invest in their children's education.

Figure 3 shows the distribution across 20 countries for which there is evidence from high quality studies on the impacts of education interventions that are intended to increase households' demand for schooling. The number of studies per country is indicated by the shade of color for that country. Note that most of these studies originate from only four countries: China, Colombia, India and Mexico.

Figure 3 Distribution of Demand Intervention Studies

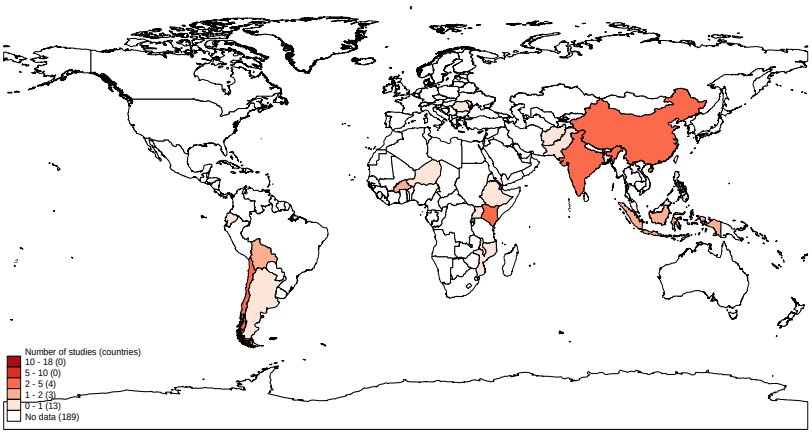


B Challenge 2: Inadequate School Inputs

Many studies have examined the impact of inputs, broadly defined, on students' educational outcomes. This reflects the fact that most of the costs of building schools are input costs, and the same is true of the costs of operating schools when teachers are considered to be inputs. It also reflects the reality that most inputs are generally easier to measure than other factors that affect students' education outcomes, and so data on inputs are more likely to be collected than are data on other factors. There are a wide variety of school input interventions across the globe. Further, in some cases it may be that a combination of different inputs are required for effective learning, in which case the marginal effect of any particular input on increasing test scores and time in school would vary according to the levels of other inputs.

Figure 4 shows the distribution of high quality studies that examined school input interventions. While these studies are from about 20 countries, they are concentrated in four countries: Chile, China, India and Kenya.

Figure 4 Distribution of School Input Intervention Studies

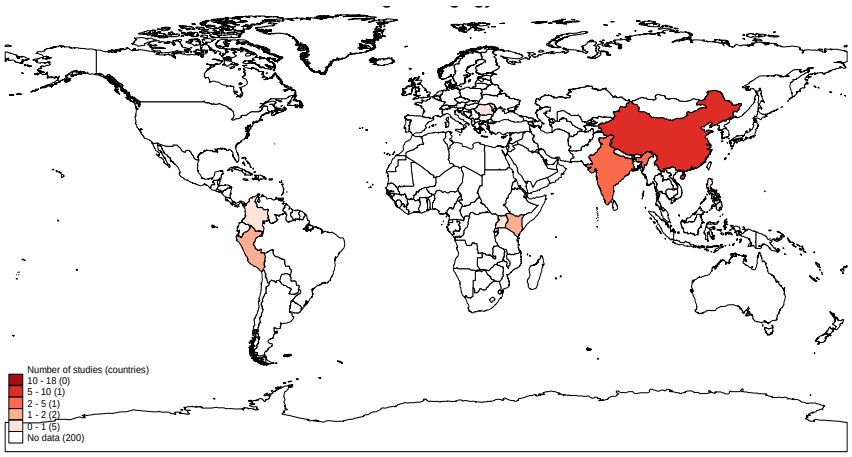


C Challenge 3: Ineffective Pedagogy

A critical determinant of the extent to which increases in schooling inputs translate into improved learning outcomes is the way in which these inputs are used in practice, and the way in which teaching and learning is organized. In many developing countries, pedagogical practices have changed very little in the past several decades. Current pedagogy often consists of teachers using a "lecturing" style, with students expected to follow the textbooks. There is relatively little scope for differentiating instruction to account for the actual level, and variation, of student ability and preparation within a classroom. Teaching effectively may be particularly challenging in many developing country contexts because of the high variation, relative to developed countries, in the initial preparation of children when they enter school.

Fortunately, there are many recent high quality studies that have focused on improvements in teacher pedagogy. Figure 5 presents the distribution of countries for which there is high quality evidence on pedagogy interventions. In comparison to demand, input and governance interventions, the number of studies across 9 countries is limited, and most are from either China or India.

Figure 5 Distribution of Teaching Pedagogy Intervention Studies

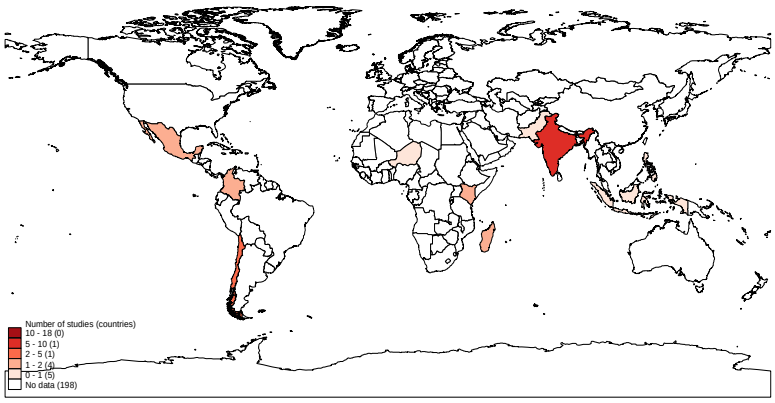


D Challenge 4: Low Quality School Governance

The fourth critical determinant of education outcomes – in addition to policies to stimulate household demand for education, adequate provision of effective school inputs, and improvements in pedagogical practices – is the quality of governance of the education system. Governance refers to a broad set of characteristics of education systems that determine how efficiently they are managed. These include goal setting, personnel policy (hiring, training, retention, and promotions), accountability and monitoring, and performance management. More broadly, governance also includes decentralization, the extent of choice and competition in school "markets", and the regulatory structure for private schools.

Figure 6 illustrates the distribution of countries for which there is high quality evidence. Of the 24 studies in 11 countries, about one third were conducted in India while none were conducted in China.

Figure 6 Distribution of School Governence Intervention Studies



V. Analysis of Interventions that Increase Time in School

This section reviews the evidence on which interventions are most effective at improving educational outcomes. The results are organized around the two broad goals of increasing time in school and increasing student learning. Under each goal, we apply the rubric developed in Section III to identify which interventions work (or do not work). We do this separately for each of the four broad categories of interventions: increasing demand, increasing school inputs, improving pedagogy and strengthening school governance.

We begin the review by considering the impact of education interventions on students' time in school, which includes daily attendance, enrollment, years/levels of completed schooling, and dropping out. We do this for each of the four broad types of education interventions, beginning with interventions that are intended to increase households' demand for schooling.

A Interventions to Increase the Demand for Schooling

This subsection reviews the evidence concerning the impact on time in school of interventions intended to increase households' demand for schooling. Demand interventions are likely to be most effective in contexts where households face major barriers to sending their children to school. Table 4 summarizes the evidence.

Table 4. Effects of Demand Side Interventions on Time in School

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Interventions that work					
Conditional cash transfers					
RCTs	0 (0)	1 (1)	4 (3)	34 (15)	15
non-RCT	0 (0)	0 (0)	0 (0)	18 (9)	9
Interventions that often work					
None					
Promising interventions that need more evidence					
Merit-based scholarship (RCTs)	0 (0)	0 (0)	1 (1)	3 (2)	2
Inform. on returns to educ. (RCTs)	0 (0)	1 (1)	1 (1)	2 (1)	2
Labeled cash transfer	0 (0)	0 (0)	0 (0)	2 (1)	1
Bicycle program (non-RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Matching remittances funds for education (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
School counseling (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Unconditional cash transfers (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Interventions that do not work					

None					
Interventions with insufficient research					
Eliminating school fees (non-RCT)	0 (0)	2 (2)	1 (1)	1 (1)	2
Career counseling (RCT)	1 (1)	0 (0)	0 (0)	0 (0)	1
Mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Mother literacy class (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Adult Literacy program (non-RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Female sanitary products (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Interventions that work

Conditional cash transfers (CCTs) provide monetary payments to parents if their children are enrolled in school and have a high rate of attendance (usually 80 or 85 percent). The most well-known CCT program is Mexico's PROGRESA program, but CCT programs are common throughout Latin America and are now being implemented in Africa and Asia as well. These programs have consistently been shown to be highly effective at increasing enrollment and attendance in school. As seen in Table 4, there are 24 RCT and high-quality studies included in our review, and among the 57 estimates from these studies the overwhelming majority (52) provide positive and statistically significant estimates indicating that CCTs increase students' time in school.

One of the studies with insignificant results came from the Philippines (Chaudhury, Friedman and Onishi, 2013), revealing a potential caveat for future interventions. The authors find that the program was less effective in addressing school enrollment in older children because both the direct cost and the opportunity cost of schooling were higher. To address low enrollment in different age-cohorts, programs should consider whether the value of the cash transfer should vary in contexts where the direct and opportunity costs of schooling differ across age groups. However, the advice to increase the size of the transfers for older children also reveals the primary criticism of CCTs: they are expensive to implement. The issues of costs is taken up below in Section VII

Promising interventions that need more evidence

There are seven interventions that appear to be promising in terms of their impacts on students' time in school. The first of these is the provision of merit-based scholarships. Two studies from the same experiment in Kenya have examined the impact of merit-based scholarships on students' time in school. First, Kremer, Miguel and Thornton (2009) conducted an experimental evaluation of a scholarship program for rural Kenyan girls in grade 6 (Kenya's primary schools enroll students from grade 1 to grade 8). At the beginning of the school year, grade 6 girls were told that those who scored in the top 15% on end-of-year exams would be given approximately $6.40 for each of the next two years (grades 7 and 8), which was enough to cover school fees. In addition, they were informed that their parents would be given an amount of money equal

to $12.80 for each of the next two years. The authors found that the program significantly increased participation (daily attendance, where girls who drop out of school have a zero attendance rate) by 3.2 percentage points. In the second study, Friedman et al. (2011) examined the educational outcomes of the same girls 4-5 years after the original program started (about two years after the program had ended). They found that the program had significantly positive impacts on enrollment in secondary school (8.6 percentage point increase) and current enrollment in any school (7.9 percentage point increase). While these results are promising, more research is needed for male students and for other countries where time in school is a concern.

A second promising demand side intervention is providing information on (estimated) returns to schooling. In the Dominican Republic, Jenson (2010) focused on the low rates of secondary school completion, which were believed to be driven by the assumption among students that the returns to additional years of schooling were very low. Jenson provided information on the estimated returns to schooling to randomly selected grade 8 boys in poor households. He found that the boys who received the information were 4.1 percentage points more likely to be in school one year after receiving the information, a result that is statistically significant at the 10% level. Moreover, four years later he found that boys who had received the information had completed, on average, 0.2 more years of schooling, which was significant at the 5% level. The effects were strongest for the least poor of those in poor households, and weakest for the poorest households, which suggests that the latter may face other barriers, such as credit constraints. However, in a second study Loyalka et al. (2013) found no impact on providing information on the returns to schooling to grade 7 students in China. The authors suggest two reasons why the intervention may have been ineffective: 1. Students may have felt that the quality of their schools was too low for their years of schooling to have a high return; and 2. Financial difficulties remained a barrier to additional schooling. Thus, while this intervention may be promising for secondary students, more research is needed on the extent to which, and under what conditions, this intervention is effective.

There were five other interventions that also appear promising but require additional studies. Two of these considered different types of cash transfers. First, a “labeled” cash transfer program in Morocco was examined by Benhassine et al. (2015). It provided monthly cash

payments of $8 to $13 to primary school students' parents, with higher amounts for higher grades. The payments were "labeled" as assistance for costs of education, but there was no formal requirement that students enroll or attend school regularly for their parents to receive the payments. The authors find that the program increased the enrollment rate by 7.4 percentage points (average over boys and girls). In a similar program that targeted secondary girls currently enrolled in school in Malawi, Baird, McIntosh and Ozler (2011) studied a cash transfer program that did *not* condition the transfers on students' educational outcomes. They found that this *unconditional* cash transfer program increased the number of terms that the girls were enrolled over the next two years from 4.79 to 5.02 terms. This increase of 0.23 terms is equivalent to only a little less than one month (a term last for approximately three months).

Another promising intervention that also targeted secondary school girls, this time in India, was examined by Muralidharan and Prakash (2013). The program offered families funds to purchase bicycles so that their secondary school daughters could ride them to attend school. The authors found this program increased secondary school enrollment on average by 5.2 percentage points, and for girls who lived more than three kilometers from the nearest school, the impact was about 9 percentage points. These are very large impacts given that the initial enrollment rate for these girls was only 17.2 percent.

Turning to Latin America, Ambler, Aycinena and Yang (2014) studied a program that matched remittances sent by Salvadoran migrants in the U.S. to students of their choosing in El Salvador, but only if those remittances are committed to educational purposes. The authors found that when the match amount is three times the amount of remittance, the program increased enrollment in a *private* school by a statistically significant 10.9 percentage points. Note, however, that the overall increase in enrollment of 3.1% was not statistically significant.

The final promising demand-based intervention is a school counseling intervention in China that was evaluated by Huan et al. (2014). This program was designed for students in grades 7 and 8 who were preparing to take the upper secondary school entrance exam. The main goal of the counseling was to reduce students' anxiety. The authors found that it reduced the dropout rate by about 2 percentage

points in the first half of the school year, but there was no effect at the end of the school year.

Interventions with insufficient research

There are seven demand side interventions for which only a little is known about their impacts on students' time in school. The first of these was an attempt to reduce the direct cost of schooling. Indeed, in the past 10-20 years many developing countries, especially in Sub-Saharan African, have eliminated school fees at the primary level. There are two studies that examined the effect of such a program on students' time in school. First, Yi et al. (2014) studied a program in China that provided an early commitment for financial aid for the next level of schooling for grade 7 and 9 students. The program weakly increased the percentage of grade 9 students who matriculated to high school, but there were no impacts for grade 7 students' dropout rates. Second, Borkum (2012) examined the impact of eliminating school fees in South African primary and secondary schools that served poor populations. Unlike the results for CCT programs, he found no effect of the elimination of school fees on enrollment at either the primary or the secondary level.

Next are five interventions that attempted to increase children's time in school by providing information or training to parents. For three interventions from the same study in India, which provided a mother literacy program and a mother class on how to develop their child's learning, Banerji, Berry and Shotland (2013) found no impact on children's time in school (as measured by enrollment and daily attendance). In a similar vein, Handa (2002) examined the impact of a literacy program for adults in Mozambique on their children's enrollment rate; he found no effect. Also, Loyalka et al. (2014) studied an information-based initiative that provided four, 45-minute career counseling sessions to grade 7 and 8 students. This program had a significantly negative impact on time in school; the authors suggest that this may reflect that students learned that upper secondary and post-secondary entrance requirements were more difficult than they had previously thought.

Finally the non-availability of female sanitary products is often posited as a factor inhibiting girls' school participation after puberty. Oster and Thornton (2011) evaluated the impact of an intervention that provided female sanitary products to girls of secondary school age in Nepal, but found that the program had no

effect on the daily attendance of secondary school girls. This may have been because only 1% of girls actually stated that the lack of sanitary products affected their school attendance.

B Interventions that Provide School Inputs

Schools that are adequately, or even well, equipped may be more attractive to students (and their parents) and thus may increase these students' time in school. Many high quality studies have examined whether this is the case, and Table 5 summarizes the results of studies that have estimated the impact of increased school inputs on measures of time in school. The following paragraphs describe these results in more detail.

Table 5. Effect of School Inputs on Time in School

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Interventions that work					
Building new schools					
RCTs	0 (0)	0 (0)	0 (0)	4 (3)	3
non-RCT	0 (0)	0 (0)	0 (0)	3 (3)	3
Interventions that often work					
School meals					
RCTs	0 (0)	0 (0)	1 (1)	2 (2)	3
non-RCT	0 (0)	3 (2)	1 (1)	0 (0)	2
Promising interventions that need more evidence					
Extra teacher/materials (non-RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Deworming medicine (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Support circles (RCT)	0(0)	0 (0)	0 (0)	1(1)	1
Take home rations (RCTs)	0 (0)	0 (0)	1 (1)	1 (1)	2
Multi-level teaching materials (RCT)	0(0)	0 (0)	0 (0)	1 (1)	1

Interventions that do not work					
None					
Interventions with insufficient research					
Textbooks (RCTs)	0 (0)	0 (0)	2 (2)	0 (0)	2
Provision of libraries (RCT)	0(0)	0 (0)	1 (1)	0(0)	1
Health insurance (non-RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
School infrastruct. investment (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Multi-level teaching materials & parent teacher partnership (RCT)	0(0)	0(0)	1 (1)	0 (0)	1
School feeding /parent-teacher partnerships* (RCT)	0 (0)	0(0)	1 (1)	0 (0)	1
Providing school uniforms (RCT)	1(1)	0 (0)	0 (0)	0(0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Input interventions that work

For communities with an insufficient number of schools, or no schools at all, six studies have examined the impact of building new schools on time in school. Each of these studies examined a different country, so evidence is available from Afghanistan, Burkina Faso, Indonesia, Mozambique, Niger and Pakistan. Building new schools reduces an important *indirect* cost of attending school, the distance to the nearest school. More time spent traveling to school is time lost that could have been used for work or other activities, and greater distances may increase transportation costs and worries about safety. All estimates from these six studies show significantly positive impacts from building new schools on students' time in school.

In the earliest study, Duflo (2001) examined the impact of a massive school construction program in Indonesia in the 1970s on the years of schooling of boys born between 1950 and 1972. She found that an additional school built per 1000 school-age children increased years of education by 0.2 years. Next, Handa (2002) estimated the impact of constructing new primary schools in Mozambique and found that the marginal probability of enrollment of boys and girls increased by 0.3% for each new primary school built within an "administrative post" area (these are relatively large areas, with on average 21 primary schools). Third, Alderman, Kim and Orazem (2003) conducted an RCT in Pakistan that provided funding to construct new (or support existing) private girls' primary schools; the support of existing schools made them affordable to poor families. In urban areas, the program increased the enrollment rate for girls by 25 percentage points, while in rural areas the enrollment rate increased by 15 percentage points.

Turning to more recent studies, Dumitrescu et al. (2011) evaluated the IMAGINE program in Niger, which included the construction of a school and complementary interventions (such as teacher training).[11] When IMAGINE schools added to or replaced existing school structures, the program had a positive impact on enrollment. In a recent study on Afghanistan, Burde and Linden (2013) examined the impact of the opening of primary schools on children of primary school age in rural villages without a school. In Ghor province, where this intervention took place, only 29% of

11 Most of the complimentary interventions were not implemented, so the main impact of the program was to build more schools.

families lived within 5 kilometers of a primary school in 2007. The program increased the enrollment rate by 51.5 percentage points for girls and by 34.6 percentage points for boys. The higher impact on girls likely reflects that they are often not allowed to travel to neighboring villages, which in most cases is necessary to enroll in school. Finally, Kazianga et al. (2013) evaluated the impact of providing "girl friendly" schools in rural villages in Burkina Faso (BRIGHT program). These schools have amenities that are particularly attractive to girls, such as sources of clean water and separate latrines for boys and girls. Overall, opening these schools increased the enrollment rate of all children (average over boys and girls) by 18.5 percentage points, and separate estimates show increases of 16.3 percentage points for boys and 21.9 percentage points for girls.

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

Input interventions that often work

Child malnutrition is a common problem in many developing countries, and there is a large amount of evidence that well-nourished children have better educational outcomes (Glewwe and Miguel, 2008). Thus many developing countries have implemented programs that provide meals to students and/or their families as well. The most common type of program that attempts to improve students' nutritional status is the provision of school meals. Five studies in five different countries (Burkina Faso, Chile, India, the Philippines and Uganda) have estimated the impact of school meals on time in school. Of the seven estimates, two are statistically significant and five are statistically insignificant. One estimate from Burkina Faso finds a significantly positive impact on enrollment of children 6-15 years old (Kazianga, de Walque and Alderman, 2012). Another estimate from Alderman et al. (2012) for Uganda is also statistically significant. This evidence indicates that in some, but certainly not most, cases school meals increase students' time in school. Further research is needed to understand what conditions are needed for school meal programs to be effective.

Promising input interventions that need more evidence

There are six promising interventions that increase time in school by providing a school input: providing an extra teacher and materials; administering deworming medicine to school children; community support circles; take home rations; providing school meals; and multi-level teaching materials. All of these interventions had one or more positive and statistically estimate.

First, Chin (2005) evaluated a program that provided extra teachers *and* additional educational materials (including blackboards) to very small primary schools in India. She found that the program significantly increased students' primary school completion rates (by one to two percentage points), but it is not possible to determine how much of this effect is due to the extra teacher and how much is due to the additional educational materials. Second, in a separate study Tan, Lane and Lassibille (1999) evaluated a number of interventions in the Philippines, one of which was providing multi-level teaching materials that were designed to help teachers pace their teaching according to differing abilities of their students. This intervention was found to be effective at reducing the dropout rate for primary school children.

Third, many individuals in developing countries have helminth (worm) infections of various types (roundworm, hookworm, whipworm and schistosomiasis). School-age children are particularly vulnerable to these types of infections, which lead to anemia and other problems that may reduce children's attentiveness in school. Miguel and Kremer (2004) implemented an RCT to estimate the impact of providing deworming medicine to primary school students in rural Kenya; they found that it reduced their absence rate by about 7-8 percentage points.

Fourth, in Malawi, Pridmore and Jere (2011) study a "Circles of Support" program that mobilizes networks of family, friends and neighbors to support vulnerable learners (children in regions with high HIV prevalence rates). The program significantly reduced the dropout rate.

Finally, two studies found mixed results for the impact of take-home rations on students' education outcomes. Adrogue and Orlicki (2013) found no impact of such rations on student attendance in Argentina, but Kazianga, de Walque and Alderman (2012) found a significantly positive 4.8 percentage point impact on the enrollment of children age 6-15 in Burkina Faso.

Interventions with insufficient research

There are seven other school input interventions, all of which have inconclusive results, as shown at the bottom of Table 5. These interventions are: 1. Provision of textbooks; 2. Provision of school libraries; 3. Student health insurance; 4. School infrastructure investments; 5. Provision of multi-level teaching materials with parent-teacher partnerships; 6. School meals combined with parent-teacher partnerships; and 7. Provision of school uniforms. While these interventions are not particularly promising, more studies are needed in order to assess their effectiveness at increasing students' time in school. To provide an example of why more information is needed, as opposed to classifying these interventions as "what does not work", consider the case of providing textbooks. Surprisingly, neither textbook program improved attendance.[12] However, both interventions had major flaws. Sabarwal et al. (2014) find no impact on test scores from the provision of textbooks to schools in Sierra Leone and attribute this to the fact that schools stored the textbooks instead of distributing them to students (a particularly egregious example of poor implementation). Glewwe et al. (2009) also find no impact on test scores from providing textbooks to students in Kenya. But they do find positive impacts on students with the highest baseline test scores, and they present evidence to support to argument that their results are consistent with the fact that the majority of children could not read the English language textbooks to begin with, and thus could not benefit from the textbooks (whereas those who could read them *did* benefit).

C Pedagogy Interventions

Once children are enrolled in school and their schools have sufficient inputs, their progress in school is likely to depend very much on the quality and appropriateness of their teachers' pedagogical practices. We now turn to the impact of interventions that are intended to change pedagogical practices on students' time in school. Table 6 shows that only two studies have evaluated the impact of changes in pedagogy on increasing measures of time in school. Unfortunately, neither study provides encouraging evidence.

12 As will be seen below in Section VI, they also did not increase test scores.

Table 6. Effect of Pedagogy Interventions on Time in School

	Negative, Signi-ficant	Neg, Insigni-ficant	Positive, Insigni-ficant	Pos., Signi-ficant	Total Studies
Interventions that work					
None					
Interventions that often work					
None					
Promising interventions that need more evidence					
None					
Interventions that do not work					
None					
Interventions with insufficient research					
Computer, electronic fames, and access to technology (RCT)	0 (0)	0 (0)	2 (1)	0 (0)	1
Teaching at the right level (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Interventions with insufficient research

Only two studies consider the impact of pedagogical interventions on measures of time in school. Banerjee et al. (2010) combined tutoring with a community information campaign in India, and Cristia, Czerwonko and Garofalo (2014) examined the impact of an intervention Peru that increased access to computers and the internet to students in public secondary schools in urban areas of Peru. Neither study provides evidence that these innovations in pedagogy affect enrollment, attendance or dropout rates. Given the scarcity of research on the impact of pedagogy on measures of time in school, we have little basis on which to recommend, or not recommend, pedagogical innovations as a mechanism for increasing time in school.

However, as will be seen in subsection VI.C, there is much more evidence showing impacts of pedagogy interventions on student learning. Indeed, both of the above interventions are shown below to have a positive impact on test scores, which suggests that, from a policy perspective, some interventions are inherently better suited for increasing learning outcomes than they are for increasing time in school.

D Interventions that Change School Governance

A final general type of education intervention are inventions aimed at changing how schools are organized and operated, which is generally referred to as "governance". Table 7 illustrates that there is very little evidence to date on what changes in school governance can increase the time that children spend in school. The following paragraphs explain this in more detail.

Table 7. Summary of Impacts on Time in School of Governance

	Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total Studies
Interventions that often work					
Private schools (vouchers)					
RCTs	0 (0)	1 (1)	1 (1)	1 (1)	2
non-RCT	0 (0)	1 (1)	0 (0)	1 (1)	2
Promising interventions that need more evidence					
Single sex school (non-RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Interventions that do not work					
Monitoring (RCTs)	0 (0)	1 (1)	1 (1)	0 (0)	2
School-based management					
RCTs	0 (0)	7 (3)	5 (3)	1 (1)	3
non-RCT	0 (0)	2 (1)	1 (1)	0 (0)	2
Interventions with insufficient research					
Teacher performance pay (RCT)	0 (0)	1 (1)	2 (1)	0(0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Interventions that often work

An important trend in primary education in developing countries over the past two decades has been the rapid growth of private schools, with recent estimates showing that private schools now account for over 20% of total primary school enrolment in low-income countries (Baum et al. 2014). This phenomenon is surprising because it is occurring within a context of increased spending on public education and near universal access to free public primary schools. Opponents of private schooling argue that it is much less accessible to children from lower income households and that it weakens the public education system. In contrast, supporters of private schools contend that private schools in developing countries have grown in response to failures of the public schooling system, that they are more accountable and

responsive to parents, and that the revealed preference of parents suggests that they are likely to be better than public schools.

Four studies examined the impact of vouchers on time in school. Of the five estimated impacts, two are significantly positive and the other three are statistically insignificant. Angrist et al. (2002) and Angrist et al. (2006) study the short and medium term effects of the PACES program in Colombia, which provided vouchers (allocated by lottery) to students to attend private secondary schools.[13] Angrist et al. (2002) found no significant impacts of the vouchers on enrollment or time in school, yet Angrist et al. (2006) found that the voucher winners have significantly higher high-school graduation rates (5.6 percentage points higher on a base of 25 to 30 percentage points). While both of these studies suggest that the PACES voucher program was effective, these results may not reflect only the differential "productivity" of private schools because the PACES program also allowed the vouchers to be topped up by parents (to attend a better school than they could have afforded without a voucher), and required students to maintain minimum academic standards to continue receiving the voucher. Thus while the results reflect a combination of private school productivity, additional education spending, and student incentives.

There are also two non-RCT studies of the impact of going to a private school. First, Hsieh and Urquiola (2006) examined the impact of using vouchers to enroll in a private school. They found no significant impacts on time in school. Second, Barrera-Osorio and Raju (2011) found a significant increase in enrollment from a program in Pakistan that offered a subsidy to private schools per student enrolled.

Promising interventions that need more evidence

There is one governance intervention that may lead to increased time in school, single-sex schools, but since there is only one study to date there is insufficient for drawing firm conclusions. More specifically, Jackson (2012) estimated the impact of single-sex schools on enrollment (as measured by taking a national examination) of grade 10 students in Trinidad and Tobago. He found that boys and girls who

13 We classify these two studies as RCTs because they are based on an actual lottery to determine eligibility for vouchers. One could also classify them as "natural experiments".

attend such schools are 7.3 percentage points more likely to take the national exam, and thus he concludes that such schools increase time in school.

Interventions that do not work

There are two school governance interventions that do not appear to increase students' time in school: monitoring teachers and community based school management. Both of these interventions are thought to be potential remedies for a common problem in many schools in developing countries, that of teacher absences. The most basic policy tool to reduce teacher absence is to increase the extent of monitoring and oversight of schools. This can include administrative (top-down) monitoring as well as community-based (bottom-up) monitoring. Two studies in India, Duflo, Hanna and Ryan (2012) and Banerjee et al. (2010), have considered the impact of teacher monitoring. While the monitoring may have worked to reduce teacher absences, these studies found no impact on student attendance.

Another approach to improve monitoring and thus the accountability of schools and teachers is to decentralize more management authority to schools and communities – an approach that is broadly referred to as "school-based management" (SBM). The theory of change associated with this approach is to empower communities to take charge of their schools and in particular to make teachers more accountable to them. Several reforms based on this approach have been attempted around the developing world, but the empirical evidence suggests that it is ineffective at increasing time in school; as seen at the bottom of Table 7, 15 of the 16 estimates from five studies are statistically insignificant.

Interventions with insufficient research

There is one last governance intervention that may work to increase time in school, but there is too little evidence to be sure: teacher pay based on student performance. A common feature of teacher pay in developing countries is the use of fixed salary schedules with little or no possibility of higher pay if teachers improve their teaching performance. Since teacher effort is a key determinant of education quality, a natural set of policy options to enhance governance is linking compensation of teachers to measures of their performance. However, it can be difficult to measure individual teacher

productivity, and learning outcomes that are measurable may swing the focus away from learning outcomes that are not easily measured.

Glewwe, Ilias, and Kremer (2010) evaluated a teacher incentive program in Kenya that provided school-level *group* incentives using prizes for high-achieving schools. The prizes were awarded, using a tournament design, to the schools that had the best average student test scores and also to those that had the highest average improvements. The authors found no impacts on reducing student absence rates. It may be that the group-nature of the incentive program (across 12 teachers) induced free riding and weakened the incentives faced by individual teachers. While performance pay based on individual teachers' performance may be more effective, to date no studies exist of the impact of such a scheme on students' time in school.

Summary of What Works for Increasing Time in School

Here we summarize the findings of this section. A graphic summary is also provided in Figure 7.

Interventions that work: The two most effective interventions to increase time in school in communities where attendance and/or enrollment are low are: (1) Provision of conditional cash transfers; and (2) Building new schools where local access is lacking. Both of these interventions reduce the "cost" of attending school for students. However, both have been shown to be relatively expensive interventions; this is discussed further in Section VII.

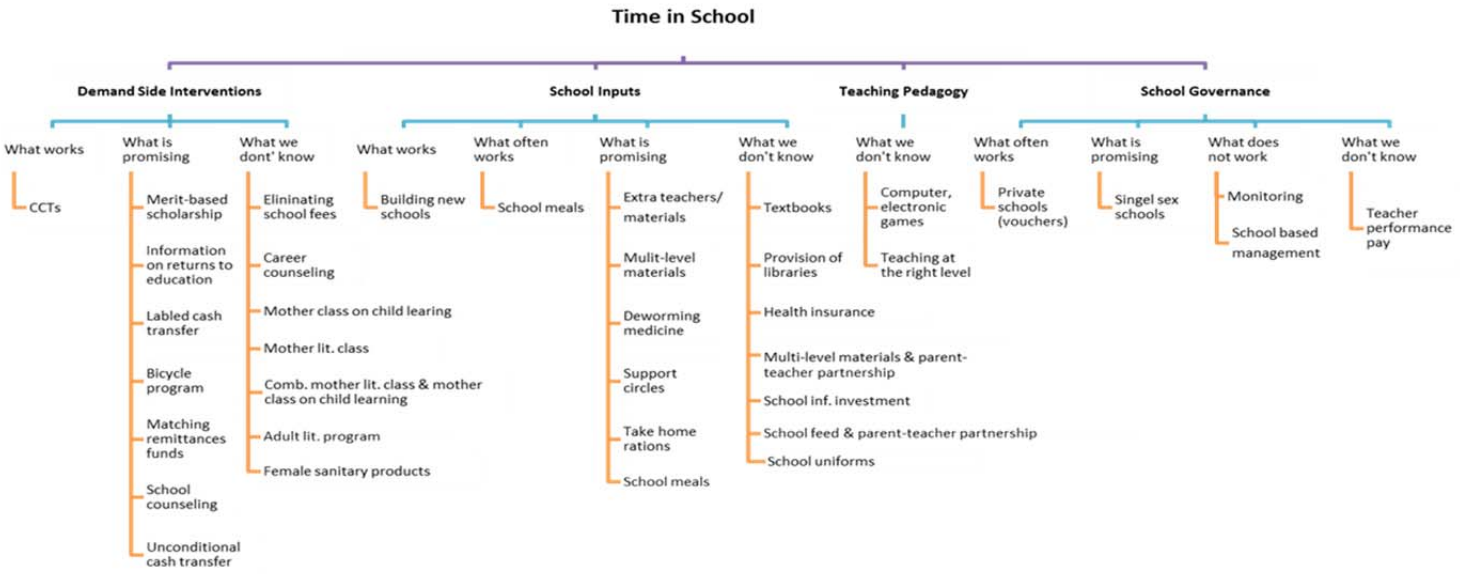
Interventions that often work: The one education intervention that appears to work at least some of the time to increase students' time in school is vouchers to attend private schools.

Promising interventions that need more evidence: There are many promising interventions when communities face a problem of low student attendance and enrollment. These include: (1) Providing information directly to students and parents on the returns to education; (2) Providing school counseling to students; (3) Merit-based scholarships; (4) Labeled cash transfers; (5) Unconditional cash transfers; (6) Funds for bicycles for student transportation; (7) Matching remittances for educational purposes; (8) Extra teachers and teaching materials; (9) Food-based interventions (school meals and take-home rations); and (10) Deworming. We recommend that more

impact evaluations on these interventions be conducted to confirm (or refute) their promising nature.

Interventions that do not work: Finally, there are two interventions that are generally ineffective for the purpose of increasing students' time in school. Changes in governance structures that either increase monitoring of teacher performance or promote school-based management are ineffective in terms of their ability to increase students' time in school.

Figure 7. Summary of Intervention Effectiveness for Time in School



VI. Analysis of Interventions that Improve Learning Outcomes

We now turn to impacts of education interventions on students' learning outcomes, as measured by test scores. We do this by considering each of the four broad types of education interventions, beginning with interventions that are intended to increase households' demand for schooling.

Table 8. Effect of Demand-side Interventions on Test Scores

	Negative, Significant	Negative, Insignificant	Positive, insignificant	Positive, Significant	Total Studies
Interventions that work					
Merit-based scholarship (RCTs)	0 (0)	1 (1)	0 (0)	6 (4)	4
Interventions that often work					
Conditional cash transfers					
RCTs	0 (0)	1 (1)	1 (1)	6 (3)	5
non-RCT	0 (0)	3 (2)	3 (2)	0 (0)	3
Promising interventions that need more evidence					
Mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Interventions that do not work					
None					
Interventions with insufficient research					
Labeled cash transfer (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Mother literacy classes (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Unconditional cash transfers (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1

Promise of high school financial. aid (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Inform. on returns to schooling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Career counseling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

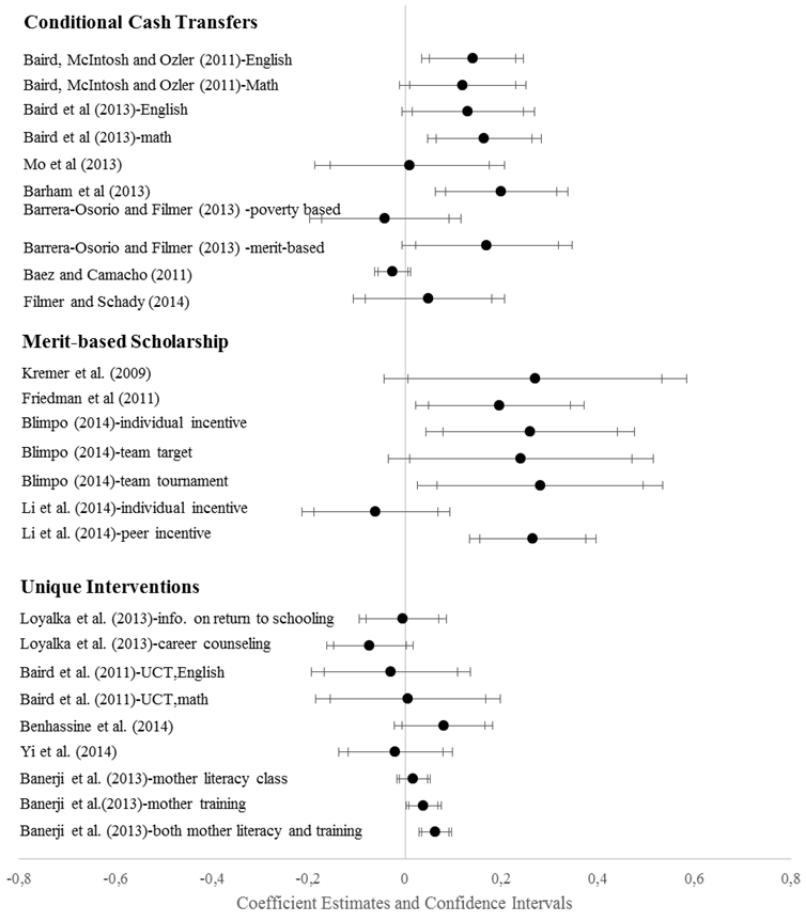
A Interventions to Increase the Demand for Schooling

Several education interventions that are intended to increase the demand for schooling have been shown to increase student learning. These include merit-based scholarships and conditional cash transfers (CCTs), although the evidence is less strong for the latter. Two other interventions –classes for mothers on child learning, and those classes combined with mother literacy classes – have been shown in a single study to increase test scores. Six other types of interventions are less promising but there is too little evidence for drawing firm conclusions. These results are summarized in Table 8. The following paragraphs describe these findings in more detail.

In addition to the summary of demand side estimates on test scores in Table 8, since test scores can be compared by expressing the size of impacts in terms of the standard deviation[14] of the distribution of test scores, we also summarize the individual estimates for the results in Table 8 in Figure 8. This figure, as well as Figures 9 – 11, report the point estimates as well as both the 90 and 95th percent confidence intervals for each paper by category. More specifically, Figure 8 reports all coefficients point estimates and confidence intervals for the individual studies included in the analysis of conditional cash transfers, merit-based scholarships, and unique interventions.

14 See "Key Definitions and Acronyms" for a definition of commonly used statistical and technical terms used in this report.

Figure 8. Estimated Impacts of Demand Side Interventions on Test Scores



Interventions that work.

Four randomized controlled trails (RCTs) have examined the impact of providing scholarships awarded on the basis of students' academic performance on student learning as measured by test scores. The findings are almost unanimous: provision of merit-based scholarships significantly improves student learning.

Three interventions conducted in Africa were found to be very effective. The two studies of the Kenya scholarship program discussed above in Section V (Kremer, Miguel and Thornton, 2009;

Friedman et al. 2011) found significantly positive estimates for test scores. More recently, Blimpo (2014) examined a program providing three different types of scholarships: scholarships based on individual-level performance with respect to a set goal, with no limit on the number of scholarships offered; scholarships based on average performance for (randomly assigned) teams of four students, again with respect to a set goal and no limit on the number of scholarships; and a "tournament" in which 84 teams of four students each (randomly assigned) from 28 schools competed for a large prize that was given only to the three top performing teams. For the first two types, the payments were $10 per person ($40 for a team of four) for a relatively low level of performance, and $40 per person ($160 for a team of four) for a high level of performance. For the third, the prizes were much larger, at $640 for each of the top three teams. All three types of incentives had similar (and statistically significant) impacts, increasing grade 10 test scores by 0.24 to 0.28 standard deviations of the distribution of test scores (hereafter denoted by σ).

In contrast, a tournament-based scholarship program in China, evaluated by Li et al. (2014), found that an individual incentive intervention in China had no statistically significant impact. However, the authors find that combining student incentives with peer tutoring (where academically higher achieving students were paired with lower achieving ones and both students were rewarded for improvements of the lower achieving students) increased the test scores of the weaker students by 0.27σ. Thus, it is possible that student incentives on their own may not be effective unless they are accompanied by some type of pedagogical support.

Interventions that often work.

Three studies of conditional cash transfer (CCT) programs found significantly positive impacts on test scores. In a study conducted in Malawi, Baird, McIntosh and Ozler (2011) and the Baird et al. (2013) evaluated two different types of recipients of the same program: girls (age 13-22) already enrolled in school at the start of the program and girls (age 13-22) *not* enrolled in school at the start of the program. For girls already enrolled in school at the start of the program, Baird, McIntosh and Ozler (2011) found that the program significantly increased English and math scores by 0.14σ and 0.12σ, respectively. Baird et al. (2013) examine the impact of the program on girls who were *not* in school when the program began and found even stronger

results. The impacts for these girls were slightly higher, 0.13σ and 0.16σ for English and math, respectively.

There is evidence from Nicaragua that improvements from test scores persist years after the administration of a CCT. The long term impact of exposure to the Nicaraguan CCT program on (former) primary school students' test scores was studied by Barham, Macours and Maluccio (2013). The children were exposed to the program when they were 9-12 years old, but the test score data were collected (for boys only) in 2010, 10 years after the start of the program. The results show that, 10 years after the start of the program, average test scores were 0.20σ higher (0.23σ for home language and 0.17σ for math) for the boys who were exposed longer (five years instead of two years) to the CCT program, and these impacts are statistically significant.

In contrast, other studies have found that CCTs do little to increase test scores, including Baez and Camacho (2011) in Colombia, Mo et al. (2013) in rural China, and Barrera-Osorio and Filmer (2013 in Cambodia. For the first two studies the lack of a positive impact could reflect selection bias in that weaker students, who did not receive CCT payments, may have dropped out of school and so were not tested (which would increase the average test scores of those who remained in school). This is unlikely to be the case for the Cambodia study, however, since students (and former students) were tested in their homes.

Promising interventions that need more evidence.

Two promising interventions come from a mother literacy program in India. One intervention found that a program to train a mother on child learning led to small but statistically significant increases in the test scores (average over literacy and mathematics) of students in grades 1-4 (0.04σ). Further, combining a mother literacy class with this program to train mothers on child learning led to a somewhat higher impact (0.06σ) on student test scores. Note, however, that these impacts after one year of the program are relatively small, as seen in Figure 8.

Interventions with insufficient research.

There are also six interventions with little evidence of any effectiveness because they are based on a single study and the results

were insignificant. Three have positive but insignificant results (labeled cash transfers, mother literacy classes, and unconditional cash transfers) and three have negative, insignificant results (promise of high school financial aid, information on returns to schooling, and career counseling).

B Interventions that Provide School Inputs

Several evaluations have been conducted of the impact on student learning of interventions that provide basic educational inputs on student learning. Table 9 summarizes the results of these evaluations, and the following paragraphs provide a more detailed discussion.

Table 9. Effect of School Inputs on Test Scores

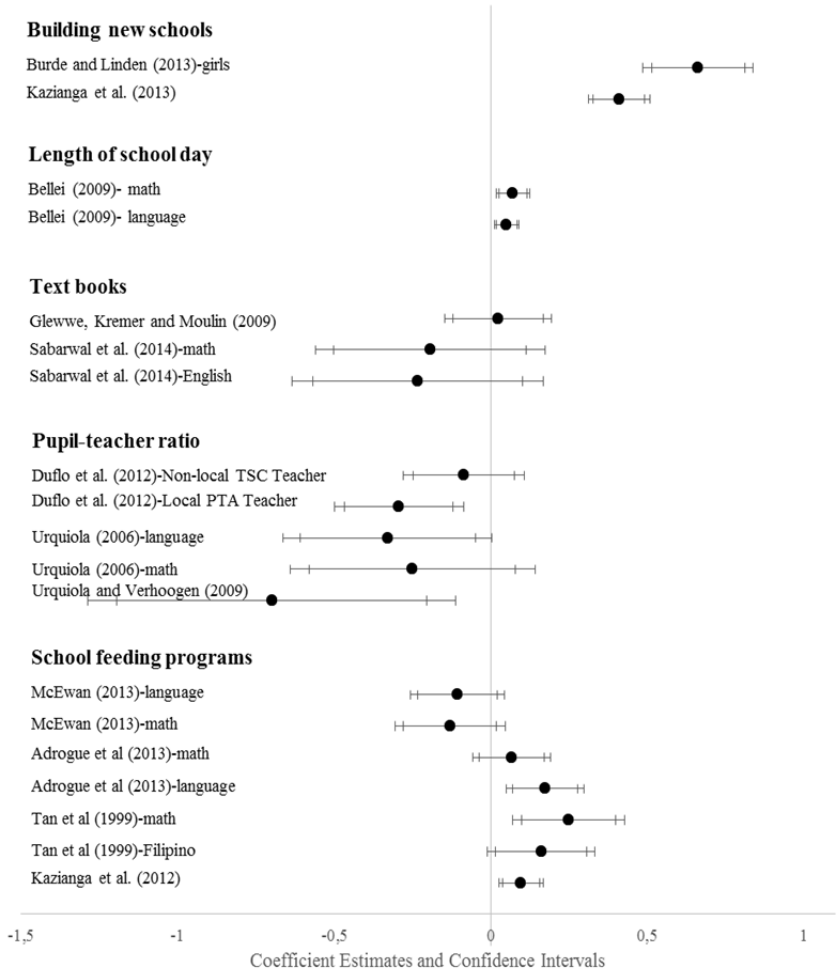
	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Interventions that work					
Pupil-teacher ratio					
RCTs	0 (0)	0(0)	1(1)	0 (0)	1
non-RCT	3 (2)	1 (1)	0 (0)	0 (0)	2
Building new schools					
RCTs	0 (0)	0 (0)	2(1)	2 (1)	2
non-RCT	0 (0)	0 (0)	0 (0)	1 (1)	1
Interventions that often work					
School meals					
RCTs	0 (0)	0 (0)	0 (0)	3 (2)	2
non-RCT	0 (0)	2 (1)	1 (2)	1 (1)	2
Promising interventions that need more evidence					
Hours per school day (non-RCT)	0 (0)	1 (1)	0 (0)	3 (2)	2
Multi-level teaching materials and parent-teacher partnerships (RCT)	0 (0)	1 (1)	0 (0)	2 (1)	1
School feeding /parent-teacher partnerships (RCT)	0 (0)	0 (0)	1 (1)	2 (1)	1
Take-home rations (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Provision of eyeglasses (RCT)	0(0)	0 (0)	0 (0)	1 (1)	1
Unexpected school block grant (RCT)	0 (0)	0 (0)	0 (0)	1(1)	1

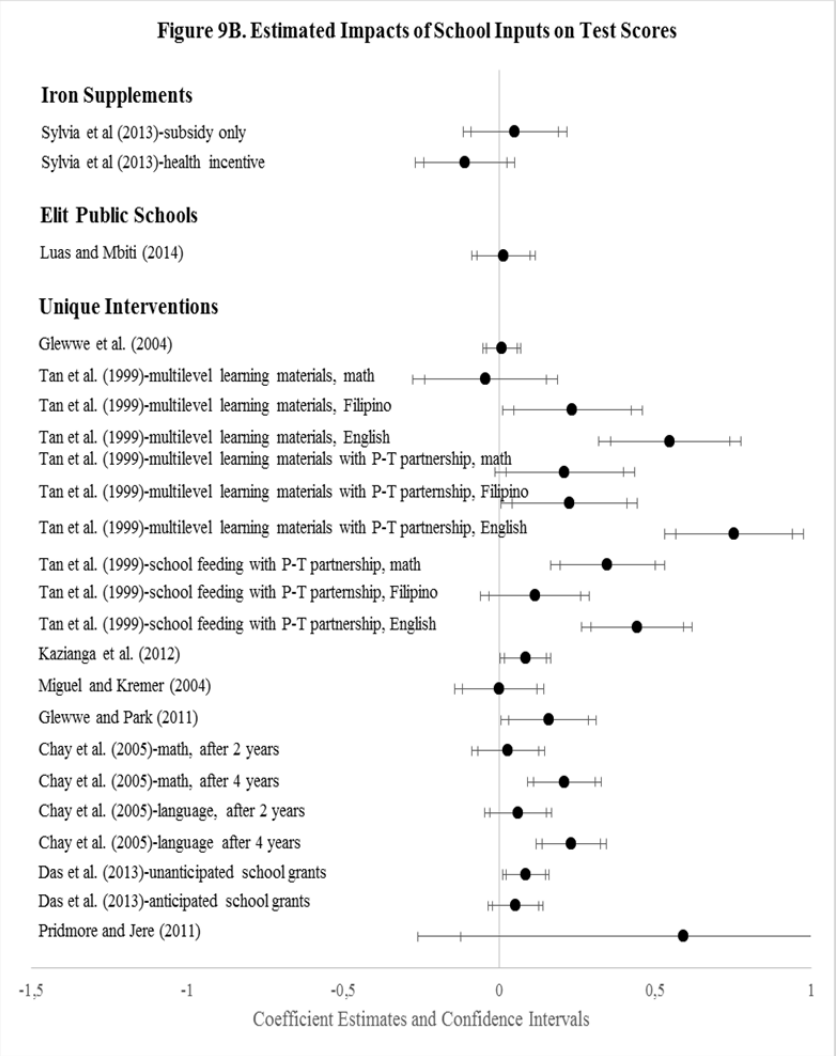
Iron supplements (all RCTs)	0 (0)	1 (1)	2 (2)	1 (1)	2
Attending an elite public school (non-RCT)	0 (0)	0 (0)	1 (1)	1 (1)	2
Infrastructure/materials/training	0 (0)	0 (0)	2 (1)	2 (1)	1
Multi-level learning materials (RCT)	0 (0)	1 (1)	0 (0)	2 (1)	1
Interventions that do not work					
None					
Interventions with insufficient research					
Flipcharts (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Expected school block grant (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Support circles (RCT)	0(0)	0 (0)	1(1)	0 (0)	1
Deworming medicine (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1
Textbooks (RCTs)	0 (0)	2 (1)	1 (1)	0 (0)	2
Provision of libraries (RCT)	1 (1)	1 (1)	0 (0)	0 (0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Figures 9A and 9B provides a summary of the estimates with 90 and 95th confidence intervals by intervention type for school input interventions reported in Table 9.

Figure 9A. Estimated Impacts of School Inputs on Test Scores





Interventions that work.

Perhaps the most basic educational "input" is a teacher, and an obvious way to measure the size of this input is by using the pupil-teacher ratio. Three high quality studies have produced five estimates of the impact of the pupil-teacher ratio on student learning. Intuitively, one would expect that interventions that reduce pupil-teacher ratios would increase learning because smaller classes would lead to more opportunities for teachers to give individual attention to students. Indeed, two different studies found three significantly

negative estimates of the relationship between pupil-teacher ratio and students' test scores, which indicates that reducing pupil-teacher ratios increase test scores. On the other hand, two of the five estimates are statistically insignificant.

The two papers that produce the expected finding that lower pupil-teacher ratio lead to higher test scores are those by Urquiola (2006) and Urquiola and Verhoogen (2009). Urquiola (2006) used the fact that schools in Bolivia with pupil-teacher ratios above 30 can apply to the education authorities for another teacher, and he presents evidence that these schools often do obtain another teacher. He finds that schools that obtain another teacher, which greatly reduces the pupil-teacher ratio, have significantly higher language scores, but the effect on math scores is not statistically significant. Urquiola and Verhoogen (2009) estimate the impact of class size on student test scores in Chile. The focus of the study is on children in grades K-8 in private schools (about half of students in Chile are enrolled in private schools). The authors found that reductions in class size led to significantly positive impacts on both math and language test scores.

Smaller, insignificant impacts were found by Duflo, Dupas and Kremer (2012, 2015), who conducted an RCT in Kenya that randomly assigned some schools an extra contract teacher, and within those schools students were randomly divided into classes that were taught by the current teacher (all of whom were civil service teachers) and those taught by the newly hired contract teacher.[15] When they compared classes of different sizes taught by civil service teachers, they found that although the reduction in class size from about 80 to about 40 led to higher test scores (about 0.09σ), this increase was not statistically significant.

Building schools has also been shown to increase test scores. Three studies have found positive impacts on student learning. Burde and Linden (2013) found that constructing new schools in rural Afghanistan had a large impact on test scores over a period of about six months, generating increases of 0.66σ for girls and 0.41σ for boys. These estimates include all children in those villages, not just those who are enrolled in school. Similarly, Kazianga et al. (2013) found that building new "girl friendly" primary schools in rural Burkina Faso in villages that previous had no primary school increased test scores by

15 The findings of this study on the impact of contract teachers are discussed below in subsection VI.D.

0.41σ. Finally, Dumitrescu et al. (2011) study the impact of new schools in Niger; while the results were positive, they were not statistically significant.

Interventions that often work.

One intervention that often, but not always, increases students' test scores is school meals. Four studies have examined the impact of school meals on student learning. As seen in Table 9, the evidence on the effect of school meals on learning outcomes is somewhat mixed. More specifically, McEwan (2013) found statistically insignificant impacts of a school feeding program on math and language test scores among grade 4 students in Chile. In neighboring Argentina, Adrogue and Orlicki (2013) found a small and statistically insignificant impact of school feeding on the mathematics test scores of students in grade three, but a larger (0.17σ) and statistically significant impact on their language scores. Tan et al. (1999) found significantly positive impacts of a school feeding program in the Philippines on the math (0.25σ) and Filipino (0.16σ) of grade 1 students. Finally, Kazianga, de Walque and Alderman found that school meals increase math scores by 0.10σ in Burkina Faso. These results suggest that school meal programs can, at least in some settings, increase student learning.

Promising intervention that need more evidence.

Table 9 shows ten promising interventions that need more research to understand whether they are generally effective at increasing test scores. More specifically, increasing the number of hours in the school day, provision of multi-level teaching materials, a program that combines school feeding with parent-teacher partnerships, take home food rations, the provision of eyeglasses, iron supplementation, and receiving an unexpected block grant all are promising interventions to improve student learning.

Two high quality studies have produced four estimates of the impact of an increase in the length of the school day on student learning in Chile (Bellei, 2009) and Ethiopia (Orkin, 2013).The evidence from these two studies is generally supportive of the common sense notion that longer school days increase student learning.

Food-based programs also show some promise. In addition to showing that school meals have a positive impact on learning, the Burkina Faso paper by Kazianga, de Walque and Alderman (2012) also shows that targeted take-home ration program increased girls' math scores by 0.08σ, a statistically significant impact. Tan et al. (1999) examined a program that combined school feeding with "parent-teacher partnerships". This combination led to two (out of three) statistically positive impacts on student learning. However, it is not possible to determine the extent to which this is due to the school meals or to the parent-teacher partnerships.

Glewwe, Park and Zhao (forthcoming) conducted an RCT to investigate the impact of the provision of eyeglasses on learning for primary school students in rural China (they did not look at enrollment since it was already very close to 100%). The authors found that the provision of eyeglasses significantly increased average test scores by at least 0.16σ.

Tan, Lane and Lassibille (1999) find that multi-level learning materials, combined with a parent teacher partnership significantly increased test scores of primary students in the Philippines. However, it is not clear whether the success of the program was due to the learning materials, the parent teacher partnership, or both.

There are two Chinese studies that examine the provision of iron supplements, Luo et al. (2012) and Sylvia et al. (2013). Three of the four estimates are positive (one significantly so), providing some initial evidence that iron supplements in China can increase student learning.

Another promising intervention is attending an elite public school. There are two studies (Kenya and Romania) that have examined the rather general policy of being admitted to an elite public school, which could be interpreted as devoting the resources required to transform a typical high school into an elite high school. Many developing (and developed) countries have elite public schools that restrict admission to the best students in the country. These elite public schools have much more qualified teachers and many other types of resources that are not found in a typical secondary school. Student placement in these schools is highly valued by many parents, and the students of these schools often have very successful careers, yet it is not clear that those students' successes were due to their attending those elite schools. Pop-Eleches and Urquiola (2013) examined the effect of "going to a better school" in Romania and they

find that students who are able to get into a better, more academically challenging school perform better on graduation tests. This indicates that "better" schools do lead to greater student learning, but it provides no information on which of the many characteristics that make a school "better" are the ones that bring about this improved performance.

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

Tan, Lane and Lassibille (1999) find that multilevel teaching materials had significantly positive impacts for two of the three tests in the Philippines. However, the multi-level learning materials intervention had many components (several different types of learning materials), so it is not clear which components led to increased student learning.

Interventions with insufficient research.

There are very few randomized evaluations of education interventions that provide relatively simple classroom materials. One of these is by Glewwe et al. (2004), who estimate the impact of classroom flipcharts on learning outcomes. They find that providing flipcharts had no impact on students' test scores.

Das et al. (2013) examine a school block grant program in India. The authors show that when schools received large unexpected grants (about $3 per student), students' test scores increased by about 0.09σ. However, *expected* grants had little or no effect, because households decreased spending on education when they knew that the schools that their child was attending would receive the grants.

Miguel and Kremer (2004) implemented an RCT to estimate the impact of providing deworming medicine to primary school students in rural Kenya. This intervention increased students' time in school, however the authors did not find an impact of the program on students' test scores.

We found only one high quality study of the impact of the provision of school libraries on students' educational outcomes, a study conducted in India by Borkum, He and Linden (2012). The authors examined the provision of both "in school" libraries and traveling libraries. The authors find that "in school" libraries had no effect on students' language scores and the traveling libraries had an unexpected negative effect (-0.22σ) on students' language scores. Thus in this setting school libraries did not lead to improved educational outcomes.

While textbooks may not increase students' time in school, one would expect that they would increase student learning. None of the three estimates (that average across all students) from the Kenya (Glewwe, Kremer and Moulin, 2009) and Sierra Leone (Sabarwal et al., 2014) is statistically significant. These findings are quite unexpected given that one would think that textbooks would have a strong impact, or conversely that lack of textbooks would have a strong negative effect, on student learning.

In Kenya, Glewwe, Kremer and Moulin (2009) investigated the reasons behind the unexpected insignificant findings and found that the official government textbooks provided were too difficult for the average child to read in the region of Kenya (Busia and Teso districts) where the study took place. Indeed, when the sample is restricted to the top 20% of students (and, in some regressions, to the top 40% of students), as measured by their pre-intervention test scores, the textbooks did improve students' learning outcomes (not shown in Table 9).

In Sierra Leone, Sabarwal et al. (2014) discovered that in this RCT, few of the textbooks reached the students. This program was implemented by the Ministry of Education, and there was little follow up action to encourage the teachers and school administrators to distribute the textbooks to students; instead, most of the textbooks were kept in storage. The authors present evidence suggesting that school administrators stored most of the textbooks because they were unsure whether textbooks would be provided in future years. Overall, this small amount of evidence suggests that textbooks can have a positive impact when they are actually provided to students, and when the textbooks are at the appropriate level for those students, which was the case for the top students in Kenya. But if textbooks are too difficult, which was the case for most students in Kenya, or are never

provided to the students, which was the case in Sierra Leone, they will have little or no effect on both time in school and student learning.

C Pedagogy Interventions

In recent years many high quality studies have assessed the impact of innovative pedagogical methods on test scores. These studies, which are summarized in Table 10, are spread across China, Colombia, India, Kenya, Peru, the Philippines, Romania and Uganda, with one third of the studies focusing on China. Of these studies, the intervention that has been found to be most effective for increasing test scores is teaching at the right level. The use of computers and/or electronic games in instruction often, but not always, effective. Tracking and streaming appears promising but needs more evidence. Too little is known about reading-intensive pedagogies and literacy curriculums to know whether they work. The following paragraphs summarize these findings.

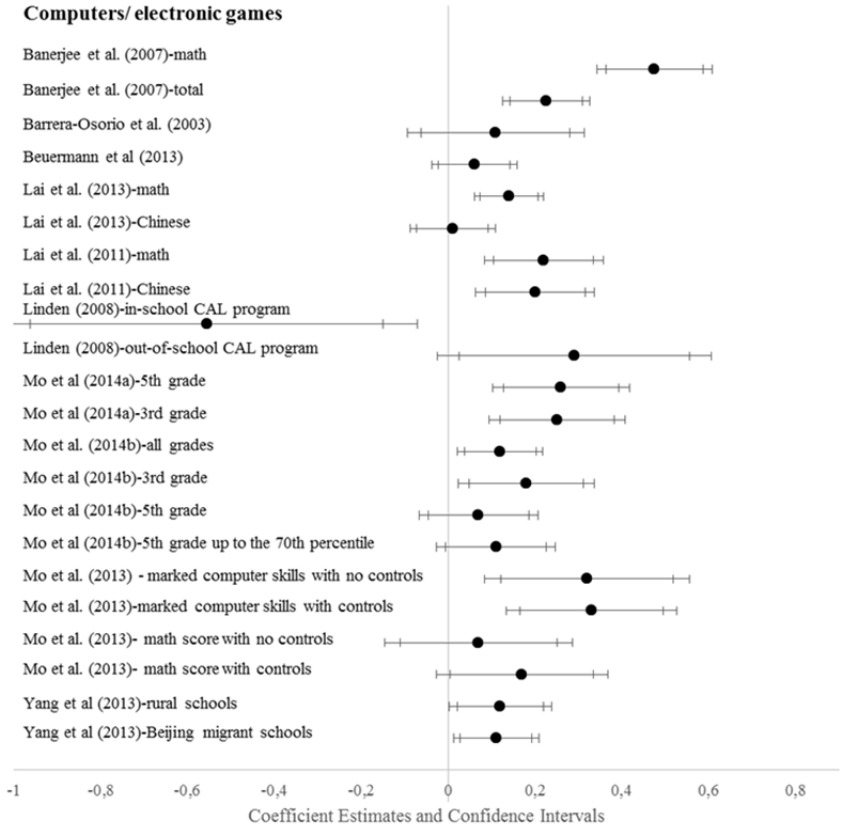
Table 10. Effect of Pedagogy Interventions on Test Scores

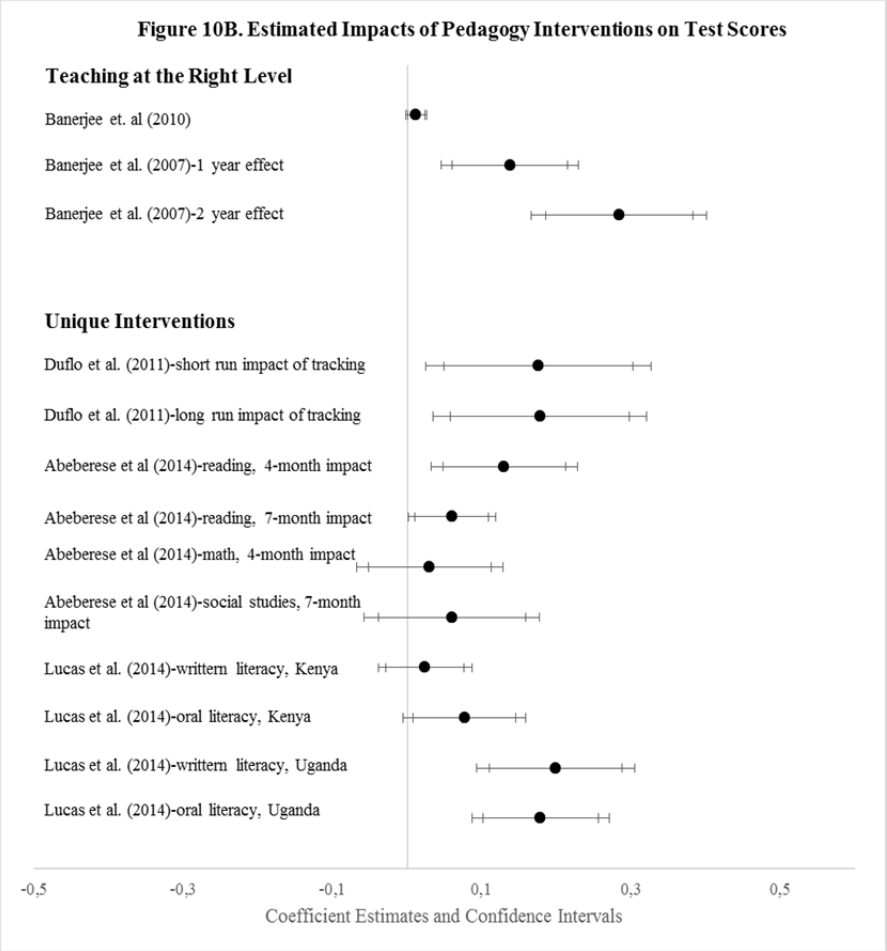
	Neg. Significant	Neg.In-significant	Positive Insignifi-cant	Positive Significant	Total Studies
Interventions that work					
Teaching at right level/ supplemental instruction (RCTs)	0 (0)	0 (0)	0 (0)	4 (3)	3
Interventions that often work					
Computers/electronic games					
RCTs	1 (1)	0 (0)	4 (4)	13 (8)	10
non-RCT	3 (1)	0 (0)	0 (0)	0 (0)	1
Promising interventions that need more evidence					
Tracking/streaming (RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Interventions that do not work					
None					
Interventions with insufficient research					
Reading-intensive pedagogy and reading materials (RCT)	0 (0)	0 (0)	2 (1)	2 (1)	1
New literacy curriculum	0 (0)	0 (0)	1 (1)	3 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Figures 10A and 10B report parameter estimates and confidence intervals for different types of pedagogy interventions reported in Table 10.

Figure 10A. Estimated Impacts of Pedagogy Interventions on Test Scores





Interventions that work.

One broad intervention that has been rigorously shown to increase learning combines remedial programs that target students who are lagging behind with "teaching at the right level", that is teaching that accounts for students' current level of skills. Ideally, such an intervention would increase student progress, and decrease the heterogeneity of student learning levels in a given grade. The evidence suggests that this may be the case, with several high-quality studies finding strong impacts of remedial instruction programs on learning outcomes, even when implemented by volunteers or informal teachers with little formal training and paid only a modest stipend that is several times lower than the salary of regular civil-service teachers.

Three studies (all conducted in India) examine the impact on students' educational outcomes of interventions that focus on "teaching at the right level", which typically involves remedial/supplemental instruction and/or tutors or volunteers to provide that instruction. First, Banerjee, Cole, Duflo, and Linden (2007) report results from an experimental evaluation of a program run by PRATHAM that was specifically targeted at the lowest performing children in public schools in the Indian cities of Mumbai and Vadodara. The program provided an informal teacher hired from the community (known as a *Balsakhi* or "friend of the child") to schools, with an explicit mandate to focus on children in 3rd and 4th grade who had not achieved even basic competencies in reading and arithmetic. The program improved students' test scores (average of math and English scores) by 0.14σ after one year of the program, and by 0.28σ after two years. Most of the gains were observed for students who were "pulled out" of their regular classroom (who were at the lower end of the learning distribution) and not for those who continued in the regular class (though the latter students did experience a reduction in class size for two hours per day). The authors therefore interpret the results as being driven by the fact that the students who were pulled out were being taught at a level corresponding to their current proficiency, as opposed to the proficiency presumed by of the textbook.

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

Third, Lakshminarayana et al. (2013), studied the impact of a program run by the Naandi Foundation. The program recruited community volunteers to provide remedial education to children in a randomly selected set of villages in Andhra Pradesh. After an initial

outreach to households to communicate program details, the volunteers provided two hours of remedial instruction per day after normal school hours in the students' school. After two years of this intervention, student test scores in program villages were 0.74σ higher than those in the comparison group, suggesting a large impact of the after-school remedial instruction program. Note, however, that the large magnitudes reported in this study also reflect high program implementation quality and monitoring over a period of two years.

While very promising, there are two challenges to interpreting this still relatively small body of evidence. First, all of these interventions occurred in India, and thus there is no evidence on whether these results generalize to other developing countries. Second, two of the three interventions studied include both extra instructional time as well as teaching that is targeted to the level of the student. However, the success of supplemental instruction at the right level offers an explanation for why increases in school inputs have not always translated into much improvement in learning outcomes; the problem may be ineffective pedagogy.

Interventions that often work.

It is widely believed that greater use of computers and other types of information and communications technology in classrooms is a promising way to rapidly improve learning outcomes in developing, and developed, countries. Interventions of this type that have been tried in developing countries range from being quite inexpensive, such as radio-based instruction, to very expensive, such as individual laptops for all students under the "One Laptop per Child" (OLPC) initiative. Eleven studies provide evidence on the impact of technology-enhanced instruction on student learning. Yet the results show widely varying magnitudes of impact. Many estimates that are significantly positive, but a few are statistically insignificant or even significantly negative, which indicates the importance of context and program design in creating effective programs of technology-aided instruction.

There are eight studies that find significant positive impacts of computer-aided learning (CAL), including two studies from India (Banerjee et al., 2007, and Linden, 2008) and six studies from China (Lai et al., 2011, Lai et al., 2013, Mo et al., 2013, 2014a, and 2014b, and Yang et al., 2013).

In India, Banerjee et al. (2007) found that a two year CAL program that provided two hours per week of computer-based math instruction in two cities in Western India was particularly effective at improving math scores, with very large positive effects of 0.48σ, but that the gains were not long-lasting (the effects fell to 0.10σ one year after the program). They also report that the CAL intervention was not as cost effective as a remedial tutoring program in the same setting. Also in India, Linden (2008) found positive effects of 0.29σ from an out-of-school CAL program after one year of the program.

In China Lai et al. (2011) estimated the impact of a one year CAL program in schools for migrant children in Beijing that provided two 40-minute sessions of remedial math instruction per week to children in grade 3. They find that the intervention increased math scores by 0.14σ, which was statistically significant. Lai et al. (2013) estimated the impact of a CAL program similar to that implemented in Beijing, but it focused on Chinese language and was implemented in a remote Western province (Qinghai); they found significant increases in both Chinese (0.20σ) and math (0.22σ) after one year.

Mo et al. (2013) studied a One Laptop per Child program on 300 migrant third graders in Beijing, where computers were loaded with game software consistent with the schools' curriculum. They find positive and significant impacts on math, language and computer skills tests. Mo et al. (2014a) examined a similar CAL program that provided remedial math instruction to boarding school students in another area of rural China and found that it led to significant improvements in math test scores for both third-grade (0.25σ) and fifth-grade (0.26σ) students. Later Mo et al. (2014b) considered a similar intervention, expanding the sample to both boarding students and students who lived at home. After one year of the program, pooled estimates of grade 3 and 5 math scores significantly increased was 0.16σ. Yang et al. (2013) also studied a CAL program in three different provinces in China and find modest (0.12σ) but significantly positive effects on test scores.

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

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

Finally, two studies have found a negative impact: Linden (2008) and Malamud and Pop-Eleches (2011). In India, Linden (2008) found that a CAL program that was implemented in-class and thus substituted for regular instruction had a strong negative impact (-0.55 standard deviations) on test scores. This is in contrast to the positive impact found from an after-school supplemental CAL program, Linden interprets these results as being driven by the difficulty of effectively modifying pedagogy within the classroom to incorporate technology, which could lead to a worsening of outcomes if effective pre-existing instructional patterns are disrupted. In Romania, Malamud and Pop-Eleches (2011) studied the impact of providing vouchers for purchasing computers to the families of middle-school students in Romania. They found that students who received the voucher had significantly lower GPA's (the results are not reported in standard deviations). The authors believe that the result was driven by that fact that students reported more time playing games and less time reading or doing homework.

These cautionary results are especially relevant for education policy, where it is tempting for politicians to want to scale up interventions like "computers for all" as a potential short-cut for addressing the challenges of education quality. Our summary of the evidence suggests that there are many good reasons to be excited about the *potential* for technology-enabled instruction to produce substantial improvements in students' learning outcomes. However, the evidence on the impact of greater use of technology in the classroom is not unanimous, and program impacts seem to depend crucially on the details of both the intervention and its implementation. In particular, it appears that the key success factor is the extent to which careful thought goes into integrating effective pedagogical techniques with technology. Much more, and much more careful, research is needed (on both process and impacts) before

committing resources to scaling up these programs - especially those involving expensive investments in hardware - with scarce public funds.

Promising interventions that need more evidence.

One pedagogy intervention that seems promising, but for which more evidence is needed, is tracking students into classrooms based on learning levels or ability. Tracking may be beneficial because it reduces the variance in student ability within a classroom, making it easier for teachers to more effectively match the content level to the students. However, students who are tracked to "lower" level classrooms may suffer further from negative peer effects and loss of self-esteem, which may place them on a permanently lower trajectory of learning. Further, some education systems may track students using data that may be noisy and not sufficiently reliable for tracking.

Nonetheless, Duflo, Dupas, and Kremer (2011) conducted an experimental evaluation of tracking in Kenya and found that tracking and streaming of pupils appears to have a positive and highly significant effect on test scores in both the short term and the long term. Students in tracking schools scored on average 0.18σ higher than students in non-tracking schools, and continued to score 0.18σ higher even one year after the tracking program ended, suggesting longer-lasting impacts than those found in many other education interventions. In addition, the authors found positive impacts for students at all quartiles of the initial test score distribution; lower-achieving students gained knowledge in basic skills while higher-achieving students gained more advanced skills, which suggests that teachers tailored their classes to the achievement level of their students. Finally, the authors are able to show that tracking did not cause adverse peer effects.

Interventions with insufficient research.

Two additional pedagogy interventions have too little evidence to determine whether, and how, they work. In the Philippines, Abeberese, Kumler and Linden (2014) provided age-appropriate reading materials and trained teachers to incorporate reading into their teaching. Four months after the start of the program the reading scores of the students in the program schools were 0.13σ higher, although this fell to 0.06σ seven months after the start of the program.

Both estimates are statistically significant. In contrast, there was no impact of the program on mathematics scores. In a study of a similar intervention, Lucas et al. (2014) evaluated a new literacy curriculum introduced into Kenya and Uganda for primary school students. The study shows the new curriculum increased written literacy test scores in Kenya by .02 σ and increased written literacy test scores in Uganda by 0.20 σ. The authors suggest that the different outcomes may result from the fact that Kenyan students had higher baseline scores and better access to learning materials, thus there was little scope for improvement. Uganda, on the other hand had much lower baseline scores and lower access to learning materials prior to the intervention.

D Interventions that Change School Governance

Some have argued that even with adequate inputs and teachers using the appropriate pedagogy, learning will not take place unless the education system is well organized and managed. This leads to the final set of education policies considered in this subsection, those related to governance. Table 11 summarizes the results for governance interventions, and the following paragraphs describe the studies in more detail.

Table 11 Effect of Governance Interventions on Test Scores

	Neg. Significant	Neg. Insignificant	Pos. Insignificant	Pos. Significant	Total Studies
Interventions that work					
None					
Interventions that often work					
School-based management					
RCTs	0 (0)	9 (3)	7 (3)	2 (2)	5
non-RCT	0 (0)	1 (1)	1 (1)	1 (1)	2
Teacher performance pay					
RCTs	0 (0)	1 (1)	2 (1)	5 (2)	3
non-RCT	0 (0)	0 (0)	1 (1)	1 (1)	1
Private school (vouchers)					
RCTs	0 (0)	0 (0)	3 (2)	2 (2)	3
non-RCT	0 (0)	2 (1)	2 (1)	0 (0)	2
Promising interventions that need more evidence					
Contract teachers (RCTs)	0 (0)	0 (0)	0 (0)	3 (2)	2
Interventions that do not work					
Monitoring (RCTs)	0 (0)	1 (1)	4 (3)	1 (1)	4
Interventions with insufficient research					
None					

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Figures 11A and 11B report coefficient estimates and confidence intervals for the studies underlying the analysis in Table 11.

Figure 11A. Estimated Impacts of Governance Interventions on Test Scores

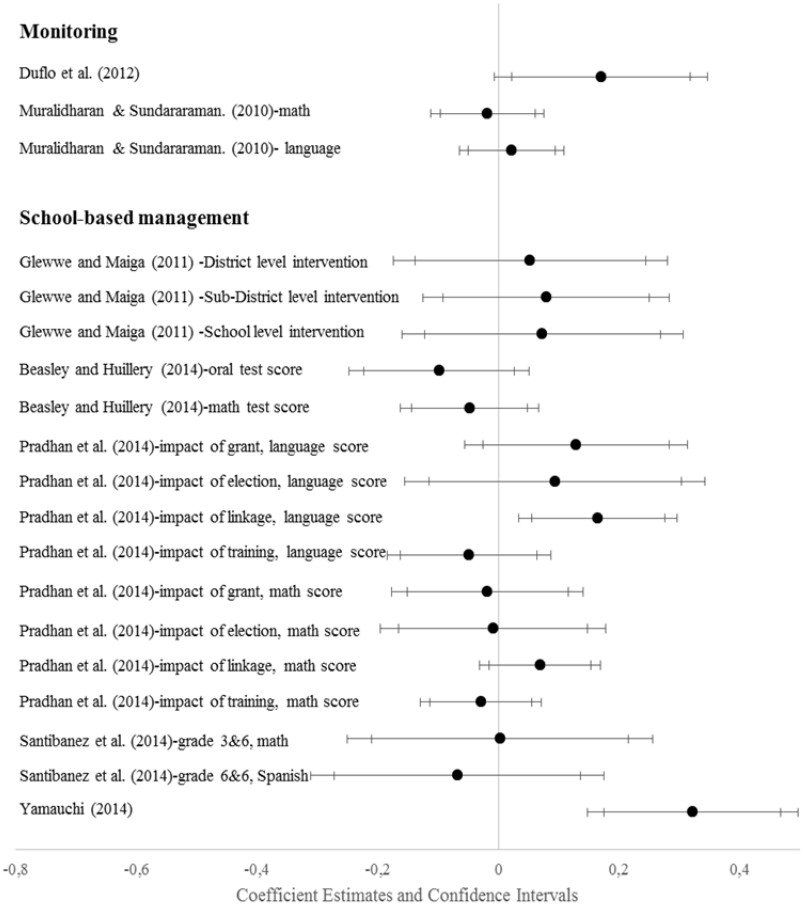
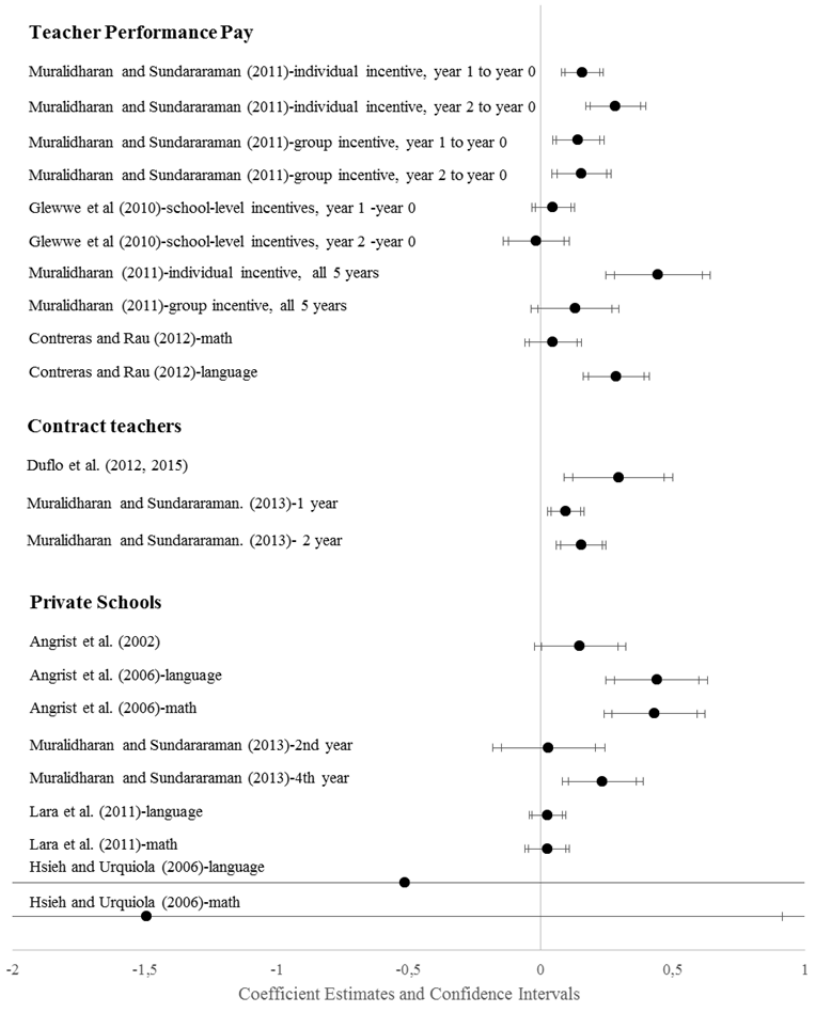


Figure 11B. Estimated Impacts of Governance Interventions on Test Scores



Interventions that often work.

We found no governance interventions that are likely to work in most circumstances, yet we did find some changes in school governance that may be effective at increasing students' learning outcomes. In particular, while school-based management programs, teacher performance pay and private school vouchers were ineffective at

increasing time in school, they may lead to increased learning under certain conditions.

The studies that examined the impact of school-based management on learning have wide coverage; they were conducted in Indonesia, Kenya, Madagascar, Mexico, Niger, and the Philippines. Of these studies, however, only three had positive and significant impacts on test scores. Pradhan et al. (2014) studied a series of interventions in Indonesia. Only one, an intervention that facilitated "linkage" meetings between school committees and village councils, had a significantly positive impact on test scores. The authors suggest that this may be because the "linkage" intervention incorporated stakeholders from a more powerful community institution (the village council). In Kenya, Duflo, Dupas, and Kremer (2012, 2015) trained school management committees to evaluate the performance of contract teachers and to have influence on the renewals of their contracts. The results show a significantly positive impact on students' test scores. In the Philippines, Yamauchi (2014) found a significantly positive impact of an SBM program (which included resources for program implementation) on students' test scores. The significantly positive effect reflects both the switch to school-based management and the additional resources to the schools.

The remainder of the school-based management studies had insignificant impacts on test scores. Beasley and Huillery (2014) evaluated a parent-empowerment program in Niger and found no impact on test scores (nor time in school). The authors suggest that in this case parents do not have enough knowledge and information to make effective decisions to improve educational quality. Lassibille et al. (2010) and Glewwe and Maiga (2011) both present experimental evaluations of the AGEMAD program in Madagascar, which aimed to strengthen school management at the district, sub-district, school and teacher levels; both studies found no impact on student test scores of these interventions. As discussed above Santibanez, Abreu-Lastra and O'Donoghue (2014) evaluated a SBM program in in Mexico, and found no impact on students' test scores (nor time in school). The authors suggest that schools are more likely to allocate grants based on immediate benefits rather than improving structural governance.

Teacher performance pay also had mixed success at increasing test scores. Two papers from one intervention in Andhra Pradesh, India, found that bonus payments to teachers based on the average improvement to student learning was successful. In particular,

Muralidharan and Sundararaman (2011) found that at the end of two years of the program, students in incentive schools performed significantly better than those in control schools by 0.27σ and 0.17σ in math and language tests, respectively. In a follow up study by Muralidharan (2011), after the program was extend for 5 years to a sub-sample, for students who completed all of their five years under the incentive program, only those in the individual teacher incentive program had significantly higher test scores than the control group. The paper estimates that the individual teacher performance pay program is 15 to 20 times more cost effective (including administrative costs) at improving learning outcomes than a common policy of reducing pupil-teacher ratios by hiring more teachers.

In a third paper on Chile, Contreras and Rau (2012) evaluated a program which provided teacher bonus payments based on students' test scores and was rolled out in a scaled up way across all public schools. Their estimates indicate that this program led to a large (0.29σ) and significant increase in students' mathematics test scores.

Similar to the weaker findings for the group teacher performance pay incentive program (relative to the individual performance pay incentive program) conducted in India, Glewwe, Ilias, and Kremer (2010), also discussed above, evaluated a teacher incentive program in Kenya that provided school-level group incentives using prizes for high-achieving schools. They found that students in treatment schools performed better on high-stakes tests but not on low-stakes tests, and also that these gains dissipated after the incentive program ended. They interpret their results as suggesting that teacher incentives may not be an effective strategy to promote long-term learning.

The final intervention that has mixed results for improving test scores is private schools/vouchers; only two of the five studies found a significant impact. As discussed above, Angrist et al. (2002) and Angrist et al. (2006) studied the short and medium term effects, respectively, of the PACES program in Colombia that provided vouchers (allocated by lottery) to students to attend private secondary schools. They found that voucher winners performed significantly better on national exams both three and seven years after receiving the voucher.

Muralidharan and Sundararaman (2015) present experimental evidence on the impact of a school-choice program in the Indian state of Andhra Pradesh. At the end of two and four years of the school

choice program, Muralidharan and Sundararaman (2015) find no difference between the test scores of lottery winners and losers on the two main subjects of Telugu (the native language of Andhra Pradesh) and math. Interestingly, they also find that private schools spend significantly less instructional time on Telugu (40% less) and math (32% less) than public schools, and instead spend more time on other subjects, including English and Hindi, the latter of which was not taught in public primary schools.

Finally, two studies from Chile found insignificant effects of attending a private school on Spanish and math test scores in that country. Lara, Mizala and Repetto (2011) found an insignificant impact of vouchers to attend private schools on the test scores of students in grade 10 (all of whom were in public primary schools but some of whom used vouchers to move into private secondary schools). Hsieh and Urquiola (2006) also do not find any significant impacts of using vouchers to enroll in a private school on the test scores of students in grades 4 and 8.

Promising interventions that need more evidence.

There are several promising intervention for increasing test scores, all of which need more evidence. The first is the use of contract teachers. Contract teachers are locally-hired teachers on fixed-term renewable contracts; these teachers are not professionally trained and are paid much lower salaries than those of regular teachers. These contracts are subject to periodic (often annual) renewal, and contracts are not renewed for under-performing teachers.

The use of contract teachers in public primary schools in developing countries has sharply increased in recent decades. Such teachers constitute about one third of public-school teachers across twelve countries in Africa (Bourdon et al. 2010), and their share among public-school teachers in India grew from 6 percent in 2003 to 30 percent in 2010 (Muralidharan et al. 2014). Unfortunately, there are only two high quality studies, one from Kenya and one from India.

In the Kenya study, Duflo, Dupas, and Kremer (2015) study a program that provided a randomly selected set of schools with an extra first grade contract teacher, reducing the class size from around 80 to about 40. The authors use this experiment to estimate the impacts of two distinct policy interventions: 1. The effects of a class size reduction when the class is taught by a civil-service teacher; and 2.

The effects of being taught, in a relatively "small" class of about 40 students, by a contract teacher versus being taught by a regular civil service teacher. The authors find that students who had the reduced class sizes *and* were also taught by a contract teacher scored significantly higher (0.29σ, averaged across subjects) than those in control schools. Even more relevant is that they found that holding class size constant, students taught by contract teachers scored significantly higher than those taught by civil-service teachers even though the contract teachers are paid much lower salaries.

In India, Muralidharan and Sundararaman (2013), study a program that provided an extra contract teacher to rural primary schools in Andhra Pradesh, Indian. After two years, students in schools with an extra contract teacher performed significantly better than those in comparison schools by 0.16σ and 0.15σ in math and language tests, respectively. This intervention combines a reduction in the pupil-teacher ratio with the introduction of a contract teacher, and so it is not possible to distinguish between the effects of these two interventions. Yet it is likely that most of the effect is from the contract teachers as opposed to the reduction in the pupil-teacher ratio. Indeed, they also found that contract teachers were significantly less likely to be absent from school than civil-service teachers (16% vs. 27%). They further suggest that the contract teachers are more effective than regular teachers, who are more qualified, better trained, and paid salaries five times higher than those of contract teachers.

Interventions that do not work.

As discussed above, a common problem in many schools is teacher absences. A common solution is monitoring. However, of the four studies (all from India) that have evaluated the impact of monitoring on learning only one found a significantly positive impact. Duflo, Hanna, and Ryan (2012) conducted a randomized evaluation of an intervention that monitored teacher attendance in informal schools in Rajasthan (India) using cameras with time-date stamps to record teacher and student attendance. The program not only monitored teachers but also paid their salaries as a function of the number of valid days of attendance. They found that this program reduced teacher absence by half, but structural estimates of a model of labor supply suggest that the mechanism for this result was not the "monitoring" per se, but rather the monetary incentives tied to the attendance. In contrast, no significant impact was found by

Muralidharan and Sundararaman (2010), who experimentally studied the impact of a program that provided schools and teachers with monitoring and feedback, but no penalty or reward based on teacher behavior or performance; they found that this program had no impact on either teacher attendance or test scores. These results suggest that monitoring is likely to be effective only when it includes positive (negative) consequences for teacher presence (absence).

Another way to improve monitoring of teachers and schools is to increase the amount of "bottom up" monitoring through the community. The evidence here is less encouraging. Banerjee et al. (2010) conducted an experimental evaluation of the impact of a community mobilization program to improve school quality in rural areas of the Indian state of Uttar Pradesh. They found no impact of various programs to build community involvement in schools in that state on community participation, teacher effort, or learning outcomes.

There is some positive evidence on the impact of community-based information campaigns (aimed at improving bottom-up monitoring), but these interventions have typically been quite intensive. Pandey, Goyal, and Sundararaman (2009) conducted an experimental evaluation of an information campaign to improve parental participation in village education committees (VEC's) in three states in India. They found positive impacts on both process measures and learning outcomes, but the estimated impacts on learning outcomes were generally statistically insignificant. Moreover, the intervention was an intensive one that involved 8-9 village level meetings in just two months, and the meetings included high-quality videos that were used to explain the rights of VEC's, as well as facilitators to answer questions.

Summary of What Works for Increasing Learning.

Interventions that work. Compared with interventions that increase time in school, we found a wider variety of effective interventions to increase learning. Thus once a community can get its children into school, there are at least four interventions that have proven to be quite effective at improving student learning. Specifically, merit-based scholarships, providing supplemental or remedial instruction ("teaching to the right level"), decreasing pupil-teacher ratios, and building new schools all have been proven effective at improving

learning outcomes for students in school. These interventions have been shown to be effective across a multitude of countries and contexts.

Interventions that often work. Also promising are several interventions that can often work. Conditional cash transfers have been shown to increase test scores in some contexts, as has the provision of school-based meals. Computer assisted learning, has increased students' test scores in many, if not most, contexts, but in other contexts is has had no effect, or even a negative effect. Changing governance structures to include school-based management, providing teacher performance pay as well as attending a private school have all been shown, under specific conditions, to increase student learning.

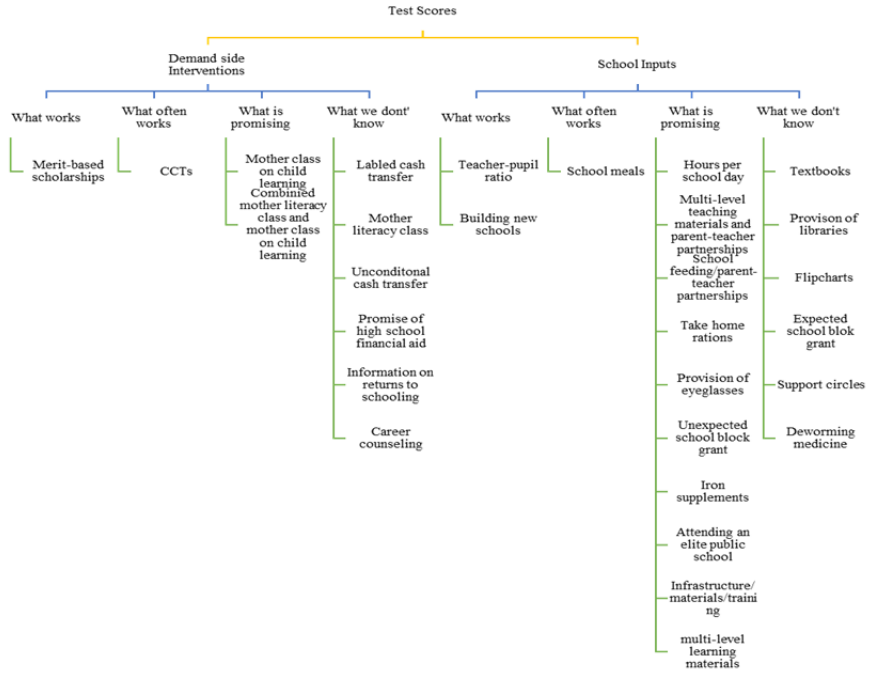
Promising interventions that needs more evidence. Many creative interventions have been evaluated to assess whether they improve student learning, and a relatively wide variety of programs have shown promise. However, interventions in this category have only one or two studies, and those with two studies have shown mixed results. For example, increasing the number of hours in a school day, combining provision of multi-level teaching materials with parent-teacher partnerships, providing iron supplementation, attending an elite public school, provision of a package of infrastructure/materials/training, and contract teachers, all have shown promise. The same is true of the provision of eyeglasses, take-home rations, and unexpected school block grants, but there is only one high quality evaluation for each of these interventions.

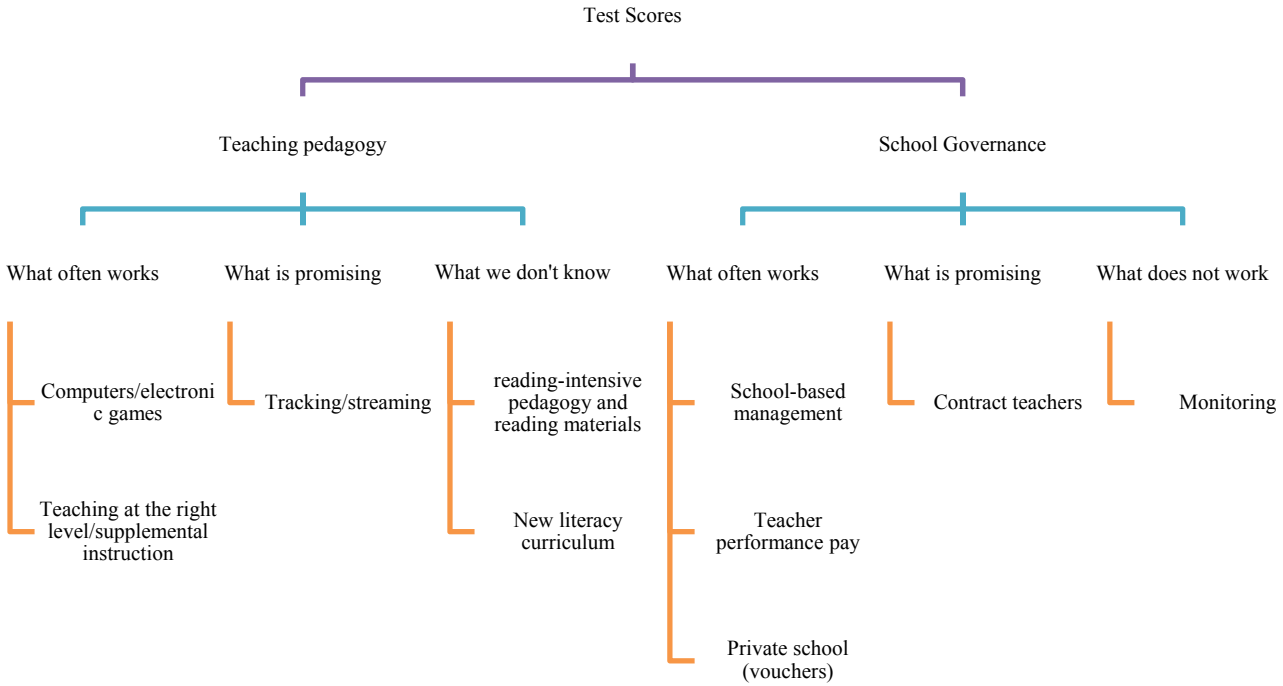
Interventions that do not work. Several interventions have been shown to be ineffective in multiple contexts, and without significant changes to program implementation they should at best be given low priority. The distribution of textbooks (without careful attention to the appropriateness and distribution of the textbooks) along with the provision of libraries both have been shown to be ineffective in improving test scores. Further, monitoring of teaching attendance, without rewards or penalties for teacher behavior and performance, is unlikely to improve teacher attendance, and thus unlikely to increase student learning.

Interventions with insufficient research. There are 14 interventions, summarized in Tables 8-11, for which too little is known about their effectiveness in improving test scores. More research is needed, especially on the more "promising" of the interventions in this group such as labeled cash transfers, which

remove the conditionality of the transfer and thereby possibly reducing the monitoring costs associated with CCTs, or reading-intensive pedagogy with accompanying reading materials.

Figure 12. Summary of Intervention Effectivness for learning outcomes




Test Scores
Teaching pedagogy
School Governance
What often works
What is promising
What we don't know
What often works
What is promising
What does not work
Computers/electronic games
Teaching at the right level/supplemental instruction
Tracking/streaming
reading-intensive pedagogy and reading materials
New literacy curriculum
School-based management
Teacher performance pay
Private school (vouchers)
Contract teachers
Monitoring

VII. The Relative Cost Effectiveness of Education Interventions on Test Scores

The discussion thus far has focused on the impact of different interventions on students' time in school and on their learning as measured by test scores. Yet for policymakers the impacts of interventions is only half of the information needed. The other half is the costs of the interventions, since the ultimate objective is to achieve a particular goal at the lowest possible cost. Thus an intervention that has the largest impact on a specific educational outcome may not be the top priority if it is also the most expensive intervention. The highest priority should be given to interventions that lead to the largest increase in the outcome per dollar (or kronor) spent.

When estimates are available on the impact of education policy interventions on students' time in school and learning, it would appear that it is a simple matter to calculate the cost of the intervention and then calculate a benefit-cost ratio. However, this is not as simple as it may seem. First, for the time in school outcomes there are many different outcome variables that are not easily compared, such as daily attendance, current enrollment, completing a given level of schooling (e.g. primary), and eventual years of schooling obtained. Indeed, for any specific time in school outcome, there are very few studies for which cost information is also available. So in this section we consider only learning outcomes as measured by test scores, which are comparable after test score impacts are transformed into standard deviations of the distribution of the test score.[16] Second, many intervention studies have multiple outcomes, including both test scores and increasing one or more time in school outcome variables. Thus, when reviewing the cost effectiveness data, one must keep in mind that focusing on test scores alone may underestimate the overall benefit of an intervention to the child, or even to the household. For example, the measures of cost effectiveness for a CCT will typically not include additional benefits the household receives from poverty reduction. Similarly, measures of cost effectiveness of school feeding

16 In fact, there are some comparability problems with this as well, as explained in Glewwe and Muralidharan (2015).

programs will typically not include the benefit of improved child health.

Cost data are available for some of the studies reviewed in the previous two sections, so it is possible to estimate benefit-cost ratios, where the benefit is increased test scores. The results are shown for several different types of interventions in Table 12. These interventions were chosen because they were among the most effective, including: merit-based scholarships, conditional cash transfer programs, computer-assisted learning, and a few miscellaneous interventions with fairly reliable cost data. The following paragraphs explain these estimates in more detail.

Three studies that found significant impacts of merit-based scholarships also report the amounts of the scholarships, although they do not report administrative and other costs of those programs. The Kenya girls' scholarship program studied both by Kremer, Miguel and Thornton (2009) and by Friedman et al. (2011) paid a total of $38 per student ($19 per year, including payments to parents, for two years). The former study found impacts of 0.27σ at the end of the two year program, which implies a cost of $14 per 0.1σ increase, and the latter study found an impact of 0.20σ 4-5 years after the program, which implies a cost of $19 per 0.1σ increase. The more recent study by Blimpo (2014) found impacts of about 0.26σ after one year for all three variants of the incentive program, but the (average) payments to students varied by the intervention: the individual incentive cost about $9 per year, which implies a cost of $3 per 0.1σ increase; the team incentive cost $3 per year, which implies a cost of $1 per 0.1σ increase; and the team tournament incentive cost about $6 per year, which implies a cost of $2 per 0.1σ increase.

Four studies of conditional cash transfer (CCT) programs found significantly positive impacts on test scores: the studies by Baird, McIntosh and Ozler (2011) and the Baird et al. (2013) of the same program in Malawi, the study by Barham et al. (2013) on a CCT program in Nicaragua. The lowest cost intervention in the Baird, McIntosh and Ozler (2011) provided payments of $5 per month for 10 months per year, and thus cost $50 per year. The payments were provided over a period of two years so the total cost was $100 per scholarship recipient. The impacts of the program on English and math scores, respectively, were 0.14σ and 0.12σ. Averaging over these impacts implies a cost of $77 per 0.1σ increase in student test scores. This is for girls who were already in school when the program began;

Baird et al. (2013) examine the impact of the program on girls who were not in school when the program began. The impacts for these girls were slightly higher, 0.13σ and 0.16σ for English and math, respectively, but that paper did not investigate the size of the impact by the size of the grant, and so the estimates are for an average grant of $10 per month; averaging over these impacts implies a cost of $138 per 0.1σ increase in student test scores. Finally, Barham et al. (2013) evaluated a CCT program in Nicaragua that made educated-related payments of $133 per household per year. A typical household had one or two children in the program, so the cost per child could be $133 or $67 per year. The impact measured is of three years in the program, so total costs are $399 (one eligible child in household) or $200 (two eligible children in household) for the duration of the program. The estimated impact, averaged over language and math, was 0.20σ. Even assuming the lower cost of $200 this implies a cost of $100 per 0.1σ increase in student test scores. Lastly Filmer and Schady (2014) evaluate a scholarship program in Cambodia that is not merit-based, and in practice is a conditional cash transfer program since it pays students conditional on their staying in school. "Scholarships" of $45 were offered to students at risk of dropping out. This program increased test scores by 0.05σ after three years of students being in the program. This implies a cost of about $90 per 0.1σ.

Four studies of the introduction of computers found significantly positive effects and also had information about the cost of the intervention. Banerjee et al. (2007) found that providing four computers per school for children to share in primary schools in India increased math test scores by 0.48σ after two years, but one year after the program ended the impact was only 0.09σ. The program cost $15 per student per year, so $30 for two years. Using the estimated effect of 0.48σ the cost was $6 per 0.1σ increase in student test scores, but using the "long-run" estimate of 0.09σ the cost increases to $33 per 0.1σ increase in student test scores. Linden (2008) also provides cost estimates for the two computer interventions that he considered. The one that was successful, which took place out of school and so did not reduce instructional time during the school day, increased students' test scores by 0.29σ after one year and cost only $5 per student per year. This implies that the cost of increasing test scores was only about $2 per 0.1σ increase in student test scores. In the case of Mo et al. (2013) they find that providing a laptop to children resulted in an increase of between .33σ for computer skills and .17σ for math skills. They report the costs to be about $5 per 0.1σ assuming a 5 year

lifespan of the laptop. Mo et al. (2014b) also found a positive overall effect of computer assisted learning on test scores and reports that this program costs about $2 per .1σ. Overall, the cost-effectiveness of interventions that provide computers to students varies widely, depending on the specifics of the intervention (recall that many had no effect, and some even had negative effects) and, for at least one study, on how long the effects persists after the child is no longer in the program.

An additional study with cost effectiveness data is that by Angrist et al. (2012), who which evaluated a program that provided vouchers that could be used to pay for private schools in Colombia. The vouchers were worth $24 to students and enabled them to partially offset the cost of attending private school. Beneficiary students in this program achieved a 0.2σ increase in test scores, making the comparable cost of the program about $12 per 0.1σ.

In Burkina Faso the program evaluated by Kazianga et al. (2013) placed well-resourced schools with a number of amenities directed at encouraging the enrollment of girls into 132 villages. Their estimates suggest that the program increased total test scores by 0.41 standard deviations. The authors estimate that enrolling an addition student in these schools cost between $63 and $70. The cost effectiveness of this program is thus about $8 per 0.1 σ.

The school building program for rural Afghan villages evaluated by Burde and Linden (2013) increased girls' achievement on combined math and language tests by 0.66 σ. The authors calculate that this program cost about $5 per 0.1 σ in test scores. Further, they calculate that the additional child-academic-year caused by the intervention for every $100 is 1.48.

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

These cost figures must be interpreted with caution, because most of them do not include administrative costs. Moreover, for the

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

Table 12. Cost-Effectiveness of Interventions in Increasing Test Scores

Intervention	Study	Country	Length of Intervention	Test Scores Measured Well after Intervention Ended?	Increase in Test Scores in Terms of US $ Cost per 0.10
Girls' merit-based scholarship	Kremer, Miguel, Thornton, 2009	Kenya	2 years	No	$14
Girls' merit-based scholarship	Friedman et al., 2011	Kenya	2 years	Yes, 4-5 years later	$19
Individual scholarship, not tournament	Blimpo (2014)	Benin	1 year	No	$3
Group scholarship, not a tournament	Blimpo (2014)	Benin	1 year	No	$1
Group scholarship, tournament	Blimpo (2014)	Benin	1 year	No	$2
Conditional cash transfer	Baird, McIntosh & Ozler, 2011	Malawi	2 years	No	$77 (for girls in school)
Conditional cash transfer	Baird et al., 2013	Malawi	2 years	No	$138 (girls not in school)
Conditional cash transfer	Barham et al., 2013	Nicaragua	3 years	Yes, 7 years later	$100
Conditional cash transfer	Filmer and Schady (2014)	Cambodia	3 years	no	$90

Conditional cash transfer	Barrera-Osorio and Filmer (2013)	Cambodia	1-3 years	no	$12
Computers (4 per school)	Banerjee et al., 2007	India	2 years	No	$6
Computers (4 per school)	Banerjee et al., 2007	India	2 years	Yes, 1 year later	$33
Computers (after school program)	Linden, 2008	India	1 year	No	$2
One laptop per child	Mo et al. (2013)	China	6 months	no	$5
Computer Assisted learning	Mo et al. (2014b)	China	1 year	no	$2
Teacher incentives	Muralidharan & Sundararaman, 2011	India	2 years	No	$1
Voucher for private schools	Angrist et al. (2002)	Colombia	1 year	Yes, 1 year later	$26
Girl-friendly schools	Kazianga et al. (2013)	Burkina Faso	N/A	no	$8
Building new schools	Burde and Linden (2013)	Afghanistan	2 years	no	$5

VIII. Policy Recommendations

Each year, Sida and other donor agencies spend billions of US dollars on programs intended to improve education outcomes in developing countries. These tax-payer resources are targeted toward many different types of interventions, but until relatively recently little rigorous evidence was generated about the actual impact of these programs.

The large number of high quality studies conducted on education programs in developing countries in the last 1-2 decades provides an unprecedented opportunity to integrate rigorous impact evaluation into policy formulation in order to make aid both more efficient and more effective. Unfortunately, much of the knowledge generated by high quality studies is written for a technical audience and thus is made available mostly through academic journals and working papers, which reduces its influence on policy. For their part, policy-makers typically face a number of barriers when formulating policies, such as political constraints, limited administrative capacity, technical feasibility issues, time pressures, and limited financial resources.

Based on the rigorous evaluations we have reviewed in this report, we have identified a number of interventions that have been shown to be generally effective at either increasing time in school or improving learning outcomes that can be used by Sida and other development agencies to maximize the impact of their development resources. However, there is still much to learn about "what works" in education in developing countries. Necessarily, these recommendations are restricted to only those interventions that have been analyzed using rigorous methods, so that while our conclusive recommendations are few in number they are made with a very high degree of confidence. Further, this analysis strongly suggests that there is ample room for organizations such as Sida to contribute to this growing body of rigorous evidence that demonstrates the effectiveness of different interventions. This section summarizes these results and provides several recommendations to aid agencies regarding planning and implementing evaluations, as well as specific recommendations for Sida.

A Priorities for Education Interventions to Increase Time in School

The most effective interventions to increase time in school in communities where attendance and/or enrollment are low are come from three of our four intervention categories. All of them reduce the actual and/or opportunity cost of attending school. Specifically, the provision of conditional cash transfers (which increase demand) and building new schools (a major school input) are both found to be effective programs to increase time in school. Two additional policies that also reduce the cost (or opportunity cost) of attendance that appear to be effective in increasing students' time in school are the provision of school meals (a school input) and providing vouchers to reduce the cost of enrolling in a private school (a school governance intervention). We found no pedagogy interventions that were effective at increasing students' time in school.

There are many more interventions that show promise to increase students' time in school in communities with low student attendance and enrollment, but more research is needed to determine whether they are effective across a variety of social, economic, and cultural contexts. Promising interventions that work by increasing the demand for schooling include: (1) Providing information directly to students and parents on the returns to education; (2) Providing school counseling to students; (3) Merit-based scholarships; (4) Labeled cash transfers; (5) Unconditional cash transfers; (6) Bicycles for student transportation; and (7) Matching remittances for educational purposes. Promising interventions that provide school inputs include: (1) Extra teachers and teaching materials; (2) Take-home rations; (3) Deworming medicine; (4) Support circles; and (5) Multi-level teaching materials. One additional promising intervention that can improve school governance is single-sex schools. We recommend that more impact evaluations be conducted on these interventions to confirm (or refute) their promising nature.

There are two interventions that are generally ineffective for the purpose of increasing students' time in school. These are changes in governance structures that either: (1) Increase monitoring of teacher performance, without rewards or penalties for teacher behavior or performance; or (2) Promote school-based management.

Finally, as seen in Tables 4-7, there are more than a dozen unique interventions that show no impact on time in school. For brevity, we do not list them here. In cases where there is only one

study and no statistically significant results there is, unfortunately, too little evidence to recommend such an intervention. Before pursuing additional funding for such interventions, more research is needed to determine their effectiveness. At minimum, interventions in this category should be tested on a smaller "pilot scale" before being scaled up to a regional or national level.

B Priorities for Education Interventions to Increase Learning

Compared with interventions that increase time in school, there is a wider variety of effective interventions to increase learning across all four intervention categories. Most promising are four interventions that have proven to be quite effective at improving student learning. The first is merit-based scholarships, which work by increasing the demand for education. Two interventions that work by providing school inputs are: (1) Building new schools where local access is lacking; and (2) Decreasing the pupil-teacher ratio. The fourth effective intervention is an improvement in teaching pedagogy, which is to provide supplemental or remedial instruction ("teaching to the right level"). Note, however, that building new schools and reducing the pupil-teacher ratio are relatively expensive interventions, and so there may be other interventions that obtain better results per dollar spent.

There are several other interventions that often work to increase student learning. These include three interventions that address policies to improve school governance (school-based management, teacher performance pay, and attending a private school) and one program in each of the other intervention categories. Conditional cash transfers (to increase demand), providing school meals (a school input) and making use of computers and electronic games (to improve teaching pedagogy) are all found to increase student learning in some contexts.

Many unique and creative interventions have been evaluated to assess whether they improve student learning, and a relatively wide variety of programs have shown promise but also require additional evaluation. Two promising interventions that may improve learning through increasing demand are: (1) Mother classes on child learning; and (2) Combining a mother literacy class with a class on child learning. Tracking/streaming of students (a change in teaching

pedagogy) and contract teachers (a school governance policy) are other interventions which show some promise. Finally, there are 10 school input interventions that show promise to increase student learning: (1) Increasing the hours of the school day; (2) Multi-level teaching materials and parent-teacher partnerships; (3) Combining a school feeding program with a parent-teacher partnership; (4) Take-home rations; (5) Provision of eyeglasses; (6) Unexpected school block grant; (7) Iron supplements; (8) Attending an elite pubic school; (9) Provision of infrastructure/materials; and (10) Multi-level learning materials.

There is only one intervention that is clearly ineffective at increasing student learning. Monitoring of teaching attendance, without rewards or penalties for teacher behavior or performance, is unlikely to improve teacher attendance, and thus unlikely to increase student learning. Finally, Tables 8-11 show over a dozen interventions for which too little is known about their effectiveness in improving test scores. Before pursuing additional funding for these interventions, more "pilot scale" research is needed to determine their effectiveness.

C Recommendations Regarding Evaluating and Implementing Interventions

While Sida and many other international development agencies often conduct evaluations, these are almost always evaluations of procedures and processes, and very rarely are they evaluations of programs' *impacts*. Moreover, almost all of the impact evaluations that aid agencies have conducted are not of particularly high quality in terms of the standards described in this report, as was seen above in the discussion of the grey literature in Section III B. Thus our first recommendation is that development agencies should conduct, or hire outside evaluation experts to conduct, high quality impact evaluations for as many of their programs and projects as possible. This subsection makes several recommendations about how to structure and implement such evaluations, and provides information on the likely costs of these types of evaluations.

A *very* important consideration when planning an impact evaluation is the timing of the evaluation relative to the project's implementation. It is *critical* that projects and programs be developed in collaboration with evaluation experts so that they are implemented in a way that allows for rigorous quantitative evaluation techniques.

Too often in the past, an evaluation of a project was an afterthought in the lifespan of a project. Ex-post impact evaluations are difficult, if not impossible, to implement in a way that generates reliable impact estimates. A useful guide for implementing one very important type of evaluation, randomized controlled trials, is provided by Glennerster and Takavarasha (2013).

A second recommendation is that international development agencies should not only estimate the impacts of their programs and projects but they should also calculate their costs. The lack of cost data and lack of a standardized system to record the costs and benefits of interventions has created a very large gap in the knowledge needed by policy makers to assess the cost-effectiveness of those projects that are shown to have positive impacts. Thus a related recommendation is that international development agencies develop a standardized system to calculate costs and benefits of intervention projects. Calculating costs and benefits can be complicated; see Dhaliwal et al. (2013) for some advice on how to collect cost data, as well as more general advice on how to conduct cost-effectiveness analysis of education programs in developing countries.

A third recommendation is that after high quality evaluations have shown that an intervention is effective, it should be implemented in new contexts in a prudent manner. More specifically, when implementing a "successful" intervention in another context, one should start small by implementing the program on a "pilot scale" in order to determine whether the program will actually work in the new context, and whether any unexpected problems arise. If the program appears successful on a pilot scale, then it can be "scaled up" to a regional level or the national level. Ideally, the pilot program should have the same geographic coverage as the planned scaled-up program. For example, if the program is intended for all rural areas in a given country then the pilot program should be done in all major "zones" of rural areas within the country.

A final set of recommendations concerns how to increase the evidence base for interventions that have two or fewer high quality studies. Perhaps the main concern is that implementing new high quality evaluations can be very expensive. While it is true that evaluations can be expensive, they should also be thought of as valuable investments in the operational efficiency of development agencies that produce knowledge on the types of programs that are effective for improving outcomes in the education sector (or any

sector of interest). The costs of randomized evaluations vary widely due to a number of factors, as discussed below. This rest of this section briefly explains the resources required to determine whether an intervention for which there is little or no high quality research is in fact an effective intervention.

Table 13 shows the resource requirements to reclassify education interventions that are either "promising interventions" or "interventions with too little research" to become either "interventions that work" or "interventions that do not work". We would argue that three or four high quality studies provide sufficient evidence on the effectiveness of a program. In particular, if there are three studies and all three show that an intervention is effective, that intervention can be classified as an "intervention that works". If there are four or more studies, and most of them (e.g. three out of four studies) show that the intervention works, then that is also an "intervention that works".

For the case of promising interventions, for which one or two studies have shown that the intervention is effective, we propose that two additional large-scale high-quality evaluations are needed to determine whether the program is effective. Given that RCTs that randomize at the school level typically cost between $500,000 and $1,000,000,[17] this would require an investment of approximately $1,000,000 to $2,000,000. While this figure may seem high, understanding the effectiveness of a program with a high degree of confidence could save hundreds of millions of dollars in the future, since international development agencies spend billions of dollars each year on education projects and programs.

Table 13 also shows the approximate cost of large-scale evaluations to determine whether "interventions with too little research" are either "interventions that work" or "interventions that do not work". Given that these interventions typically have only one high quality study, we propose that three additional high quality evaluations would be required to reliably determine the effectiveness of an "intervention with too little research". The cost of three additional RCTs would likely require an investment of between $1,5000,000 to $3,000,000 for each type of intervention. While this may seem to be a relatively expensive endeavor, it also has the

17 This range is based on the experience of the authors. It does include the cost of the intervention, but in most cases the cost of the evaluation exceeds the cost of the intervention due in part to the former's high data collection costs .

potential to increase the effectiveness of hundreds of millions of dollars of donor spending in future years.

Table 13. Cost of Further Evaluation by Effectiveness Category

Current categorization	Number of existing studies	Number of additional studies needed to determine if a program works	Per evaluation cost (range)	Cost estimate (range)
Promising Interventions	1 or 2	2 additional	$USD500,000 - $USD1,000,000	1 to 2 million USD
Interventions with too little research	1	3 additional	$USD500,000 - $USD1,000,000	1.5 to 3 million USD

It is important to note that the range of the cost of an evaluation used in Table 13 is a rough estimate. The actual cost of evaluations can vary widely, depending on a number of important factors. such as the number of schools, students, and/or households included in the evaluation, the number of rounds of surveys, the costs of transportation in-country, the detail required in the survey (more detailed household surveys are more expensive), the local enumerator costs, the number of treatment arms in an evaluation, and professional evaluator and analysis costs. Determining these costs is relatively straightforward (albeit highly dependent on local costs of labor and transportation) and thus will provide a clearer idea of how much additional evaluation will cost.

Some highly innovative interventions should probably be initially evaluated on a small scale, both qualitatively and quantitatively, to understand how they work and whether they show any promise of being effective. These "pilot" programs also offer the opportunity to rigorously evaluate innovative programs that are completely untested to see whether they should be evaluated on a larger scale. Funding for pilot interventions can be at lower levels, since the sample size and scale of the intervention are likely to be smaller. We estimate that an evaluation for a pilot type intervention could cost between $100,000 and $250,000 depending on the complexity of the program and the evaluation. USAID, for example, has a three tier funding system for their "Development Innovation

Ventures" program: Stage 1: Proof of Concept (average investment of $119,000); Stage 2: Testing at Scale (average investment of $700,000); and Stage 3: Widespread Implementation (average investment of $5.5 million). A staged evaluation funding program, such as this, is a prudent way to scale up projects of unknown effectiveness.

Note finally that the recommended number of additional large-scale evaluations needed in Table 13 is an approximation; development assistance organizations should feel free to conduct more than the recommended number of additional evaluations if they have clear reasons for doing so, but we would caution against making decisions based on fewer than the recommended number of additional evaluations in Table 13.

D Other Recommendations to International Development Agencies

We also have other recommendations for international development agencies. We start with two general recommendations that apply to almost all agencies, and then turn to a few specific recommendations for Sida.

First, international development agencies should consider the cost and benefits of "outsourcing" their impact evaluation work, as opposed to conducting most or all of that work "in house". There are several advantages of outsourcing evaluations. The first is that it may be less expensive; in principle evaluation experts compete among each other for contracts to do this work, and thus they have an incentive to offer a "good price" for their services. The second is that evaluation experts or organizations that are not a part of the development agency may be more objective in their evaluations, since neither they (nor their colleagues) have a vested interest in the success (or failure) of the program they are evaluating. Third, outsourced evaluations by an evaluation expert will have the advantage of results being presented in conferences and even published in peer reviewed journals which can generate positive externalities of the shared knowledge of what works, which will benefit both researchers and NGO agencies. While the academic content may be quite technical, evaluation experts can be easily contracted to produce policy pieces that are written in way that is understandable for policy makers, donors and government officials.

On the other hand, there are good reasons not to rely on outsourced work for all evaluations. First, international development agencies that increase the technical skills of in-house staff in order for them to conduct evaluations will make those staff members better judges of the quality of out-sourced studies. Second, developing an in-house capacity may attract highly skilled evaluators to work directly for international development agencies. Finally, in the long run, as the skills of in-house staff improve and the staff gain experience in effective impact evaluation studies, it may be more efficient to reallocate a larger proportion of the evaluations to an in-house team. The World Bank is an example of a large international organization with a highly skilled group of staff who conduct very rigorous impact evaluations.

Overall, the considerations in the previous two paragraphs suggest that international development agencies should consider outsourcing a substantial fraction of impact evaluations to outside experts, while at the same time developing its own highly skilled in-house technical staff

A second recommendation for international development agencies concerns how to use the results of an impact evaluation of a specific program's effectiveness, and in particular how to treat agency staff whose projects have been found to be ineffective. It is very likely that many programs will be found to ineffective in terms of their impact on education outcomes, because so much is unknown about what works. As shown above, there are a number of unique interventions that show promise and warrant investment in additional research. Thus it would be unreasonable to expect program implementers to have high rates of program effectiveness. Until more is learned about what works, experimentation is needed in education programs. This is what can produce new types of effective programs, but experimentation is inherently risky so one should not penalize program staff who are willing to try new programs, many of which in the end may not work.

We end this section, and thus this report, with some recommendations that are specific to Sida. They are based on the fact that Sida the last few years has provided education assistance in mainly eight developing countries: Afghanistan, Bangladesh, Bolivia, Cambodia, Kosovo, Liberia, Rwanda and Tanzania. For two of these countries, Kosovo and Tanzania, we have found no high quality studies, although there may well be high quality studies for Tanzania

in the next few years. Yet we have found one high quality study for Afghanistan, one for Bangladesh, two for Bolivia, and three for Cambodia. Thus our first recommendation is to use the currently available high quality studies of these four countries as an initial guide for what types of education policies to fund. Second, neighboring countries can provide useful additional guidance on the types of education policies to implement. For example, interventions that are effective in India are likely to be effective in Bangladesh, and interventions that are effective in Kenya are likely to be effective in Tanzania.

A third and final recommendation is that Sida should conduct new impact evaluations in these six countries, perhaps in cooperation with other development agencies that work in those countries (in order to reduce the costs to Sida). Table 13 gives a rough idea of how many evaluations would be needed and what the costs would be, to produce *generalizable* results, but for a *specific country* only one well implemented evaluation is needed to determine whether a particular intervention is effective in that country. However, some caution is needed in interpreting the results of a single evaluation of a particular intervention. An evaluation that is effective on a small scale may not be effective when implemented at the national level, for two reasons. First, the results a medium-scale evaluation (i.e. on the will cost between $500,000 and $1,000,000) are usually restricted to one or two regions of a country, and those regions may differ from the rest of the country. One implication of this is that such evaluations should be done in one or two "typical" regions of a given country. Second, such evaluations are usually carefully implemented and monitored by committed practitioners, and this level of care and commitment is not necessarily available at the national level, which implies that many programs that are effective at a sub-national scale are less effective on the national scale.

It is also possible that a program that is ineffective in a sub-national evaluation could be more effective if some changes are made. It is always possible that implementation is weak the first time a program is tried, or that some other type of "bad luck" resulted in an ineffective program. Collection of data on how the program was implemented, as well as various types of qualitative data, should provide useful information on whether the program should be evaluated again in the same country, or perhaps permanently abandoned.

These recommendations will require substantial effort, and as such they should be debated before being blindly followed. Yet given the current weak evidence base for "what works" to improve education outcomes in developing countries, more high-quality evaluations are essential for using donor funds more effectively to increase learning, and thus economic and social progress, in developing countries. While much has been learned in the past two decades, much more needs to be learned to ensure a bright future for today's, and tomorrow's, children.

Bibliography

Bibliography of Impact Evaluation Studies
*114 total studies: 75 *RCT and 39 High Quality (RDD-DD design)*

*Abeberese, Ama Baafra, Todd J. Kumler and Leigh L. Linden. "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*, Volume 49, Number 3, Summer 2014, pp.611-633

Adrogue, Cecilia, and Maria Eugenia Orlicki. 2013. "Do In-School Feeding Programs Have an Impact on Academic Performance? The Case of Public Schools in Argentina." Education Policy Analysis Archives 21, no. 50:

Afridi, Farzana. "The impact of school meals on school participation: evidence from rural India." Journal of Development Studies 47, no. 11 (2011): 1636-1656.

*Alderman, Harold, Daniel O. Gilligan, and Kim Lehrer. "The impact of food for education programs on school participation in northern Uganda." Economic Development and Cultural Change 61, no. 1 (2012): 187-218.

*Alderman, Harold, Jooseop Kim and Peter F. Orazem (2003) "Design, evaluation, and sustainability of private schools for the poor: the Pakistan urban and rural fellowship school experiment" *Economics of Education Review* 22: 265-274.

*Ambler, Kate Diego Aycinena and Dean Yang (2014) "Channeling remittances to education: a field experiment among migrants from El Salvador" *NBER Working Paper #* 20262 [published as Ambler, Aycinena and Yang (2015)

*Ambler, Kate, Diego Aycinena and Dean Yang (2015) "Channeling Remittances to Education: A Field Experiment among Migrants from El Salvador" *American Economic Journal: Applied Economics*. 7(2): 1-27.

*Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). 'Vouchers for Private Schooling in Colombia: Evidence

from a Randomized Natural Experiment', American Economic Review, 92(5): 1535-58.

*Angrist, J., Bettinger, E., and Kremer, M. (2006). 'Long-term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia', American Economic Review, 96(3): 847-62.

Asadullah, M. Niaz. 2005. The effect of class size on student achievement: Evidence from bangladesh. Applied Economics Letters 12 (4) (March 2005): 217-21, http://www.tandf.co.uk.floyd.lib.umn.edu/journals/titles/13504851.asp.

Attanasio, Orazio, Emla Fitzsimons, Ana Gomez, Martha Isabel Gutierrez, Costas Meghir, and Alice Mesnard. (2010) "Children's schooling and work in the presence of a conditional cash transfer program in rural Colombia." *Economic Development and Cultural Change* 58(2): 181-210.

Baez, Javier E. and Adriana Camacho. (2011). "Assessing the long-term effects of conditional cash transfers on human capital: evidence from Colombia." World Bank Policy Research Paper No. 5681. Washington, DC: World Bank.

*Baird, Sarah J., Ephraim Chirwa, Jacobus de Hoop, and Berk Özler, (2013). "Girl Power: Cash Transfers and Adolescent Welfare. Evidence from a Cluster-Randomized Experiment in Malawi," *NBER Working Papers* 19479, National Bureau of Economic Research.

*Baird, Sarah, Craig McIntosh and Berk Ozler (2011). "Cash or Condition? Evidence from a Cash Transfer Experiment," *The Quarterly Journal of Economics*. 26(4): 1709-1753.

*Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Gelnnerster and Stuti 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, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. (2007) "Remedying Education: Evidence from two randomized experiments in India." Quarterly Journal of Economics August: 1235-1264.

*Banerji, Rukmini, James Berry, and Marc Shotland. (2013) "The impact of mother literacy and participation programs on child

learning: evidence from a randomized evaluation in India." Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).

*Barham, Tania, Karen Macours, and John A. Maluccio. (2013) "More Schooling and More Learning?: Effects of a Three-Year Conditional Cash Transfer Program in Nicaragua after 10 Years." *Inter-American Development Bank*.

*Barrera-Osorio, Marianne Bertrand, Leigh Linden, Francisco Perez-Calle. (2011) "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics* 3): 167-195.

*Barrera-Osorio, Felipe, and Deon Filmer. (2013) "Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches." *World Bank Policy Research Working Paper* 6541.

*Barrera-Osorio and Leigh Linden. (2009) "The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia." World Bank Policy Research Working Paper #4836. Impact Evaluation Series No. 29.

Barrera-Osorio, F., & Raju, D. (2011). Evaluating public per-student subsidies to low cost private schools: regression-discontinuity evidence from Pakistan. World Bank Policy Research Working Paper Series. Available at: http://documents.worldbank.org/curated/en/2011/04/14103503/evaluating-public-per-student-subsidies-low-cost-private-schools-regression-discontinuity-evidence-pakistan

*Beasley, Elizabeth, and Elise Huillery. "Willing but Unable: Short-Term Experimental Evidence on Parent Empowerment and School Quality." J-PAL working paper (2014). http://www.povertyactionlab.org/publication/willing-unable-short-term-experimental-evidence-parent-empowerment-and-school-quality

Behrman, Jere R., Jorge Gallardo-Garcia, Susan W. Parker, Petra E. Todd and Viviana Velez-Grajales (2012) "Are conditional cash transfers effective in urban areas? Evidence from Mexico." *Education Economics*. 20(3): 233-259.

*Behrman, Jere R., Susan W. Parker, and Petra E. Todd. (2009) "Schooling impacts of conditional cash transfers on young children:

Evidence from Mexico." *Economic Development and Cultural Change*. 57(3): 439477.

*Behrman, Jere R, Susan W. Parker, and Petra E. Todd (2011) "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades." *Journal of Human Resources* 46(1): 93-122.

Bellei, Cristian. "Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile" Economics of Education Review 28 (2009) 629-640.

*Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen (2014) "Turning a shove into a nudge? A "Labeled Cash Transfer" JPAL working paper. [Published as: Benhassinie, Devoto, Duflo, Dupas and Pouliquen (2015)]

*Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. (2015). "Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education." *American Economic Journal: Economic* Policy, 7(3): 86-125.

*Beuermann, Diether, Julian Cristia, Yyannu Cruz-Aguayo, Santiago Cueto, and Ofer Malamud. (2013) "Home Computers and Child Outcomes: Short-term impacts from a randomized experiment in Peru. NBER Working Paper Series #18818.

*Blimpo, Moussa P. (2014) "Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin." *American Economic Journal: Applied Economics* 6(4): 90-109.

Borkum, Evan. (2012) "Can eliminating school fees in poor districts boost enrollment? Evidence from South Africa." *Economic Development and Cultural Change* 60(2): 359-398.

*Borkum, Evan, Fang He, and Leigh L. Linden. The Effects of School Libraries on Language Skills: Evidence from a Randomized Controlled Trial in India. No. w18183. National Bureau of Economic Research, 2012.

*Burde, Dana and Leigh L. Linden (2013) "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools" *American Economic Journal: Applied Economics*. 5(3):27-40.

*Chaudhury, Nazmu, Jed Friedman and Junko Onishi (2013) "Philippines Conditional Cash Transfer Program : Impact

Evaluation 2012" Report number 75533-PH. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/13244 License: CC BY 3.0 IGO.

Chaudhury, Nazmul, and Dilip Parajuli. (2010) "Conditional cash transfers and female schooling: the impact of the female school stipend programme on public school enrolments in Punjab, Pakistan." *Applied Economics* 42(28): 3565-3583.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." The American Economic Review 95, no. 4 (2005): 1237-1258.

Chen, Yuyu and Ginger Zhe Jin. 2012. "Does health insurance coverage lead to better health and educational outcomes? Evidence from rural China" Journal of Health Economics, Vol.31(1), pp.1-14

Chin, Aimee. 2005. Can redistributing teachers across schools raise educational attainment? Evidence from operation blackboard in India. Journal of Development Economics 78 (2) (December 2005): 384-405.

Contreras, Dante and Tomas Rau. "Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile." Economic Development and Cultural Change Vol 61 No 1 (October 2012), pp 219-246.

Cristia, Julian, Alejo Czerwonko, and Pablo Garofalo. "Does technology in schools affect repetition, dropout and enrollment? Evidence from Peru." Journal of Applied Economics 17 (2014): 89-112.

*Dammert, Ana (2009) "Heterogeneous Impacts of Conditional Cash Transfers: Evidence from Nicaragua" *Economic Development and Cultural Change* 58(1): 53-83

*Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics*, 5(2): 29-57.

De Janvry, Alain, Frederico Finan, and Elisabeth Sadoulet. (2012) "Local electoral incentives and decentralized program performance." *Review of Economics and Statistics* 94(3): 672-685.

*Dubois, Pierre, Alain de Janvry and Elisabeth Sadoulet (2012) "Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico" *Journal of Labor Economics* 30(3): 555-589.

Duflo, Esther (2001) "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment" *American Economic Review*. 91(4): 795-813

*****Duflo, Esther, Pascaline Dupas, and Michael Kremer. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." The American Economic Review 101 (2011): 1739-1774.

*Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools." No. w17939. National Bureau of Economic Research, 2012. Available at http://www.nber.org/papers/w17939 [published as Duflo, Dupas and Kremer (2015)]

*Duflo, Esther, Pascaline Dupas, and Michael Kremer. "School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools." Journal of Public Economics, Volume 123, March 2015, Pages 92–110

*Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." American Economic Review, 102(4): 1241–78.

*Dumitrescu, Anca and Dan Levy, Cara Orfield, and Matt Sloan. 2011. "Impact Evaluation of Niger's IMAGINE Program." Washington, DC: Mathematica Policy Research. Available at: http://cipre.mathematica-mpr.com/our-publications-and-findings/publications/impact-evaluation-of-nigers-imagine-program

Filmer, Deon, and Norbert Schady. (2008) "Getting girls into school: evidence from a scholarship program in Cambodia." *Economic development and cultural change*. 56(3): 581-617.

Filmer, Deon and Norbert Schady. (2014) "The medium-term effects of scholarships in a low-income country." *Journal of Human Resources*, 49(3): 663-694.

*Friedman, Willa, Michael Kremer, Edward Miguel, and Rebecca Thornton. (2011). *Education as Liberation?* No. w16939. National Bureau of Economic Research.

*Galiani, Sebastian, and Patrick J. McEwan. (2013) "The heterogeneous impact of conditional cash transfers." *Journal of Public Economics* 103: 85-96.

Garcia, Sandra and Jennifer Hill (2010) "Impact of conditional cash transfers on children's school achievement: evidence from Colombia" *Journal of Development Effectiveness* 2(1): 117-137.

Gertler, Paul J., Harry Anthony Patrinos and Marta Rubio-Codina (2012) "Empowering parents to improve education: Evidence from rural Mexico" Journal of Development Economics 99:68-79

*Gitter, Seth and Bradford Barham (2008) "Women and Targeted Cash Transfers in Nicaragua" *World Bank Economic Review*, 22(2): 271-290

*Glewwe, Paul, Nauman Ilias, and Michael Kremer. "Teacher Incentives." American Economic Journal: Applied Economics (2010): 205-227.

Glewwe, P., & Kassouf, A. L. (2012). The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, 97(2), 505-517.

*Glewwe, Paul, Michael Kremer, Sylvie Moulin, & Eric Zitzewitz. (2004). Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya. Journal of Development Economics, 74(1): 251-68

*Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many children left behind? Textbooks and test scores in Kenya." *American Economic Journal: Applied Economics* 1 (1) (January 2009): 112-35, http://www.aeaweb.org.floyd.lib.umn.edu/aej-applied/.

*Glewwe, Paul, and Eugenie WH Maïga. "The impacts of school management reforms in Madagascar: do the impacts vary by teacher type?" Journal of Development Effectiveness 3, no. 4 (2011): 435-469.

*Glewwe, Paul, Albert Park, and Meng Zhao. "A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China." J-PAL Working Paper (2011).

Handa, Sudhanshu (2002) "Raising primary school enrolment in developing countries. The relative importance of supply and demand" *Journal of Development Economics* 69:103-128

*Hidalgo, Diana, Mercedes Onofa, Hessel Oosterbeek, and Juan Ponce. "Can provision of free school uniforms harm attendance? Evidence from Ecuador." Journal of Development Economics 103 (2013): 43-51.

Hsieh, Chang-Tai, and Miguel Urquiola. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." Journal of Public Economics 90, no. 8 (2006): 1477-1503.

*Huan, Wang, James Chu, Prashant Loyalka, Xin Tao, Yaojiang Shi, Qinghe Qu, Chu Yang and Scott Rozelle (2014) "Can School Counseling Reduce School Dropout in Developing Countries?" REAP Working Paper #275

Jackson, C. Kirabo (2012) "Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago" Journal of Public Economics, Volume 96, Issues 1–2, February 2012, Pages 173–187

Jayachandran, Seema (2014) "Incentives to teach badly: After-school tutoring in developing countries" Journal of Development Economics. 108:190-205

*Jensen, Robert (2010) "The (Perceived) Returns to Education and the Demand for Schooling" *The Quarterly Journal of Economics*. 25(2): 515-548.

Kazianga, Harounan, Dan Levy, Leigh L. Linden and Matt Sloan (2013) "The effects of "girl-friendly" schools: evidence from the BRIGHT school construction program in Burkina Faso" *American Economic Journal: Applied Economics* 5(3): 41-62.

*Kazianga, Harounan, Damien de Walque, and Harold Alderman. (2012) "Educational and Child Labour Impacts of Two Food-for-Education Schemes: Evidence from a Randomized Trial in Rural Burkina Faso" *J Afr Econ* 21 (5): 723-760 first published online April 12, 2012 doi:10.1093/jae/ejs010

*Kremer, Michael, Edward Miguel, and Rebecca Thornton. (2009). "Incentives to learn." *The Review of Economics and Statistics 91, no. 3: 437-456.*

*Lai, Fang, Renfu Luo, Linxiu Zhang, Xinzhe Huang, Scott Rozelle. 2011. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing" Rural Education Action Program working paper 228.

*Lai, Fang, Linxiu Zhang, Qinghe Qu, Xiao Hu, Yaojiang Shi, Matthew Boswell, Scott Rozelle. 2013. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China" Rural Education Action Program working paper.

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

Lara, Bernardo, Alejandra Mizala, and Andrea Repetto. 2011. "The Effectiveness of Private Voucher Education: Evidence from Structural School Switches." Educational Evaluation and Policy Analysis 33, no. 2: 119-137.

*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.

Levy, Dan and Jim Ohls (2010) "Evaluation of Jamaica's PATH conditional cash transfer program" *Journal of Development Effectiveness* 2(4): 421-441

*Li, Tao, Li Han, Linxiu Zhang and Scott Rozelle (2014) "Encouraging classroom peer interactions: Evidence from Chinese migrant schools." *Journal of Public Economics* 111: 29-45.

*Linden, Leigh (2008). "Complement or Substitute? The Effect of Technology on Student Achievement in India" J-PAL Working Paper

*Loyalka, Prashant, Chengfang Liu, Yingquan Song, Hongmei Yi, Xiaoting Huang, Jianguo Wei, Linxiu Zhang, Yaojiang shi, James

Chu and Scott Rozelle (2013) "Can information and counseling help students from poor rural areas go to high school? Evidence from China." *Journal of Comparative Economics* 41: 1012-1025.

Lucas, Adrienne and Isaac Mbiti (2014) "Effects of School Quality and Student Achievement: Discontinuity Evidence from Kenya" American Economic Journal: Applied Economics 6(3): 234-263

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

*Luo, Renfu, Yaojiang Shi, Linxiu Zhang, Chengfang Liu, Scott Rozelle, Brian Sharbono, Ai Yue, Qiran Zhao, and Reynaldo Martorell. "Nutrition and educational performance in rural China's elementary schools: Results of a randomized control trial in Shaanxi Province." Economic Development and Cultural Change 60, no. 4 (2012): 735-772.

Malamud, Ofer and Cristian Pop-Eleches. "Home Computer Use and the Development of Human Capital" *The Quarterly Journal of Economics* (2011) 126, 987-1027.

McEwan, Patrick. "The impact of Chile's school feeding program on education outcomes" Economics of Education Review 32 (2013) 122-139.

*Miguel, Edward and Michael Kremer (2004) "Worms: Identifying impacts on education and health in the presence of treatment externalities" Econometrica 72(1):159-217

*Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M., & Rozelle, S. (2013). Can One-to-One Computing Narrow the Digital Divide and the Educational Gap in China? The Case of Beijing Migrant Schools. World Development Vol. 46, pp. 14–29, 2013

*Mo, D., Zhang, L., Lui, R., Qu, Q., Huang, W., Wang, J.,. Rozelle, S. (2014b). Integrating computer-assisted learning into a regular curriculum: evidence from a randomized experiment in rural schools in Shaanxi. Journal of Development Effectiveness. Volume 6, Issue 3, 2014

*Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S., & Brinton, C. (2013). "School Dropouts and Conditional Cash Transfers: Evidence from

a Randomised Controlled Trial in Rural China's Junior High Schools." *The Journal of Development Studies*, 49(2): 190-207.

*Mo, Di, Linxiu Zhang, Jiafu Wang, Weiming Huang, Yaojiang Shi, Matthew Boswell, Scott Rozelle. (2014a). "The Persistence of Gains in Learning from Computer Assisted Learning (CAL): Evidence from a Randomized Experiment in Rural Schools in Shaanxi Province in China. REAP Working Paper 268.

*Muralidharan, Karthik. "Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India." JPAL Working Paper (2011).

Muralidharan, Karthick and Nishith Prakash (2013) "Cycling to school: increasing secondary school enrollment for girls in India." *NBER Working Paper* #19305

*Muralidharan,Karthik and Venkatesh Sundararaman. "Teacher Performance Pay: Experimental Evidence from India" Journal of Political Economy, Vol. 119, No. 1 (February 2011), pp. 39-77

*Muralidharan, Karthik, and Venkatesh Sundararaman. 2013. Contract Teachers: Experimental evidence from India. JPAL Working Paper. Available at: http://www.povertyactionlab.org/evaluation/extra-contract-teachers-andhra-pradesh-india

*Muralidharan, Karthik and Venkatesh Sundararaman. 2013. The Aggregate Effect of School Choice: Evidence from a Two-stage Experiment in India. NO. 19441. National Bureau of Economic Research working paper. http://www.nber.org/papers/w19441 [published as Muralidharan and Sundararaman (2015)]

*Muralidharan, Karthik and Venkatesh Sundararaman. The Aggregate Effect of School Choice: Evidence from a Two-stage Experiment in India. The Quarterly Journal of Economics (2015) 130 (3): 1011-1066.

*Muralidharan, Karthik, and Venkatesh Sundararaman. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." The Economic Journal 120, no. 546 (2010): F187-F203.

*Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa and Jose Luis Evia. (2002) "An Impact Evaluation of Education, Health and Water Supply Investments by the Bolivian Social Investment Fund" *The World Bank Economic Review*. 16(2):241-274.

Orkin, Kate. The Effect of Lengthening the School Day on Children's Achievement in Ethiopia. 2013. YL-WP119

*Oster, Emily, and Rebecca Thornton. 2011. "Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics*, 3(1): 91-100.

*Pandey, Priyanka, Sangeeta Goyal and Venkatesh Sundararaman (2009) "Community participation in public schools: impact of information campaignes in three Indian states." *Education Economics*, 17(3): 355-375.

*Pridmore, Pat & Catherine Jere (2011) Disrupting patterns of educational inequality and disadvantage in Malawi, Compare: A Journal of Comparative and International Education, 41:4, 513-531.

Pop-Eleches and Miguel Urquiola. "Goring to a Better School: Effects of Behavioral Responses" American Economic Review 2013, 103(4): 1289-1324.

*Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, Armida Alisjahbana, and Rima Prama Artha. "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia." American Economic Journal: Applied Economics 6, no. 2 (2014): 105-26.

*Sabarwal, Shwetlena; Evans, David K.; Marshak, Anastasia. 2014. *The permanent input hypothesis: the case of textbooks and (no) student learning in Sierra Leone*. Policy Research working paper, no. WPS 7021. Washington, DC: World Bank Group

Santibanez, Lucrecia, Raul Abreu-Lastra and Jennifer L. O'Donoghue (2014) "School-based management effects: Resources or governance change? Evidence from Mexico" *Economics of Education* Review 39:97-109

*Schultz, Paul T. (2004) "School subsidies for the poor: evaluating the Mexican Progresa poverty program." *Journal of Development Economics* 74:199-250.

*Sylvia, Sean, Renfu Luo, Linxiu Zhang, Yaojiang Shi, Alexis Medina, and Scott Rozelle. "Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China." Economics of Education Review 37 (2013): 1-12.

*Tan, Jee-Peng, Julia Lane, and Gerard Lassibille. 1999. Student outcomes in Philippine elementary schools: An evaluation of four experiments. World Bank Economic Review 13 (3) (September 1999): 493-508, http://wber.oxfordjournals.org.floyd.lib.umn.edu/.

Urquiola, Miguel. 2006. Identifying class size effects in developing countries: Evidence from rural Bolivia. Review of Economics and Statistics 88 (1) (February 2006): 171-7,

Urquiola, M., & Verhoogen, E. (2009). Class-Size Caps, Sorting, and the regression discontinuity Design. American Economic Review, 99(1), 179-215.

Yamauchi, Futoshi. 2014. An alternative estimate of school-based management impacts on students' achievements: evidence from the Philippines. Policy Research working paper no. WPS 6747; Impact Evaluation series no. IE 113. Washington, DC: World Bank Group.

*Yang, Yihua, Linxiu Zhang, Junxia Zeng, Xiaopeng Pang, Fang Lai, Scott Rozelle. (2013) "Computers and the academic performance of elementary school-aged girls in China's poor communities". Computers & Education 60 (2013) 335-346.

*Yi, Hongmei, Yingquan Song, Chengfang Liu, Xiaoting Huang, Linxiu Zhang, Yunli Bai, Baoping Ren, Yaojiang Shi, Prashant Loyalka, James Chu, Scott Rozelle. (2014) "Giving Kids a Head Start: The Impact and Mechanisms of Early Commitment of Financial Aid on Poor Students in Rural China." *Journal of Development Economic*. 113: 1–15.

*Zhang, L., Lai, F., Pang, X., Yi, H., & Rozelle, R. The impact of teacher training on teacher and student outcomes: evidence from a randomized experiment in Beijing migrant schools. Journal of Development Effectiveness. Volume 5, Issue 3, 2013

Additional References (not listed above)

Baird, Sarah, Francisco H. G. Ferreira, Berk Özler, Michael Woolcock (2013). "Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review." Campbell Systematic Reviews 2013:8.

Birdthistle, I. K. Dickson, M. Freeman and L. Javidi (2011) "What impact does the provision of separate toilets for girls at schools have on their primary and secondary school enrolment, attendance and completion?: A systematic review of the evidence. London: EOOI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bouillon, Cesar Patricio and Luis Tejerina (2007) "Do we know what works? : A systematic review of impact evaluations of social programs in Latin America and the Caribbean" Inter-American development Bank.

Carr, Stuart, Chez Leggatt-Cook, Michael Clarke, Malcolm MacLachlan, T.S. Papola, Jesim Pais, Steve Thomas, Eilish McAuliffe, and Charles Normand (2011) "What is the evidence of the impact of increasing salaries on improving the performance of public servants, including teachers, doctors/nurses, and mid-level occupations, in low and middle-income countries: Is it time to give pay a chance?" London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Conn, Katharine (2014). Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations. 3IE Systematic Review: 2014. Available at: http://www.3ieimpact.org/en/evidence/systematic-reviews/details/326/

Dhaliwal, Iqba, Esther Duflo, Rachel Glennerster and Caitlin Tulloch (2013) "Comparative Cost-Effectiveness to Inform Policy in Developing Countries", in P. Glewwe, ed., *Education Policy in Developing Countries*. University of Chicago Press.

Evans, David K. and Anna Popova (2015) "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." World Bank Policy Research Working Paper 7203.

Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton University Press: Princeton, NJ.

Glewwe, Paul, Eric Hanushek, Sarah Humpage and Renato Ravina. 2013. "School Resources and Educational Outcomes in Developing Countries", in P. Glewwe, ed., *Education Policy in Developing Countries*. University of Chicago Press.

Glewwe, Paul and Michael Kremer (2006) "Schools, Teachers, and Education Outcomes in Developing Countries" Ch. 16 in E. Hanusheck and F. Welch, eds, Handbook of the Economics of Education, Volume 2, Elsevier.

Glewwe, Paul and Karthik Muralidharan "Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications" in E. Hanusheck, S. Machin and L. Woessmann, eds., Handbook of the Economics of Education, Volume 5. North Holland/Elsevier (forthcoming).

Guerrero, Gabriela, Juan Leon, Mayli Zapata, Claudia Sugimaru, Santiago Cueto (2012) "What works to improve teacher attendance in developing countries? A systematic review." London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Hanushek, Eric and Ludger Woessmann (2015) The Knowledge Capital of Nations: Education and the Economics of Growth. MIT Press: Cambridge, MA.

Jasper, Christian, Thanh-Tam Le and Jamie Bartram (2012) "Water and Sanitation in schools: A Systematic Review of the Health and educational Outcomes" *International Journal of Environmental Research and Public Health*. Vol. 9: 2772-2787.

Kabeer, N. C. Piza and L. Taylor (2012) "What are the economic impacts of conditional cash transfer programmes? A systematic review of the evidence. Technical report. London: EPPI-Centre, Social Science Research Unity, Institute of Education, University of London.

Krishnaratne, Shari, Howard White and Ella Carpenter (2013) "Quality education for all children? What works in education in developing countries". 3IE Working Paper 20. Available at: http://www.3ieimpact.org/evaluation/working-paper-eva/working-paper-20/

Kristjansson, Betsy, Vivian Robinson, Mark Petticrew, Barbara MacDonald, Julia Krasevec, Laura Janzen, Trish Greenhalgh, George A. Wells, Jessie MacGowan, Anna P. Farmer, Beverley J. Shea, Alain Mayhew and Peter Tugwell (2006). "School Feeding for Improving the Physical and Psychosocial Health of Disadvantaged Students". Campbell Systematic Reviews 2006:14.

Kremer, Michael, Conner Brannen and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World", *Science* 340:297-300.

Lochner, Lance (2011) "Non-Production Benefits of Education: Crime, Health, and Good Citizenship" NBER Working Paper No. 16722.

McEwan, Patrick (2015) "Improving learning in primary schools of developing countries: a meta-analysis of randomized experiments" Review of Educational Research 85(3): 353-394

Morgan, Claire, Anthony Petrosino, and Trevor Fronius. (2012) "A systematic review of the evidence of the impact of eliminating school user fees in low-income developing countries." London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Morgan, Claire, Anthony Petrosino and Trevor Fronius (2013) "A systematic review of the evidence of the impact of school voucher programms in developing countries." Londin: EPPI-Centre, Social Science Research Unity, Institue of Education, University of London.

Murnane, Richard, and Alejandro Ganimian. 2014. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations". NBER Working Paper 20284.

Saavedra, Juan Esteban and Sandra Garcia (2012) "Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries: A Meta-analysis." RAND Labor and Population working paper No. WR-921-1.

Spire, Elizabeth, Pia Britto, Terri Pigott, Eugene Roehlkepartain, Michael McCarthy, Yael Kidron, Janis Glover, Daniel Wagner, Julia Lane, and Mengli Song (2014) "Parental, Familial and Community Support Interventions to Improve Children's Literacy in Developing Countries: A Systematic Review" The Campbell Collaboration.

Taylor-Robinson D.C., N. Maayan, K. Soares-Weiser and S. Donegan, Garner P. (2012) "Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance (Review)" The Cochrane Library, Issue 11.

UNESCO Institute for Statistics (UIS) and UNICEF (2015). Fixing the Broken Promise of Education for All: Findings from the Global Initiative on Out-of-School Children. Montreal: UIS

UNESCO (2014) Education for All Global Monitoring Report. Teaching and Learning: Achieving Quality for All. France. UNESCO.

Yoong, Joanne, Lila Rabinovich, Stephanie Diepeveen (2012) "The impact of economic resource transfers to women versus men: a systematic review. Technical report." London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Key Definitions and Acronyms

Confidence Interval: A confidence interval gives a range of values in which the "true" value is most likely to lie. For example, the probability that the "true" value lies outside of the 95 percent confidence interval is 5%. Figures 8-11 provide both the 90 and 95 percent confidence intervals.

Difference-in-difference estimator (DD): A statistical technique used in econometrics which attempts to mimic an experimental research design using observational study data. It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the program participants to the average change over time for a comparable group of individuals who did not participate in the program.

Grey literature: A type of information or research output produced by organizations, outside of commercial or academic publishing, and distribution, channels.

Longitudinal data: data that following individuals or households over multiple time periods.

Matching estimators: An estimation technique that compares similar cases (e.g. students) to estimate the average effect of a treatment or intervention by comparing participants to non-participants in a program. Comparisons are made only between participants and nonparticipants who are judged to be sufficiently similar in their observed characteristics.

Omitted variable bias: This occurs when a variable that is correlated with both the dependent and one or more included independent variables is left out of (omitted from) a regression equation. The "*bias*" is created when the model compensates for the missing factor by over or underestimating the effect of one of the other factors.

Point Estimate: A point estimate is a "best estimate" of a program impact (or, more generally, of any type of parameter) using a sample of data from a population that includes program participants (and usually non-participants as well).

Randomized Control Trial (RCT): A type of scientific experiment, where the people being studied are randomly allocated to a control group and to one or more treatment groups. Random assignment of

intervention(s) is done after subjects have been assessed for eligibility and recruited, but before the intervention to be studied begins.

Regression analysis: A statistical process for estimating the relationships among variables.

Regression discontinuity design (RDD): A quasi-experimental evaluation method that estimates the causal effects of interventions by assigning or using a policy assignment as a cutoff or threshold above or below which an intervention is assigned. By comparing observations lying close to, and on both sides, of the threshold, it is possible to estimate the impact of the program for those individuals who are close to the threshold. An example is a program that provides assistance to households with an income less than a certain threshold. The impact of the program is estimated by comparing (potential) program participants whose income is slightly below the cutoff threshold to non-participants whose incomes are slightly above the cutoff.

Standard deviation: The standard deviation is a statistic that tells you how tightly various observations are clustered around the mean (average value) in a set of data. When the examples are tightly bunched together, so that the bell-shaped curve describing the data is steep, the standard deviation is small. When the examples are spread apart, so that the bell curve is relatively flat, the standard deviation is large.

Standardized test score: The number of standard deviations an observation is *above* or below the mean (average value). A positive standard score indicates an observation above the mean, while a negative standard score indicates an observation below the mean.

Statistically significant: A result is statistically significant if the relationship between two variables is caused by something other than random chance. For example, a significance level of 0.05 indicates that there is only a 5% chance of concluding that a difference exists when there is no actual difference.

Appendix 1

Tables of Coefficient Estimates for the Four Intervention Categories

Part A: Demand Side Interventions

Table A1. Summary of Impacts on Test Scores of Demand Side Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Information-Based Interventions					
Information on returns to schooling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Career counseling (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Cash Transfer Programs					
Conditional cash transfer					
RCTs	0 (0)	1 (1)	1 (1)	6 (3)	5
Non-RCT	1 (1)	3 (2)	3 (2)	0 (0)	3
Unconditional cash transfers (RCT)	0 (0)	1 (1)	1 (1)	0 (0)	1
Labeled cash transfer (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1

Promise of high school financial aid (RCT)	0 (0)	1 (1)	0 (0)	1 (1)	1
Scholarship Programs					
Merit-based scholarship (RCT)	0 (0)	1 (1)	0 (0)	6 (3)	4
Other Household Interventions					
Mother literacy classes (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Table A2. Demand for Schooling Estimates of Effect of CCTs on Test Scores

Author	Estimate description	Impact of CCTs on test scores (SD)	Standard error	Observations	Country
Baird, McIntosh and Ozler (2011)*	Girls in school at baseline, Age 13-22, English	0.140	0.054	2057	Malawi
	Girls in school at baseline, Age 13-22, Math	0.120	0.067	2057	Malawi
Baird et al (2013)*	Girls who dropped out at baseline, Age 13-22, English	0.131	0.070	729	Malawi
	Girls who dropped out at baseline, Age 13-22, Math	0.164	0.060	729	Malawi
Mo et al (2013)*	Boys & Girls, Grade 7, Math	0.010	0.10[18]	268	China
Barham et al (2013)*[19]	Boys, Age 9-12 at baseline, Combined Score	0.20	0.07	899	Nicaragua
Barrera-Osorio and Filmer	Boys and girls, poverty based, grade 6 math	-0.041	0.08	883	Cambodia

[18] This is a t-statistic.
[19] See paper summary for separate estimates for home language and math.

(2013)*					
	Boys and girls, merit based, grade 6 math	0.170	0.09	940	Cambodia
Baez and Camacho (2011)	Boys and girls, grade 11 overall	-0.025	0.019	131744	Colombia
Filmer and Schady (2014)	Scholarship effects on test scores	0.049	.080	1,869	Cambodia
Garcia and Hill (2010)	Boys and girls, age 7-12, grade 5 math test scores	1.460[20]	1.200	300	Colombia
	Boys and girls, age 7-12, grade 5 math language scores	1.320[21]	0.850	300	Colombia
	Boys and girls, age 13-17, grade 9 math language scores	-2.170[22]	0.980	111	Colombia
	Boys and girls, age 13-17, grade 9 math language scores	-1.480[23]	3.090	112	Colombia

*RCT

20 Not in terms of a standard deviation.
21 Not in terms of a standard deviation.
22 Not in terms of a standard deviation.
23 Not in terms of a standard deviation.

Table A3. Demand for Schooling Estimates of Effect of Merit-based Scholarships on Test Scores

Author	Estimate description	Impact of merit based scholarship on test scores (SD)	Standard error	Observations	Country
Kremer, Miguel and Thornton (2009)*	Girls grade 6 year end exam, Busia	0.27	0.16	2106	Kenya
Friedman, Kremer, Miguel and Thornton (2011)*	girls, age 17-21, Busia district only, four-five year follow-up survey (table 2)	0.196	0.089	1385	Kenya
Blimpo (2014)*	grade 10, combined score, individual incentive	0.260	0.110	1274	Benin
	grade 10, combined score, team target	0.240	0.140	1274	Benin
	Grade 10, combined score, team tournament	0.280	0.130	1274	Benin
Li et al. (2014)*	Boys and girls, grade 3-6, individual incentive	-0.061	0.078	829	China
	Boys and girls, grade 3-6, peer incentive	0.265	0.067	747	China

A4. Demand for Schooling Estimates of Unique Interventions on Test Scores

Author	Estimate Description	Impact of unique interventions on test scores (SD)	Standard Error	Observations	Country
Information-Based Interventions					
*Loyalka et al (2013)	Impact of info. on returns to school-ing boys & girls, grade 7, math	-0.005	0.046	10451	China
	Impact of career counseling on boys & girls, grade 7, math	-0.073	0.046	10451	China
Cash Transfer Programs					
Baird, McIntosh and Ozler (2011)	Impact of UCT on girls (in-school at baseline), age 13-22, English	-0.030	0.084	2057	Malawi
	Impact of UCT on girls (in-school at baseline), age 13-22, math	0.006	0.098	2057	Malawi
*Benhassine et al. (2014)	Impact of labeled cash transfer on boys and girls math test	0.080	0.052	3316	Morocco
*Yi et al. (2014)	Impact of promise of high school financial aid given	-0.02	0.06	1672	China

	to grade 7 students on math achievement				
Other Household Interventions					
*Banerji, Berry and Shotland (2013)	Impact of mother lit. classes on boys & girls combined math & lit. tests	0.0176	0.0176	18282	India
	Impact of mother training on child learning on boys and girls combined math and literacy tests	0.0387	0.0183	18282	India
	Impact of both mother literacy classes and mother training on child learning on boys and girls combined math and literacy tests	0.0632	0.0171	18282	India

*RCT study

Table A5. Summary of Impacts on Time in School of Demand Side Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Information-Based Interventions					
Information on returns to educ. (RCT)	0 (0)	1 (1)	1 (1)	2 (1)	2
Career counseling (RCT)	1 (1)	0 (0)	0 (0)	0 (0)	1
School Counseling (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Cash Transfer Programs					
Conditional cash transfer					
RCT	0 (0)	1 (1)	4 (3)	34 (15)	15
Non-RCT	0 (0)	0 (0)	0 (0)	18 (9)	9
Unconditional cash transfers (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1
Labeled cash transfer (RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Eliminating school fees	0 (0)	2 (2)	1 (1)	1 (1)	2
Scholarship Programs					
Merit-based scholarship (RCT)	0 (0)	0 (0)	1 (1)	3 (2)	2
Other Household Interventions					
Mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1

Mother literacy class (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Combined mother literacy class and mother class on child learning (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Adult literacy program (non-RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Female sanitary products (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Bicycle program (non-RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Matching remittances funds for education (RCT)	0 (0)	0 (0)	1 (1)	1 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

A6. Demand for Schooling Estimates of Information on Returns to Education on Time in School

Author	Estimate description	Impact of Information on returns to educ. on time in school	Standard error	Observations	Country
Loyalka et al. (2013)*	Boys & Girls, grade 7, dropout rate	0.011[24]	0.009	11,633	China
Jensen (2010)*	Boys, grade 8, returned next year	0.041	0.023	2241	Dominican Republic
	Boys, grade 8, finished secondary school	0.023	0.020	2205	Dominican Republic
	Boys, grade 8, increase in years of schooling	0.200	0.082	2082	Dominican Republic

*RCT study

[24] This positive effect on dropout rate is reported as a negative impact on time in school in summary table.

A7. Demand for Schooling Estimates of CCTs on Time in School

Author	Estimate description	Impact of CCTs on measure of time in school	Standard error	Observations	Country
Baird, McIntosh and Ozler (2011)*	Girls (in-school at baseline), age 13-22, terms enrolled over two years	0.535	0.129	852	Malawi
	Girls (in-school at baseline), age 13-22, attendance	0.080	0.035	319	Malawi
Baird et al (2013)*	Girls (dropout at baseline), age 13-22, number of terms enrolled	2.348	0.163	749	Malawi
Mo et al (2013)*	Boys and Girls, Grade 7, probability of dropout	-0.08[25]	-2.53[26]	300	China
Barham et al (2013)*	Boys, aged 9-12, increase in grades attained over late treatment group – long term (age and stratification controls)	0.501	0.274	980	Nicaragua
Behrman, Parker, and Todd (2009)*	Girls, age 6-14 in 2003, impact of 1.5 years of program exposure after initial treatment	0.049	0.08	NA	Mexico

[25] Counted as a positive impact on time in school in summary table.
[26] This is a t-statistic.

	on grades of school completed				
	Boys, age 6-14 in 2003, impact of 1.5 years of program exposure after initial treatment on grades of school completed	0.051	0.06	NA	Mexico
	Girls, age 6-14 in 2003, impact of 5.5 years of program exposure after initial treatment on grades of school completed	0.240	0.060	NA	Mexico
	Boys, age 6-14 in 2003, impact of 5.5 years of program exposure after initial treatment on grades of school completed,	0.320	0.070	NA	Mexico
Behrman, Parker and Todd (2011)*	Girls, age 15-21 in 2003, impact of 1.5 years of program exposure five years after initial treatment on grades completed	0.201	0.047	7870	Mexico
	Boys, age 15-21 in 2003, impact of 1.5 years of program exposure five years after initial treatment on grades completed	0.180	0.045	8384	Mexico
	Girls, age 17-18 in 2003, impact of 5.5 years of program exposure five years after initial treatment on grades completed.	0.750	0.180	NA	Mexico
	Boys, age 17-18 in 2003, impact of 5.5 years of program exposure five years after	0.930	0.210	NA	Mexico

	initial treatment on grades completed.				
Schultz (2004)*	Enrollment rate (completed grade 6)	0.087	0.040[27]	4800	Mexico
Barrera-Osorio, et al (2011)*	Attendance (basic treatment)	.032	.007	5799	Columbia
	Attendance (savings treatment)	.027	.008	5799	Columbia
	Attendance (tertiary treatment)	.056	.020	930	Columbia
	Enrollment (basic treatment)	.011	.010	8980	Columbia
	Enrollment (savings treatment)	.040	.011	8980	Columbia
	Enrollment (tertiary treatment)	.037	.019	1735	Columbia
Filmer and Schady (2008)*	Girls, grade 8 enrollment	0.238	0.129	3623	Cambodia
Barrera-Osorio and Filmer (2013)*	boys and girls, percentage likely to reaching grade 6	0.170	0.04	883	Cambodia
	boys and girls, more grade completed	0.332	0.11	831	Cambodia
	boys and girls, percentage likely to	0.120	0.04	940	Cambodia

[27] This is not a standard error. Rather it is a significant level indicating the statistical probability that the observed difference could have occurred randomly.

	reaching grade 6				
	boys and girls, more grade completed	0.182	0.10	897	Cambodia
Gitter, Barham (2007)*	Boys and girls, percentage increase in enrollment	0.166	0.030	4593	Nicaragua
Galiani and McEwan (2013)*	Children grades 1-4, percentage increase in enrollments.	0.083	0.028	120411	Honduras
Chaudhury Friedman and Onishi (2013)*	Increase in enrollment rate, 6-11 year olds	0.045	0.014	1570	Philippines
	Increase in enrollment rate, 12-14 year olds	0.039	0.024	809	Philippines
	Increase in enrollment rate, 15-17 year olds	-0.027	0.041	713	Philippines
	Percentage point increase in those who attended >85%, 6-11 year olds	0.038	0.017	1463	Philippines
	Percentage point increase in those who attended >85%, 12-14 year olds	0.049	0.020	680	Philippines
	Percentage point increase in those who attended >85%, 15-17 year olds	0.076	0.021	410	Philippines

Dammert (2009)*	Increase in school attendance percentage points over the baseline for girls, after one year of program	.117	0.030	1359	Nicaragua
	Increase in school attendance percentage points over the baseline for boys, after one year of program	.060	0.030	1359	Nicaragua
Dubois, de Janvry and Sadoulet (2012)*	Probability of grade continuation during school year 1997-1998, primary school, girls	0.031	0.008	13894	Mexico
	Probability of grade continuation during school year 1997-1998, primary school, boys	0.031	0.008	13894	Mexico
	Probability of grade continuation during school year 1997-1998, secondary school, girls	0.034	0.005	13894	Mexico
	Probability of grade continuation during school year 1997-1998, secondary school, boys	0.032	0.006	13894	Mexico
Chaudhury and	Percent change in enrollment in grade 6-8	8.66	1.96[28]	NA	Pakistan

[28] This is a t-statistic.

Parajuli (2010)	females				
Baez and Camacho (2011)	High school completion rate, percentage point(s) more likely to finish	0.039	0.005	624,028	Colombia
Glewwe and Kassouf (2012)	Boys and girls, grade 1-4 percentage change in enrollment	0.0274	0.0018	699255	Brazil
	Boys and girls, grade 5-8 percentage change in enrollment	0.0317	0.0031	182191	Brazil
	Boys and girls, grade 1-4 percentage change in dropping out	-0.310[29]	0.058	698229	Brazil
	Boys and girls, grade 5-8 percentage change in dropping out	-0.267[30]	0.075	182006	Brazil
DeJanvry, Finan and Sadoulet (2012)	Boys and girls, percentage change in dropout rates	-0.096[31]	0.004	344107	Brazil
Attanasio et al (2010)	Boys and girls, grade 8-13 percentage change in enrollment, rural	0.0282	0.0111	3648	Colombia
	Boys and girls, grade 8-13 percent-age	0.0140	0.0066	2579	Colombia

[29] Counted as a positive impact on time in school in summary table.
[30] Counted as a positive impact on time in school in D2.
[31] Counted as a positive impact on time in school in D2.

	change in enrollment, urban				
	Boys and girls, grade 14-17 percent-age change in enrollment, rural	0.0662	0.0232	1873	Colombia
	Boys and girls, grade 14-17 percent-age change in enrollment, urban	0.0470	0.0123	1439	Colombia
Filmer and Schady (2014)	Additional grades of school completed after receiving a scholarship (measured in grades)	0.598	.140	1,869	Cambodia
Behrman et al. (2012)	Boys age 6-20, increase in grades attained after 1 year exposure, urban	0.110	0.020	4470	Mexico
	Girls age 6-20, increase in grades attained after 1 year exposure, urban	0.110	0.020	4489	Mexico
	Boys age 6-20, percentage point increase in enrollment rates after 1 year exposure, urban	3.600	1.000	4604	Mexico
	Girls age 6-20, percentage point increase in enrollment rates after 1 year exposure, urban	2.800	1.100	4626	Mexico

Garcia and Hill (2010)	Boys and girls, age 7-12, days absent from school last month	-0.504[32]	0.086	5850	Colombia
Levy and Ohls (2010)	Boys and girls, age 7-17, increase in days attended over a 20-day reference period	0.506	0.220	6790	Jamaica

*RCT

[32] This negative coefficient is reported as positive and significant in summary table because a decrease in days absent is a positive increase in time in school.

Table A8. Demand for Schooling Estimates of Eliminating School Fees on Time in School

Author	Estimate description	Impact of intervention on measures of time in school	Standard error	Observations	Country
Borkum (2012)	Cost of schooling on 2007 Enrollment, measured on 10th day of school, Primary schools, Eastern Cape	-.004	.014	4179	South Africa
	Cost of schooling on 2007 Enrollment, measured on 10th day of school, Secondary schools, Eastern Cape	.001	.021	761	South Africa
*Yi et al. (2014)	Impact of promise of high school financial aid given to grade 7 students on percentage point dropout rate.	3.0	1.8	1892	China
	Impact of promise of high school financial aid school given to grade 9 students on percentage point increase in matriculation to high school	7.9*	4.6	380	China

*RCT study

Table A9. Estimates of effect of Merit-based Scholarships on Time in School

Author	Estimate Description	Impact of merit-based Scholarship on measure of time in school	Standard Error	Observations	Country
Kremer, Miguel and Thornton (2009)*	Girls grade 6, average student school participation rate, Busia	0.032	0.018	2033	Kenya
Friedman, Kremer, Miguel, Thornton (2011)*	Girls age 13-17; age 17-21 in follow up, attended some secondary school	0.086	0.041	1385	Kenya
	Girls age 13-17; age 17-21 in follow up, still in school	0.078	0.044	1385	Kenya
	Girls age 13-17; age 17-21 in follow up, grades completed.	0.086	0.103	1385	Kenya

*RCT study

A10. Demand for Schooling Estimates of Unique Interventions on Time in School

Author	Estimate description	Impact of unique intervention on measures of time in school	Standard error	Observations	Country
Information-Based Interventions					
*Loyalka et al. (2013)	Impact of career counseling on boys & girls, grade 7, dropout rate [33]	0.017	0.007	11,633	China
*Huan et al. (2014)	Impact of school counseling on percentage point dropout rate for boys and girls, after one semester of treatment[34]	-0.02	0.009	7495	China
	Impact of school counseling on percentage point dropout rate for boys and girls, after two semesters of treatment[35]	-0.01	0.008	7495	China
Cash Transfer Programs					
Baird, McIntosh and Ozler (2011)*	Impact of UCT on girls (in-school at baseline), age 13-22, enrollment	0.231	0.136	852	Malawi

[33] The coefficient estimate indicates the intervention significantly increased the dropout rate. This is reported as a negative effect on time in school in summary table.
[34] The coefficient estimate for dropout rate is negative. It will be reported as a positive impact on time in school in summary table.
[35] The coefficient estimate for dropout rate is negative. It will be reported as a positive impact on time in school in summary table.

	Impact of UCT on girls (in-school at baseline), age 13-22, attendance	0.058	0.037	319	Malawi
*Benhassine et al. (2014)	Impact of labeled cash transfer on boys and girls percentage point increase in attending school after 2 years of program duration	0.074	0.016	11074	Morocco
	Impact of labeled cash transfer on boys and girls percentage point decrease in dropped out rate after 2 years of program duration[36]	-0.076	0.012	5998	Morocco
Other Household Interventions					
*Banerji, Berry and Shotland (2014)	Impact of mother literacy classes on boys and girls enrollment	0.0122	0.0118	25053	India
	Impact of mother training on child learning on boys and girls enrollment	0.0136	0.0116	25053	India
	Impact of both mother literacy classes and mother training on child learning on boys and girls enrollment	0.0144	0.0118	25053	India

[36] The coefficient estimate for dropout rate is negative. It will be reported as a positive impact on time in school in summary table.

Handa (2002)	Boys and girls, adult literacy program on probability of current enrollment	0.135	1.11[37]	6609	Mozambique
Oster and Thornton (2011)*	Impact of sanitary products on attendance	-0.010	.017	31,693	Nepal
Muralidharan and Prakash (2013)	Impact of bicycle program on girls, age 14-16, secondary school enrollment	0.0515	0.0252	30112	India
*Ambler, Aycinena and Yang (2014)	Impact of 3:1 matching remittance funds for education on percentage point increase in enrollment	0.0309	0.0398	728	El Salvador

*RCT study

[37] Z-statistic.

Part B: School Input Interventions

Table B1. Summary of Impacts on Test Scores of School Inputs

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Interventions that Increase Access to Schools					
Building new schools					
RCTs	0 (0)	0 (0)	2 (1)	2 (1)	2
Other high quality studies	0 (0)	0 (0)	0 (0)	1 (1)	1
Hours per school day	0 (0)	1 (1)	0 (0)	3 (2)	2
Pedagogical Materials and Facilities					
Textbooks (RCTs)	0 (0)	2 (1)	1 (1)	0 (0)	2
Flipcharts (RCT)	0(0)	0 (0)	1(1)	0 (0)	1
Provision of libraries (RCT)	1(1)	1 (1)	0 (0)	0(0)	1
Multilevel learning materials (RCT)	0(0)	1 (1)	0 (0)	2 (1)	1
Multi-level teaching materials/parent teacher partnerships (RCT)	0 (0)	0 (0)	0 (0)	3 (1)	1
Teacher Quantity and Quality					
Pupil-teacher ratio					
RCTs	0 (0)	0 (0)	1 (1)	0 (0)	1
Non-RCT	3 (2)	1(1)	0 (0)	0 (0)	2
Provision of Food					
School meals					
RCTs	0 (0)	0 (0)	0 (0)	3 (2)	2

Other high quality studies	0 (0)	2 (1)	1 (1)	1 (1)	2
Take home rations (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
School feeding /parent-teacher partnerships (RCT)	0 (0)	0 (0)	1 (1)	2 (1)	1
Medical Services					
Deworming medicine (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Iron supplements (RCTs)	0 (0)	1 (1)	2 (2)	1 (1)	2
Provision of eyeglasses (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Large-scale Provision of Resources					
Attending an elite public school (Non-RCT)	0(0)	0(0)	1(1)	1(1)	2
Infrastructure/materials/training (Non-RCT)	0 (0)	0 (0)	2(1)	2(1)	1
Unexpected school block grant (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Expected school block grant (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Support circles (RCT)	0(0)	0 (0)	1(1)	0 (0)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

B2. School Input Estimates of Building New Schools on Test Scores

Author	Estimate description	Impact of building new schools on test scores (SD)	Standard error	Observations	Country
Burde and Linden (2013)*	Girls, primary school, combined math and language skills	0.661	.0900	687	Afghanistan
	Boys, primary school, combined math and language skills	0.413	Not reported	709	Afghanistan
Dumitrescu et al. (2011)*	Boys and girls math test score	0.028	n/a	13686	Niger
	Boys and girls French test score	0.044	n/a	13969	Niger
Kazianga et al. (2013)	Boys and girls, combined math and French	0.409	0.050	17970	Burkina Faso

*RCT study

Table B3. School Input Estimates of Length of School Day on Test Scores

Author	Estimate description	Impact of length of school day on test scores (SD)	Standard error	Observations	Country
Bellei (2009)	Full day effect on math test scores	.07	≤.027	210,235	Chile
	Full day effect on language test scores	.05	≤.020	210,235	Chile
Orkin (2013)	Four Additional Hours on Reading Test Score (odds ratio)	0.96	0.35	1794	Ethiopia
	Four Additional Hours on Numeracy Score (odds ratio)	2.74	0.87	1794	Ethiopia

Table B4. School Input Estimates of Textbooks on Test Scores

Author	Estimate description	Impact of textbooks on test scores (SD)	Standard error	Observations	Country
Glewwe & Kremer (2009)	Primary School (all subjects)	.023	.087	24,132	Kenya
Sabarwal et al. (2014)	Grade 5 Math Scores	-0.194	0.187	7,746	Sierra Leone

	Grade 5 English Scores	-0.234	.204	7,746	Sierra Leone

Table B5. School Input Estimate of Pupil-Teacher Ratio on Test Scores

Author	Estimate description	Impact of P-T ratio on test scores (SD)	Standard error	Observations	Country
Duflo, Dupas, Kremer (2012)	Child in Extra teacher School, Total Score, (Non-local TSC Teacher)	0.087	0.098	6533	Kenya
	Child in ETP School, Total Score, (Local PTA Teacher)	0.294	0.105	6,533	Kenya
Urquiola (2006)	Pupil-teacher ratio on Language Score	-0.33	.17	3594	Bolivia
	Pupil-teacher ratio on Math Score	-0.25	.20	3594	Bolivia
Urquiola & Verhoogen (2009)	Pupil-teacher ratio on Math Score	-0.7***	0.3	1623	Chile
	Pupil-teacher ratio on language Score	-0.6**	0.3	1623	Chile

Table B6. School Input Estimate of School Feeding Programs on Test Scores

Author	Estimate description	Impact of school feeding program on test score (SD)	Standard error	Observations	Country
McEwan (2013)	Grade 4 language score in 2005 (points)	-0.1071*	0.076*	4785	Chile
	Grade 4 math score in 2005 (points)	-0.1303*	0.090*	4767	Chile
Adrogue et al (2013)	Math test score, Grade 3; School and Provincial time-varying fixed-effects.	0.0668*	0.063*	1793	Argentina
	Language test score, Grade 3; School and Provincial time-varying fixed-effects.	0.1722*	0.063*	1793	Argentina
Tan et al (1999)	School Feeding, Grade 1, Math Score, IV estimation	0.248	.091*	1676	Philippines
	School Feeding, Grade 1, Filipino Score, IV estimation	0.160	.088*	1676	Philippines
Kazianga, deWalque and Alderman (2012)	School meals on math score (proportion of correct answers 0-1)	0.096	0.036	7615	Burkina Faso

B7. School Input Estimate of Iron Supplements on Test Scores

Author	Estimate description	Impact of iron supplements program on test score (SD)	Standard error	Observations	Country
Sylvia et al (2013)	Children, Grades 4 and 5, Subsidy only group, with additional controls (anemia)	0.05	0.085	5656	China
	Children, Grades 4 and 5, Health incentive group, with additional controls (anemia)	-0.11	0.081	5656	China
Luo et al (2012)	Children in multivitamin Supplement group, Math Score	1.09	1.86[38]	3,661	China
	Children in anemia information Treatment group, Math Scores	0.16	0.22[39]	3,661	China

[38] t-statistic
[39] t-statistic

Table B8. School Input Estimate of Attending an Elite Public School on Test Scores

Author	Estimate description	Impact of attending an elite public school on test score (SD)	Standard error	Observations	Country
Pop-Eleches and Miguel Urquiola (2013)	Baccalaureate grade after being just above a cutoff	.018[40]	.002	1,256,038	Romania
Luas and Mbiti (2014)	Boys and girls, impact of having graduated from an elite government school on secondary exit exam	0.014	0.052	12704	Kenya

[40] Not in terms of a SD.

Table B9. School Input of Unique Interventions on Test Scores

Author	Estimate description	Impact of unique interventions on test scores (SD)	Standard error	Observations	Country
Pedagogical Materials and Facilities					
Glewwe et al. (2004)*	Effect of flip chart school on flip-chart subject tests, grades 6-8	0.008	0.0308	141,698	Kenya
Borkum, He and Linden (2012)*	Effect of all libraries on language test scores	-0.044	0.049	n/a	India
	Effect of traveling libraries on language test scores	-0.218	0.088	n/a	India
Tan, Lane and Lassibille (1999)*	Multilevel learning materials, Grade 1, Math Score, IV estimation	-0.045	0.38[41]	1676	Philippines
	Multilevel learning materials, Grade 1, Filipino Score, IV estimation	0.234	2.05[42]	1676	Philippines
	Multilevel learning materials, Grade 1, English Score, IV estimation	0.548	4.71[43]	1676	Philippines

[41] t-statistic
[42] t-statistic
[43] t-statistic

Tan, Lane and Lassibille (1999)*	Multilevel learning materials with P-T partnership, Grade 1, Math Score, IV estimation	0.21	1.83[44]	1676	Philippines
	Multilevel learning materials with P-T partnership, Grade 1, Filipino Score, IV estimation	0.225	2.02[45]	1676	Philippines
	Multilevel learning materials with P-T partnership, Grade 1, English Score, IV estimation	0.754	6.64[46]	1676	Philippines
Provision of Food					
Kazianga, de Walque and Alderman (2012)*	Take-home rations on math score (proportion of correct answers 0-1)	0.084	0.041	7615	Burkina Faso
Tan, Lane and Lassibille (1999)*	School Feeding with P-T partnership, Grade 1, Math Score, IV estimation	0.347	3.74[47]	1676	Philippines
	School Feeding with P-T partnership,	0.114	1.28[48]	1676	Philippines

[44] t-statistic
[45] t-statistic
[46] t-statistic
[47] t-statistic
[48] t-statistic

	Grade 1, Filipino Score, IV estimation				
	School Feeding with P-T partnership, Grade 1, English Score, IV estimation	0.442	4.89[49]	1676	Philippines
Medical Services					
Miguel and Kremer (2004)*	Impact of deworming after two years on test score	0.001	0.073	24958	Kenya
Glewwe and Park (2011)*	Effect of eyeglasses on average test scores	0.158	0.078	2474	China
Large-scale Provision of Resources					
Chay, McEwan and Urquiola (2005)	Effect of infrastructure/materials/training on math scores after 2 years	0.25	0.53	2,644	Chile
	Effect of infrastructure/materials/training on math scores after 4 years	2.09	0.6	2,591	Chile
	Effect of infrastructure/materials/training on language scores after 2 years	0.54	0.49	2,644	Chile
	Effect of infrastructure/materials/training on language scores after 4 years	2.10	0.52	2,591	Chile

[49] t-statistic

*Das et al. (2013)	Impact of unanticipated school block grant one year after intervention on boys and girls combined test score	0.085	0.038	27704	India
	Impact of anticipated school block grant two years after intervention on boys and girls combined test score	0.053	0.045	19872	India
Pridmore and Jere (2011)*	Impact of circles of support on math test scores	0 .59	(-0.253 - 0.442)[50]	2767	Malawi

*RCT

[50] Confidence Interval

Table B10. Summary of Impacts on Time in School of School Inputs

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Interventions that Increase Access to Schools					
Building new schools					
RCTs	0 (0)	0 (0)	0 (0)	4 (3)	3
Non-RCT	0 (0)	0 (0)	0 (0)	3 (3)	3
Providing school uniforms (RCT)	1(1)	0 (0)	0 (0)	0 (0)	1
Pedagogical Materials and Facilities					
Textbooks (RCTs)	0 (0)	0 (0)	2 (2)	0 (0)	2
Provision of libraries (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Multi-level teaching materials (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Multi-level teaching materials/parent teacher partnerships (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Teacher Quantity and Quality					
Extra teacher/materials (Non-RCT)	0 01)	0 (0)	0 (0)	2 (1)	1
Provision of Food					
School meals					
RCTs	0 (0)	0 (0)	1 (1)	2 (2)	3
Other high quality studies	0 (0)	3 (2)	1 (1)	0 (0)	2
Take home rations	0 (0)	0 (0)	1 (1)	1 (1)	2
School feeding /parent-teacher partnerships* (RCT)	0 01)	0 (0)	1 (1)	0 (0)	1

Medical Services					
Deworming Medicine (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1
Health Insurance (Non-RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Large-scale Provision of Resources					
School infrastructure investments (RCT)	0 (0)	0 (0)	1 (1)	0 (0)	1
Circles of Support (RCT)	0 (0)	0 (0)	0 (0)	1 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Table B11. School Input Estimates of Building New Schools on Time in School

Author	Estimate description	Impact of building new schools on measures of time in school	Standard error	Observations	Country
Alderman, Kim and Orazem (2003)*[51]	girls, age 5-8, enrollment rate, urban	24.8	0.700	NA	Pakistan
Burde and Linden (2013)*	Girls primary enrollment rate	.515	0.082	693	Afghanistan
	Boys primary enrollment rate	.346	Not reported	797	Afghanistan
Dumitrescu et al. (2011)*	Boys and girls percentage point increase in primary enrollment	0.037	0.022	13969	Niger
Handa (2002)	Boys and girls, number of new schools built on probability of current enrollment	0.003	3.68[52]	6409	Mozambique

[51] See summary of paper for rural results. Control group was not used in rural results, so not presented here.
[52] Z-statistic. Coefficient estimate is from a probit.

Kazianga et al. (2013)	Boys and girls, enrollment	0.185	0.025	17970	Burkina Faso
Duflo (2001)	Boys and girls, primary school, years of education for each school build per 1000 children	0.188	0.0289	78470	Indonesia

*RCT study

Table B12. School Input Estimate of Textbooks on Time in School

Author	Estimate description	Impact of textbooks on measure of time in school	Standard error	Observations	Country
Glewwe, Kremer & Moulin (2009)	Primary School Dropouts (reported as positive for time in school above)	-0.7	NA	NA	Kenya
	Primary School Repeaters (reported as positive for time in school above)	-4.1	Significant	NA	Kenya
Sabarwal et al. (2014)	Grade 5 Total Attendance (treatment on the treated)	0.051	0.075	231	Sierra Leone

Table B13. School Input Estimate of School Feeding Program on Time in School

Author	Estimate description	Impact of school feeding program on measure of time in school	Standard error	Observations	Country
Alderman et al. (2012)	School Feeding Program, Ages 6-13 (attendance percentage point)	0.072	0.047	2,266	Uganda
McEwan (2013)	Enrollment in grade 1	-0.736	1.2	3437	Chile
	Enrollment in grades 1-8	-4.139	11.066	3437	Chile
Afridi (2011)	On-site feeding attendance (pp)	.018	.038	112	India
	On-site feeding enrollment (pp)	-10.026	17.395	112	India
Tan et al (1999)	Grades 1-5, School Feeding (probability of dropping out)	-0.254	0.56	8229	Philippines
Kazianga, de Walque and Alderman (2012) RCT	School meals on enrollment of children 6-15	0.039	0.017	8753	Burkina Faso

Table B14. School Input of Unique Interventions on Time in School

Author	Estimate description	Impact of unique interventions on measures of time in school	Standard error	Obser-vations	Country
Interventions that Increase Access to Schools					
Hidalgo et al. (2013)*	Effect of provision of school uniforms on attendance (in percentage points)	-0.022	0.009	9851	Ecuador
Pedagogical Materials and Facilities					
Borkum, He and Linden (2012)*	Impact of libraries on attendance	0.000	0.004	NA	India
Tan, Lane and Lassibille (1999)*	Impact of multi-level teaching materials on probability of dropping out	-0.428	1.71[53]	8229	Philippines
	Impact of multi-level teaching materials with parent teacher partnerships on probability of dropping out	-0.410	1.15[54]	8229	Philippines
Teacher Quantity and Quality					
Chin (2005)	Impact of providing an extra teacher and teacher learning packet on primary school completion, girls	0.0120	0.0028	153,024	India

[53] t-statistic
[54] t-statistic

	Impact of providing an extra teacher and teacher learning packet on primary school completion, boys	0.0100	0.0027	186,607	India
Provision of Food					
Tan, Lane and Lassibille (1999)*	Impact of school feeding with parent teacher partnerships on probability of dropping out	-0.311	1.40	8229	Philippines
Medical Services					
Miguel and Kremer (2004)*	Impact of deworming on school participation after two years	0.034	0.021	56487	Kenya
Chen and Zhe Jin (2012)	Impact of health insurance on school enrollment	0.004	1.09	1420685	China
Large-scale Provision of Resources					
Newman et al. (2002)	Impact of school infrastructure improvement on grade 7 dropout rate	-3.900	0.260[55]	NA	Bolivia
Pridmore and Jere (2011)	Impact of circles of support on dropout rates	0.46	(.367-.827)[56]	2767	Malawi

[55] p-value
[56] Confidence Interval

Part C. Teaching Pedagogy Intervention

Table C1. Summary of Impacts on Test Scores of Pedagogy Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Teaching at right level/Supplemental instruction (all RCTs)	0 (0)	0 (0)	0 (0)	4 (3)	3
Tracking/Streaming (RCT)	0 (0)	0 (0)	0 (0)	2 (1)	1
Computer, Electronic Games, and Access to Technology					
RCTs	1 (1)	0 (0)	4 (4)	13 (8)	10
Other high quality studies	3 (1)	0 (0)	0 (0)	0 (0)	1
Short-term reading program (RCT)	0 (0)	0 (0)	2 (1)	2 (1)	1
New literacy curriculum (RCTS)	0 (0)	0 (0)	1 (1)	1 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Table C2. Pedagogy Estimate of Teaching at the Right Level/Supplemental Instruction on Test Scores

Author	Estimate description	Impact of intervention on test scores (SD)	Standard error	Observations	Country
Banerjee et. al (2010)*	Boys and girls, percentage likely to develop math or reading skills	0.012	0.007	15609	India
Banerjee et al. (2007)*	Boys and girls, grade 3-4 combined math and English, after 1 year	0.138	.047	12855	India
	Boys and girls, grade 3-4 combined math and English, after 2 years	0.284	.060	21936	India
Jayachandran (2014)	Boys and girls, grade 10, passing school-leaving exam	-0.117[57]	0.045	28650	Nepal

*RCT

[57] This is not a standardized test score measure. It is an indicator of whether a student passed the school-leaving exam.

Table C3. Pedagogy Estimates of Computer, Electronic Games and Access to Technology on Test Scores

Author	Estimate description	Impact of intervention on test scores (SD)	Standard error	Observations	Country
Banerjee et al. (2007)*	Math (Year 3) grade 4	0.475	0.068	5523	India
	Total (Year 3) grade 4	0.225	0.051	5523	India
Barrera-Osorio et al. (2009)*	Total Normalized Scores (Spanish & Math)	0.109	0.104	5201	Colombia
Beuermann et al. (2013)*	Raven's progressive matrices	0.06	0.05	2756	Peru
Lai et al. (2011)*	Math test score, 3rd grade students, baseline and household controls	0.14	0.04	2157	China
	Chinese test score, 3rd grade students, baseline and household controls	0.01	0.05	2157	China
Lai et al. (2013)*	Mandarin testing, 3rd Grade, baseline and household controls	0.20	0.07	1717	China
	Math testing, 3rd Grade, baseline and household controls	0.22	0.07	1717	China
Linden (2008)*	Total Score In-School CAL program	-0.556	0.247	1640	India

	Total Score Out-of-School CAL program	0.290	0.161	1640	India
Mo et al. (2014a)*	Fifth grade math test scores	0.26	0.08	1388	China
	Third grade math test scores	0.25	0.08	1038	
Mo et al. (2014b)*	Standardized post-CAL math test score – All grades	0.12	0.05	2613	China
	Standardized post-CAL math test score – Third grade	0.18	0.08	1124	China
	Standardized post-CAL math test score – Fifth grade	0.07	0.07	1489	China
	Standardized post-CAL math test score – Fifth grade up to the 70th percentile in post-CAL math scores	0.11	0.07	909	China
Mo et al. (2013)*	Marked computer skills, no controls	0.32	0.12	250	China
	Marked computer skills, household and demographic controls	0.33	0.10	250	China
	Standardized math score, no controls	0.07	0.11	250	China
	Standardized math score, household and demographic controls	0.17	0.10	250	China
Yang et al. (2013)*	Standardized post CAL test score – standardized baseline score Rural	0.12	0.06	2613	China

	China (T2)				
	Standardized post CAL math test – Beijing migrant schools.	0.11	0.05	2157	China
Malamud & Pop-Eleches (2011)[58]	Effect of home computer voucher on Math GPA	-0.243	0.104	4,179	Romania
	Effect of home computer voucher on Romanian GPA	-0.266	0.112	4,201	Romania

*RCT estimate

[58] Note: Coefficients for Malamud & Eleches are not reported as changes in standard deviation, but rather nominal GPA points.

Table C4. Pedagogy Estimate of Unique Intervention on test scores.

Author	Estimate description	Impact of unique intervention on test scores (SD)	Standard error	Observations	Country
Duflo et al. (2011)	Impact of student enrolled in tracking school on short-run test score	0.176	0.077	5,001	Kenya
	Impact of student enrolled in tracking school on long run test score	0.178	0.073	5,001	Kenya
Abeberese et al (2014)*	Impact of short-term reading program on reading test after 4 months	0.13	0.05	5,228	Philippines
	Impact of short-term reading program on reading test after 7 months	0.06	0.03	4,887	Philippines
	Impact of short-term reading program on math test after 4 months	0.03	0.05	5,228	Philippines
	Impact of short-term reading program on social studies test after 7 months	0.06	0.06	4,887	Philippines
Lucas et al. (2014)*	Impact of new literacy	0.024	0.032	5302	Kenya

	treatment on written literacy test				
	Impact of new literacy treatment on oral literacy	0.077	0.042	5305	Kenya
	Impact of new literacy treatment on written literacy test	0.199	0.054	3596	Uganda
	Impact of new literacy treatment on oral literacy	0.179	0.047	3575	Uganda

*RCT

Table C5. Summary of Impacts on Time in School of Pedagogy Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Teaching at the right level (RCT)	0 (0)	1 (1)	0 (0)	0 (0)	1
Computer, electronic games, and access to technology (RCT)	0 (0)	0 (0)	2 (1)	0 (0)	1

1. Figures are number of estimates; figures in parentheses are number of papers/studies.

Table C6. Pedagogy Estimates of Unique Interventions on Time in School

Author	Estimate description	Impact of unique intervention on measure of time in school	Standard Error	Observations	Country
Banerjee et. al (2010)*	Impact of tutoring with information campaign on days present in last 14	-0.314	0.371	5,555	India
Cristia et al. (2014)	Impact of technology in schools on grade 7 dropout rate	-0.038	0.191	33,583	Peru
	Impact of technology in schools on grade 7 enrollment	0.007	1.753	6,749	Peru

Part D. School Governance Interventions

Table D1 – Summary of Impacts on Test Scores of Governance Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Monitoring (All RCTs)	0 (0)	1 (1)	5 (3)	1 (1)	4
School-based management					
RCTs	0 (0)	9 (3)	7 (3)	2 (2)	5
Other high quality studies	0 (0)	1 (1)	1 (1)	1 (1)	2
Teacher performance pay					
RCTs	0(0)	1 (1)	2 (1)	5(2)	3
Other high quality studies	0 (0)	0 (0)	1 (1)	1 (1)	1
Contract teachers (al RCTs)	0 (0)	0 (0)	0 (0)	3 (2)	2
Private school (vouchers)					
RCTs	0 (0)	0 (0)	3 (2)	2 (2)	3
Other high quality studies	0 (0)	2 (1)	2 (1)	0 (0)	2

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Table D2. Governance Intervention Estimates of Monitoring on Test Scores

Author	Estimate description	Impact of monitoring on test scores (SD)	Standard error	Observations	Country
Duflo, Hanna and Ryan (2012)*	Boys and girls, test scores	0.17	.09	1760	India
Muralidharan and Sundararaman (2010)*	Boys and girls, math test scores	-0.018	0.048	24386	India
	Boys and girls, language (Telugu) test scores	0.022	0.044	24405	India
Banerjee et al. (2010)*	Boys and girls, percentage likely to read stories	0.003[59]	0.010	15609	India
	Boys and girls, percentage likely to subtract or divide	0.006[60]	0.009	15592	India
Pandey, Goyal and Sundararaman	Boys and girls, percentage who can completed a literacy or math task, averaged across gender, and grade,	0.0015[61]	NA	NA	India

[59] Not a standard deviation measure of a test score.
[60] Not a standard deviation measure of a test score.
[61] Not a standard deviation measure of a test score.

(2009)*	MP and UP district				
	Boys and girls, percentage who can completed a literacy or math task, averaged across gender, and grade, Karnataka district	0.0205[62]	NA	NA	India

[62] Not a standard deviation measure of a test score.

Table D3. Governance Estimate of School Based Management on Test Scores

Author	Estimate description	Impact of monitoring on test scores (SD)	Standard error	Observations	Country
Glewwe and Maiga (2011) *	District level intervention, change between Year 2 and Year 1 score, Student and teacher controls with interactions.	0.053	0.116	19,281	Madagascar
	Sub-District level intervention, change between Year 2 and Year 1 score, Student and teacher controls with interactions.	0.079	0.104	19,281	Madagascar
	School level intervention, change between Year 2 and Year 1 score, Student and teacher controls with interactions.	0.073	0.119	19,281	Madagascar
Lassibille et al (2010)*[63]	Total score, direct to teachers and school directors	1.5	2.5	72,720	Madagascar

[63] Lassibille et al. (2010) estimates are in terms of raw percentage of correct responses, not a standard deviation.

	Total score, via subdistrict and district administrators	-0.4	2.4	72,720	Madagascar
	Total score, via district administrators	-0.6	2.5	72,720	Madagascar
Duflo, Dupas, and Kremer (2012, 2015)*	SBM program improve the effective of Extra Teacher Program (ETP)	Positive, significant	NA	6533	Kenya
Beasley and Huillery (2014)*	Normalized Oral Test Scores	-0.0984	0.0759	499	Niger
	Normalized Math Test Scores	-0.0477	0.0584	763	Niger
	End of Primary Test Pass Rate	-0.0244	0.0227	557	Niger
Pradhan et al (2014)*	Language test score, impact of grant	0.129	0.094	11463	Indonesia
	Language test score, impact of election, IV	0.094	0.127	11463	Indonesia
	Language test score, impact of linkage	0.165**	0.067	11463	Indonesia
	Language test score, impact of training	-0.049	0.069	11463	Indonesia

	Math test score, impact of grant	-0.018	0.081	11463	Indonesia
	Math test score, impact of election, IV	-0.009	0.095	11463	Indonesia
	Math test score, impact of linkage	0.069	0.051	11463	Indonesia
	Math test score, impact of training	-0.029	0.051	11463	Indonesia
Santibanez, Abreu-Lastra and O'Donoghue (2014)	Boys and girls, grade 3&6, whole sample, math	0.00259	0.129	18210	Mexico
	Boys and girls, grade 3&6, whole sample, Spanish	-0.0677	0.124	18210	Mexico
Yamauchi (2014)	Change in total test score	0.321816	0.088899	2,406	Philippines

* RCT study

Table D4. Governance Estimate of Teacher Performance Pay on Test Scores

Author	Estimate description	Impact of teacher incentive on test scores (SD)	Standard error	Observations	Country
Muralidharan and Sundararaman. (2011)*	Children in Individual Incentive school, Year 1 to Year 0, Combined Math and Language	0.156	0.040	42,145	India
	Children in Individual Incentive school, Year 2 to Year 0, Combined Math and Language	0.283	0.058	29,760	India
	Children in Group Incentive school, Year 1 to Year 0, Combined Math and Language	0.141	0.050	42,145	India
	Children in Group Incentive school, Year 2 to Year 0, Combined Math and Language	0.154	0.057	29,760	India
Glewwe, Ilias and Kremer (2010)*	School-level Incentives, NGO exam, Year 1 – Year 0	0.046	0.041	39,900	Kenya
	School-level Incentives, NGO Exam, Year 2 – Year 0	-0.017	0.064	18,736	Kenya

Muralidharan (2011)*	Individual Incentive, Combined Score, All 5 Years	0.444	0.101	3456	India
	Group Incentive, Combined Score, All 5 Years	0.129	0.085	3456	India
Contreras and Rau (2012)	Teacher incentive effects on LANGUAGE scores	.046	.055	29,271	Chile
	Teacher incentive effect on MATH scores	.286	.064	29,315	Chile

* RCT study

Table D5. Governance Intervention Estimate of Contract Teachers on Test Scores

Author	Estimate description	Impact of contract teacher on test scores (SD)	Standard error	Observations	Country
Duflo, Dupas and Kremer (2012, 2015)*	Child in ETP School, Total Score, PTA (contract) Teacher	0.294	0.105	6,533	Kenya
Muralidharan and Sundararaman. (2013)*	Combined Score, After 1 Year, School and HH controls	0.094	0.035	12531	India
	Combined Score, After 2 Year, School and HH controls	0.152	0.048	11329	India

* RCT study

Table D6. Governance Intervention Estimate of Private School Voucher on Test Scores

Author	Estimate description	Impact of private school voucher on test scores (SD)	Standard Error	Observations	Country
Angrist et al. *(2002)	pooled test scores (math, reading, writing), 1995 applicant cohort in three Bogota neighborhoods[64]	0.148	0.088	846	Colombia
Angrist, Bettinger & Kremer*(2006)	college-entrance-exam score, language	0.438	0.0979	3541	Colombia
	college-entrance-exam score, math	0.430	0.0978	3541	Colombia
Muralidharan and Sundararaman* (2013, 2015)	Combined score ,2nd year assessment	0.029	0.108	13,765	India
	Combined score, 4th year assessment	0.233	0.078	18,926	India
Lara, Mizala and	Language test score	0.0238	0.0352	27,218	Chile
Repetto (2011)	Math test score	0.0239	0.0427	27,303	Chile

[64] There are three estimates reported by the authors. I only pick the one with the largest sample size to report in this table. The paper summary on page 20 shows all three estimates.

Hsieh and Urquiola (2006)	Language test scores	-0.515	1.318	84	Chile
	Math test scores	-1.495	1.447	84	Chile

* RCT study

Table D7. Summary of Impacts on Time in School of Governance Interventions

Author	Estimate description	Impact of private school voucher on test scores (SD)	Standard Error	Observations	Country
Angrist et al. *(2002)	pooled test scores (math, reading, writing), 1995 applicant cohort in three Bogota neighborhoods[65]	0.148	0.088	846	Colombia
Angrist, Bettinger & Kremer*(2006)	college-entrance-exam score, language	0.438	0.0979	3541	Colombia

[65] There are three estimates reported by the authors. I only pick the one with the largest sample size to report in this table. The paper summary on page 20 shows all three estimates.

	college-entrance-exam score, math	0.430	0.0978	3541	Colombia
Muralidharan and Sundararaman* (2013, 2015)	Combined score ,2nd year assessment	0.029	0.108	13,765	India
	Combined score, 4th year assessment	0.233	0.078	18,926	India
Lara, Mizala and	Language test score	0.0238	0.0352	27,218	Chile
Repetto (2011)	Math test score	0.0239	0.0427	27,303	Chile
Hsieh and Urquiola (2006)	Language test scores	-0.515	1.318	84	Chile
	Math test scores	-1.495	1.447	84	Chile

* RCT study

Table D7. Summary of Impacts on Time in School of Governance Interventions

	Negative, Significant	Negative, Insignificant	Positive, Insignificant	Positive, Significant	Total Studies
Monitoring (all RCTs)	0 (0)	1 (1)	1 (1)	0 (0)	2
School-based management					
RCTs	0 (0)	7 (3)	5 (3)	1 (1)	3
Other high quality studies	0 (0)	2 (1)	1 (1)	0 (0)	2
Private school (vouchers/subsidies)					
RCTs	0 (0)	1 (1)	1 (1)	1 (1)	2
Other high quality studies	0 (0)	1 (1)	0 (0)	1 (1)	2
Teacher performance pay (All RCTs)	0 (0)	1 (1)	2 (1)	0 (0)	1
Single sex school (High Quality)	0 (0)	0 (0)	0 (0)	1 (1)	1

Note: Figures are number of estimates; figures in parentheses are number of papers/studies.

Table D8. Governance Intervention Estimates of Monitoring on Time in School

Author	Estimate description	Impact of monitoring on measure of time in school	Standard error	Observations	Country
Duflo, Hanna and Ryan (2012)*	Difference in attendance for children who did not leave	.04	.03	23,693	India
Banerjee et al (2010)*	Boys and girls, days present in last 14 days	-0.599	0.351	5555	India

Table D9. Governance Intervention Estimate of School Based Management on Time in School

Author	Estimate description	Impact of school based management on measure of time in school	Standard error	Observations	Country
Lassibille et al (2010)*[66]	Attendance rate, direct to teachers and school directors	4.1	1.9	72,720	Madagascar
	Attendance rate, via subdistrict and district administrators	1.5	2.3	72,720	Madagascar
	Attendance rate, via district administrators	3.0	2.1	72,720	Madagascar
	Dropout rate, direct to teachers and school directors	-0.6[67]	3.9	72,720	Madagascar
	Dropout rate, via subdistrict and district administrators	-0.6	1.5	72,720	Madagascar
	Dropout rate, via district administrators	-1.8	1.5	72,720	Madagascar
Pradhan et al (2014)*	Dropout, impact of grant	-0.005	0.005	517	Indonesia

[66] Lassibille et al. (2010) estimates are in terms of percentage points rather than standard deviation.
[67] Reversal of sign, similar for the next two estimates

	Dropout, impact of election, IV	-0.005	0.011	517	Indonesia
	Dropout, impact of linkage	-0.002	0.006	517	Indonesia
	Dropout, impact of training	0.007	0.006	517	Indonesia
Beasley and Huillery (2014)*	Enrollment (2008/2009 school year)	1.366	2.445	988	Niger
	Dropout (2007/2008 school year)	-0.00559	0.00520	748	Niger
	Number of Pupils Attempting End of Primary Test	1.67	1.107	557	Niger
Santibanez, Abreu-Lastra and O'Donoghue (2014)	Dropout, Boys and girls, grade 3, whole samples	-0.02361	0.102841	18269	Mexico
	Dropout, Boys and girls, grade 6, whole samples	-0.04018	0.111144	18151	Mexico
Gertler et al. (2012)	boys and girls, intra-year dropout rate	0.001	0.002	125700	Mexico

* RCT study

Table D10. Governance Intervention Estimate of Private School Voucher on Time in School

Author	Estimate description	Impact of private school vouchers on measure of time in school	Standard Error	Observations	Country
Angrist et al. *(2002)	Enrollment rate, 7th and 8th graders	-0.002	0.016	1577	Colombia
	Years in school, 7th and 8th graders	0.015	0.044	1577	Colombia
Angrist, Bettinger and Kremer*(2006)	high school graduation	0.056*	0.014	3542	Colombia
Barrera-Osorio & Raju (2011)	Number of students, bandwidth 3% pts	122.6	45.75	319	Pakistan
	Number of students, bandwidth 4.5% pts	85.20	42.77	319	Pakistan
	Number of students, bandwidth 6% pts	88.84	41.32	319	Pakistan
Hsieh and Urquiola (2006)	Total Years of Schooling	-1.68	1.36	85	Chile

(* RCT study)

Table D11. Governance Intervention Estimate of Unique Interventions on Time in School

Author	Estimate description	Impact of teacher incentive on measure of time in school	Standard Error	Observations	Country
Glewwe, Ilias and Kremer (2010)*	Impact of teacher performance pay on dropout rate after year one	-0.008[68]	0.012	13,347	Kenya
	Impact of teacher performance pay on dropout rate after year two	-0.008	0.011	12,007	Kenya
	Impact of teacher performance pay on dropout rate after year three	0.002	0.009	9,479	Kenya
Jackson (2012)	Impact of single-sex school on student taken national examinations (indicator of not dropping out)	0.073**	0.018	179710	Trinidad and Tobago

* RCT study

[68] Reversal of signs for dropout rate

Appendix 2

Description of Grey Literature Search

Young Lives is a research organization focusing on childhood poverty in developing countries and is coordinated by a team based at the University of Oxford led by Professor Jo Boyden. By searching its list of working papers, we found seven studies that are related to education and are potentially relevant to our literature review. Upon further review, two studies have already been included in our list of studies, while the rest five studies are not qualified due to their lack of emphasis on educational outcomes or their weak estimation strategy. Therefore, there are no additional high quality studies to add from this source.

Mathematica Policy Research conducts analysis to assess the effectiveness of policies on a variety of topics around the globe. Of the 67 international studies posted on the website for the Mathematica Center for International Policy Research and Evaluation, five studies are on education program evaluations in developing countries and all of them are high quality studies. Of these five studies, two of them were already included in our list of literature, and one of them was found in a 3IE systematic review. Therefore, two additional high quality studies were added from this source (Levy et al. (2009) and Dumitrescu et al. (2011)).

Save the Children is an international non-governmental organization that promotes children's rights and their opportunity to learn in developing countries. We searched the Save the Children documents database for quantitative studies related to education interventions using the keywords "quantitative analysis and education" and found 66 documents. After eliminating meta-analyses, papers without educational outcomes, and other irrelevant topics, we were left with five possible papers to review further. Four of the five documents reviewed a multi-country program called a "Literacy Boost" program. Countries implemented a baseline and post-intervention survey in the recipient areas. While these studies were quantitative, only one used a "high-quality" empirical approach, namely the analysis of the Literacy Boost program in Ethiopia. Authors in this study found that the program had a weak effect on letter identification and reading comprehension but a strong and relatively large effect on reading accuracy. The last of the five relevant papers examines using tablets and satellite internet to improve

numeracy skills in several developing countries. This study did not use any of the empirical methods included in our high-quality classification. Therefore, we did not find additional high quality study to add from this source.

The Millennium Challenge Corporation (MCC) was created by the U.S. Congress in January 2004 and it delivers U.S foreign assistance, focusing on good policies, country ownership and results. As part of the MCC's result's framework, independent impact evaluations are conducted. Currently the list includes 51 impact evaluations, of which only eight have been completed and 43 are planned. Of the eight completed impact evaluations, only two are education related. One program in Burkina Faso was designed to improve girl's primary education through the BRIGHT School Construction Program. The results of this impact evaluation by Kazianga et al. (2013) had already been included in our presentation of demand side interventions from our sweep of the academic literature. A second program in Niger launched in March 2008, also sought to improve girls' primary school education. This program was also listed as having a completed evaluation; however, in December 2009 the MCC Board of Directors suspended the Niger program due to political events that were inconsistent with the criteria used to determine a country's eligibility. In March 2012, the MCC Board of Directors authorized a new grant, but no details are provided. No impact evaluations for the Niger program are available through MCC, however, we believe program was evaluated by Mathematica Policy Research by Dumitrescu et al. (2011) as mentioned above. Therefore, no additional high quality study was added from this source.

USAID studies were search through both the "Development Experience Clearinghouse" and the "USAID Development Data Library". The Development Experience Clearinghouse contains mainly qualitative reports, and covers over 50 years of program descriptions. Using the same keywords used to search EconLit and ERIC, no report was found that conducts quantitative analysis to identify the treatment effect of an educational program. The papers that fit our quantitative criteria have already been included in our list of literature. Most papers include basic summary statistics or program descriptions, but lack any quantitative investigation of the program. The second branch of the search covered the Development Data Library, which is a public repository of data, and therefore contains no additional studies.

Swedish International Development Cooperation Agency (Sida) was searched using the keyword "education" and the search was

limited to publications only. Of the 75 search results we did not find any quantitative studies that use high quality empirical methods to evaluate the impacts of educational programs.

Finally, we also examined the websites of other aid agencies including DFID (Department for International Development), Norway Agency for Development Cooperation, Oxfam, UNICEF, and the World Bank but did not find any high quality quantitative program evaluations.

Appendix 3

Key Words Used in Search

KW=education AND KW=("class size" OR "school size" OR "Student teacher ratio" OR "Pupil teacher ratio" OR "School expenditure*" OR “expenditure per pupil” OR "texbook*" OR "instructional material*" OR "Workbook*" OR "exercise book*" OR "computer*" OR "laptop*" OR "internet" OR "school infrastructure" OR "Facilities" OR "Building condition*" OR "Laborator*" OR "lab" OR "labs" OR "Librar*" OR "Desk*" OR "Teaching tools" OR "teaching guide*" OR "blackboard*" OR "chalk*" OR "electricity" OR "table*" OR "bench*" OR "chair*" OR "roof*" OR "wall*" OR "floor*" OR "window*" OR "bathroom*" OR "plumbing" OR "teacher quality" OR "teacher efficacy" OR "teacher knowledge" OR "teacher salar*" OR "teacher training" OR "teacher experience" OR "teacher education" OR "teacher absenteeism" OR "teacher gender" OR "class preparation" OR "lesson planning" OR "homework" OR "evaluation" OR "follow-up" OR "monitoring of pupil performance" OR "testing" OR "remedial program*" OR "teaching practices" OR "instructional time" OR "length of instructional program" OR "hours" OR "school day" OR "curriculum" OR "principal quality" OR "principal training" OR "principal education" OR "principal experience" OR "staff assessment*" OR "teacher assessment" OR "school inspection*" OR "parent* involvement" OR "production function" OR "school resources" OR "school inputs" OR "School quality" OR "Pedagogical inputs" OR "pedagogical resources" OR “grant*” OR “fee*” OR “multi grade teaching” OR “grouping” OR “tracking” OR “streaming” OR “scholarship*” OR “rewards” OR “teaching at the right level” OR “double shift” OR “drinking water” OR “toilet*” OR “school uniform” OR “teacher sex” OR “teacher ethnicity” OR “teacher race” OR “teacher salar*” OR “teacher incentive*” OR “teaching assistant*” OR “volunteer” OR “scripted lessons” OR “student led learning” OR “group work” OR “remedial instruction*” OR “teacher contract” OR “tenure” OR “extend contract” OR “incentive* for student performance” OR “promotion” OR “tutor*” OR “principal gender” OR “principal sex” OR “classroom observation” OR “distance to school” OR “conditional cash transfers” OR “unconditional cash transfers” OR “take home food ration*” OR “school meal” OR “deworming” OR “iron” OR “iodine” OR “zinc” OR “multi vitamin” OR “malaria” OR “eyeglasses” OR “decentralization” OR “school voucher” OR “public school*” OR

“private school*” OR “school choice” OR “information campaign” OR “school performance” OR "facilit*" OR "evaluation*" OR "hour*" OR "teacher assessment*" OR "school resource*" OR "school input*" OR "pedagogical input*" OR "pedagogical resource*" OR “reward*” OR “double shift*” OR “volunteer*” OR “scripted lesson*” OR “teacher contract*” OR “extend contract*” OR “classroom observation*” OR “distance to school” OR “conditional cash transfer*” OR “unconditional cash transfer*” OR “school meal*” OR “school voucher*” OR “school choice*”)

AND AB=("developing countr*" OR "Least-Developed Countries" OR "Afghanistan" OR "Albania" OR "Algeria" OR "Angola" OR "Antigua and Barbuda" OR "Argentina" OR "Armenia" OR "Azerbaijan" OR "Bahamas" OR "Bahrain" OR "Bangladesh" OR "Barbados" OR "Belarus" OR "Belize" OR "Benin" OR "Bhutan" OR "Bolivia" OR "Bosnia and Herzegovina" OR "Botswana" OR "Brazil" OR "Brunei Darussalam" OR "Bulgaria" OR "Burkina Faso" OR "Burundi" OR "Cambodia" OR "Cameroon" OR "Cape Verde" OR "Central African Republic" OR "Chad" OR "Chile" OR "China" OR "Colombia" OR "Comoros" OR "Congo" OR "Costa Rica" OR "Côte d'Ivoire" OR "Croatia" OR "Djibouti" OR "Dominica" OR "Dominican Republic" OR "Ecuador*" OR "Egypt*" OR "El Salvador" OR “Salvadoran” OR "Equatorial Guinea" OR "Eritrea" OR "Ethiopia*" OR "Fiji*" OR "Gabon*" OR "Gambia*" OR "Georgia*" OR "Ghana*" OR "Grenada*" OR "Guatemala*" OR "Guinea" OR "Guinea-Bissau" OR "Guyana" OR "Haiti" OR "Honduras" OR "India" OR "Indonesia" OR "Iran" OR "Iraq" OR "Jamaica" OR "Jordan" OR "Kazakhstan" OR "Kenya" OR "Kiribati" OR "Kosovo" OR "Kuwait" OR "Kyrgyz Republic" OR "Lao People's Democratic Republic" OR "Latvia" OR "Lebanon" OR "Lesotho" OR "Liberia" OR "Libya" OR "Lithuania" OR "Macedonia" OR "Madagascar" OR "Malawi" OR "Malaysia" OR "Maldives" OR "Mali" OR "Mauritania" OR "Mauritius" OR "Mexico" OR "Moldova" OR "Mongolia" OR "Montenegro" OR "Morocco" OR "Mozambique" OR "Myanmar" OR "Namibia" OR "Nepal" OR "Nicaragua" OR "Niger" OR "Nigeria" OR "Yugoslav" OR "Oman" OR "Pakistan" OR "Panama" OR "Papua New Guinea" OR "Paraguay" OR "Peru" OR "Philippines" OR "Poland" OR "Qatar" OR "Romania" OR "Russia" OR "Rwanda" OR "Samoa" OR "São Tomé and Príncipe" OR "Saudi Arabia" OR "Senegal" OR "Serbia" OR "Seychelles" OR "Sierra Leone" OR "Solomon Islands" OR "South Africa" OR "Sri Lanka" OR "St. Kitts and Nevis" OR "St. Lucia" OR

"St. Vincent and the Grenadines" OR "Sudan" OR "Suriname" OR "Swaziland" OR "Syrian Arab Republic" OR "Tajikistan" OR "Tanzania" OR "Thailand" OR "Timor-Leste" OR "Togo" OR "Tonga" OR "Trinidad and Tobago" OR "Tunisia" OR "Turkey" OR "Turkmenistan" OR "Uganda" OR "Ukraine" OR "United Arab Emirates" OR "Uruguay" OR "Uzbekistan" OR "Vanuatu" OR "Venezuela" OR "Vietnam" OR "Yemen" OR "Zambia" OR "Zimbabwe" OR "North Korea" OR "Cuba" OR "Syria")

NOT AB=("U.S." OR "US" OR "District of Columbia" OR "D.C." OR "DC" OR "Washington DC" OR "Washington D.C." OR "Alabama" OR "Alaska" OR "Arizona" OR "Arkansas" OR "California" OR "Colorado" OR "Connecticut" OR "Delaware" OR "Florida" OR "Georgia" OR "Hawaii" OR "Idaho" OR "Illinois" OR "Indiana" OR "Iowa" OR "Kansas" OR "Kentucky" OR "Louisiana" OR "Maine" OR "Maryland" OR "Massachusetts" OR "Michigan" OR "Minnesota" OR "Mississippi" OR "Missouri" OR "Montana" OR "Nebraska" OR "Nevada" OR "New Hampshire" OR "New Jersey" OR "New Mexico" OR "New York" OR "North Carolina" OR "North Dakota" OR "Ohio" OR "Oklahoma" OR "Oregon" OR "Pennsylvania" OR "Rhode Island" OR "South Carolina" OR "South Dakota" OR "Tennessee" OR "Texas" OR "Utah" OR "Vermont" OR "Virginia" OR "Washington" OR "West Virginia" OR "Wisconsin" OR "Wyoming" OR "UK" OR "U.K." OR "United Kingdom" OR "Great Britain" OR "Britain" OR "England" OR "Scotland" OR "Wales" OR "Northern Ireland" OR "Europe" OR "Cyprus" OR "Malta" OR "Australia" OR "Austria" OR "Belgium" OR "Canada" OR "Czech Republic" OR "Denmark" OR "Estonia" OR "Finland" OR "France" OR "Germany" OR "Hungary" OR "Iceland" OR "Ireland" OR "Israel" OR "Italy" OR "Japan" OR "South Korea" OR "Korea" OR "Taiwan" OR "Luxembourg" OR "Netherlands" OR "New Zealand" OR "Norway" OR "Portugal" OR "Slovak Republic" OR "Slovenia" OR "Spain" OR "Switzerland" OR "Sweden" OR "Greece")

Previous EBA-reports

2014:01, *Randomized controlled trials: strengths, weaknesses and policy relevance,* Anders Olofsgård.

2014:02, *Sweden´s development assistance for health – policy options to support the global health 2035 goals,* Gavin Yamey, Helen Saxenian, Robert Hecht, Jesper Sundewall and Dean Jamison

2014:03, *International party assistance – what do we know about the effects?* Lars Svåsand

2014:04, *The African Development Bank: ready to face the challenges of a changing Africa?* Christopher Humphrey

2014:05, *Svenskt statligt internationellt bistånd i Sverige: en översikt*, Expertgruppen för biståndsanalys

2015:01, *Rethinking Civil Society and Support for Democracy*, Richard Youngs

2015:02, *Utvärdering av svenskt bistånd – en kartläggning*, Expertgruppen för biståndsanalys

2015:03, *Concentration difficulties? An analysis of Swedish aid proliferation*, Rune Jansen Hagen

2015:04, *Youth, entrepreneurship and development*, Kjetil Bjorvatn

2015:05, *Has Sweden injected realism into public financial management reforms in partner countries?* Matt Andrews

2015:06, *Now open for business: joint development initiatives between the private and public sectors in development cooperation*, Sara Johansson de Silva, Ari Kokko and Hanna Norberg

2015:07, *Making development work: the quality of government approach*, Bo Rothstein and Marcus Tannenberg

2015:08 *Business and Human Rights in development cooperation – has Sweden incorporated the UN guiding principles*? Rasmus Kløcker Larsen and Sandra Atler

2015:09 *In search of double dividens from climate change intervention evidence from forest conservation and household energy transition.* Gunnar Köhlin, Subhrendu K Pattanyak, Erin Sills, Eskil Mattson, Madelene Ostwald, Ariana Salas, Daniel Ternald

2016:01 *Support to regional cooperation and integration in Africa - what works and why?* Fredrik Söderbaum and Therese Brolin