2021

Girls' Education at Scale

Review of Evidence

David K. Evans
Amina Mendez Acosta
Fei Yuan

2021

Research, evaluation and learning are core components of Co-Impact's work. As we embark on a concerted effort to contribute to achieving gender equality at scale in the global south, we need to learn about and build on the existing evidence and knowledge. As part of our broader learning effort, we commissioned a series of rapid reviews of literature by area experts to help us understand major trends as well as new directions about what we know works - and doesn't work - to achieve gender-equitable outcomes at scale in the global south. While these reviews serve as a core component of our evolving thinking, they do not represent official opinions of Co-Impact.

Given that these reviews are focused on critical evidence of initiatives that have been evaluated at scale, we understand there are experiences and knowledge that may not be captured in these documents. We hope to invest in additional reviews in the future to cover other areas of inquiry, and also to build on a wider spectrum of evidence and perspectives.

This important work underpins the development of our own research and learning strategy, in which we will prioritize the questions and needs of practitioners working to achieve gender equitable outcomes, and also to amplify the voices and experiences of women, girls, and other marginalized groups. We hope that this evidence and knowledge, in turn, will contribute to building the global evidence base.

Abstract:

Many educational interventions boost outcomes for girls in settings where girls face education discrimination, but which of those interventions are proven to function effectively at large scale? In contrast to earlier reviews, this review focuses on large-scale programs and policies—those that reach at least 10,000 students—and on final school outcomes such as completion and student learning rather than intermediate school outcomes such as enrollment and attendance. Programs and policies that have boosted access and/or learning at scale across multiple countries include school fee elimination, school meals, making schools more accessible, and improving the quality of pedagogy. Other interventions, such as providing better sanitation facilities or safe spaces for girls, show promising results but either have limited evidence across settings or focus on post-educational outcomes (such as income earning) in their evaluations. We discuss three aspects of considering evidence-based solutions to local problems—constraints to girls' education, potential solutions, and program costs---as well as lessons for scaling programs effectively. If education systems seek to expand and improve girls' education at scale, there are tested tools that have performed effectively in multiple settings, even as education leaders, partners, and researchers collaborate to continue innovating and testing new programs at scale.

Introduction

Gender equality is a stated objective in much of the world: indeed, the fifth Sustainable Development Goal is "achieve gender equality and empower all women and girls" (United Nations 2015). Education is a crucial human capital investment that opens the door to subsequent economic opportunity. As a result, gender equality in education is one crucial step—albeit not the only one—towards achieving gender equality in life outcomes more broadly (see Box 1).

Girls' education is often touted as one of the best investments in international development (Kim 2016). But across low- and middle-income countries, adult women still have less education than men. Among young women and men in their early 20s, girls in South Asia and Sub-Saharan Africa still have less educational opportunity, whereas in other regions, girls have gained more ground (Evans et al. 2021a). These average shifts mask important differences across countries, within regions of countries, and across levels of schooling. In some parts of the world and at some levels, improved gender equality in education means more and better education for girls; in others, it means more and better education for boys.

How to achieve gender equality in education at scale? Evidence on how to expand and improve education—including girls' education—in low- and middle-income countries has expanded dramatically in recent years (Cameron et al. 2016; Sabet and Brown 2018; World Bank 2018). This review examines evidence from large-scale interventions, usually implemented by or in partnership with the government, to improve gender equality in education. It also discusses the quality of the evidence, where and how different solutions may apply differently, and signals where more evidence may be needed.

Box 1: Gender equity versus gender equality in education

Gender equality in education and gender equity in education are different, related concepts. Equality may be associated with achieving similar education outcomes for boys and for girls. Equity, rather, is associated with "fairness or justice in the provision of education" (Espinoza 2007). While the current study focuses on achieving gender equal outcomes in education (e.g., gender parity in school completion or learning), that may require gender inequality in resources spent. One might also argue that, if women face greater challenges than men later in life, gender inequality in education may be needed to achieve gender equality in later life outcomes.

Our results show that programs and policies that have increased school completion or boosted learning for girls at scale in areas where girls face educational disadvantage include, among others, the elimination of fees or providing scholarships or stipends, reducing the distance to school or facilitating travel to school, providing school meals, improving the pedagogy of teachers through a range of inputs, and interventions that help students receive instruction at their level of learning.[2]

---

[2] In this study, we do not distinguish strictly between programs and policies, as policies (such as the elimination of school fees) are virtually always accompanied by a program (such as providing grants to schools to compensate for lost fees).

We also discuss interventions that explicitly address issues faced principally by girls. These include sanitation and menstrual health, gender sensitization training, and safe spaces for girls. However, most of these interventions either have not been implemented at large scale or have not been evaluated with a focus on educational outcomes like improved learning and school completion. However, we provide evidence on the interventions they do shift (e.g., girls' mental health and in some cases, their post-school transitions). Future, large-scale evaluations of such interventions will allow a better understanding of how well such programs can be implemented at scale and whether they shift educational outcomes.

In the discussion section of the paper, we propose a three-step framework—constraints, solutions, and costs—for how policymakers and their partners can apply this evidence across different contexts. We also discuss lessons for scaling interventions effectively.

These findings complement those of other, recent reviews related to girls' education (e.g., Evans and Yuan forthcoming or the girls' education section of Evans and Mendez Acosta 2021), but the focus on at-scale programs gives greater salience to programs that have been implemented by multiple countries nationwide, such as school fee elimination, school construction, or school meals.

The COVID-19 pandemic introduces large challenges for education overall: for example, as of February 2021, schools in South Sudan had been closed for 16 percent of a child's average lifetime schooling careers (Evans et al. 2021b). There are various channels by which the pandemic may be particularly harmful for girls: with a higher burden of housework while schools are closed, greater risks from possible adolescent pregnancies, and discriminatory treatment when resources for education are reduced (Mendez Acosta and Evans 2020). Beyond projections, there is little to no data yet available on gender differences in learning loss or dropout rates. But insofar as there are differential challenges, targeted programs to bring girls back to school and remediate their learning may be valuable.

The challenges

The broad policy challenge

Inequality in educational achievement is a massive global challenge, but the nature of the challenge varies dramatically across settings (Figure 1). For example, in low-income countries (like Afghanistan or Mali), boys are more likely to complete primary, lower secondary, and upper secondary education than girls. The gap grows with each level of education, doubling from a four percentage point gap in primary school (63 percent versus 67 percent) to an eight percentage point gap in lower and upper secondary school.

Figure 1: Gender inequities in education differ at the primary level versus the secondary level, and they vary in low-income versus middle-income countries.
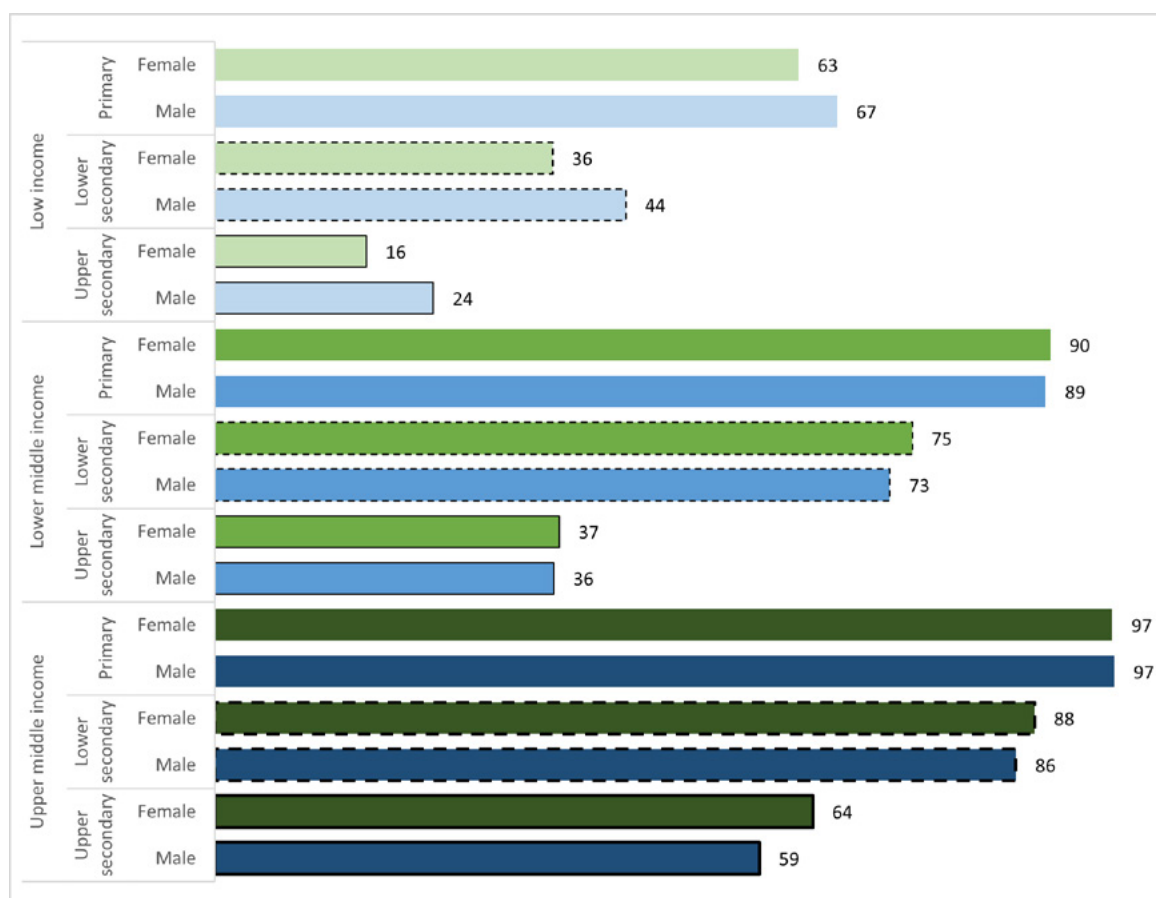
Source: Authors' construction. The school completion rates for this figure are from the World Development Indicators (most recent available year) for primary and lower secondary education, and they are from UNESCO for upper secondary education. The data were downloaded in January 2021.

In lower middle-income countries (like Bolivia or Ghana), girls and boys complete their basic education at essentially the same rates. But in upper middle income countries (like Malaysia or Mexico), while there is parity in primary school completion (at 97 percent – almost every child is completing primary school), girls are five percentage points more likely to complete upper secondary school than boys. As countries grow more prosperous, gender gaps favoring boys disappear and girls even begin to pull ahead.

There are exceptions at every level of income. In Madagascar and Burkina Faso, both low-income countries, girls are more likely to complete primary and lower secondary school than boys. Angola, Benin, and Pakistan are lower-middle income countries with a remaining sizeable gender gap favoring boys at the primary level. In some upper-middle income countries such as Bulgaria, Gambia, and Guatemala, boys are at least three percentage points more likely to finish upper secondary schooling than girls.

Beyond access, there are also differences in learning outcomes. The World Bank's harmonized test scores show that girls tend to have lower test scores than boys in low-income countries, with a concentration in Sub-Saharan Africa. In most middle-income countries, including most countries in Latin America, girls outperform boys on exams (World Bank 2020b, 19–20). Likewise, beyond differences in the gender gap across

national levels of income, there will be differences in the gender gap across income levels within countries, between rural and urban areas (Evans 2019), and across other parameters. This array of parameters across which gender inequality can linger explains why a middle-income country, despite having achieved gender parity on average, will still need to worry about gender equality in education.

At the same time, the specific challenges that girls (and boys) face change over time. For example, in a study of improved sanitation in Indian schools, Adukia (2017) found that the gains from sex-specific latrines principally appeared once girls had already reached puberty; for younger girls, the presence of a latrine helped, but whether or not it was sex-specific mattered less. Concern about sexual violence may be concentrated (albeit not limited to) older students. So just as countries will differ in whether they need to principally focus on closing gaps favoring boys or girls, the steps will vary based on the level of schooling.

This all reminds us that the challenge of gender equitable education is not a single challenge. Countries vary enormously in whether boys or girls are ahead in education and, therefore, need special attention and resources. Furthermore, even in countries where outcomes are similar, there may be important differences that require distinct attention. While this study explores – principally – interventions that have sought to improve girls' education at large scale, no one intervention will apply in all cases.

The knowledge challenge

The growth in evidence from evaluations of interventions on how to improve education has been dramatic in recent years, with a 15-fold increase in studies between 2000 and 2016 (Figure 2). Many of these evaluations either focus on girls education or present evidence on girls education within the context of a program that benefits both boys and girls (Evans and Yuan forthcoming). Beyond methodological differences (e.g., experimental versus quasi-experimental evaluations), this evidence includes evaluations of various types of programs. Policymakers and partners can learn different things from each type of program evaluation.

Figure 2: The growth in cumulative experimental and quasi-experimental education impact evaluations over time

Legend: ■ Cumulative studies ■ Studies in given year
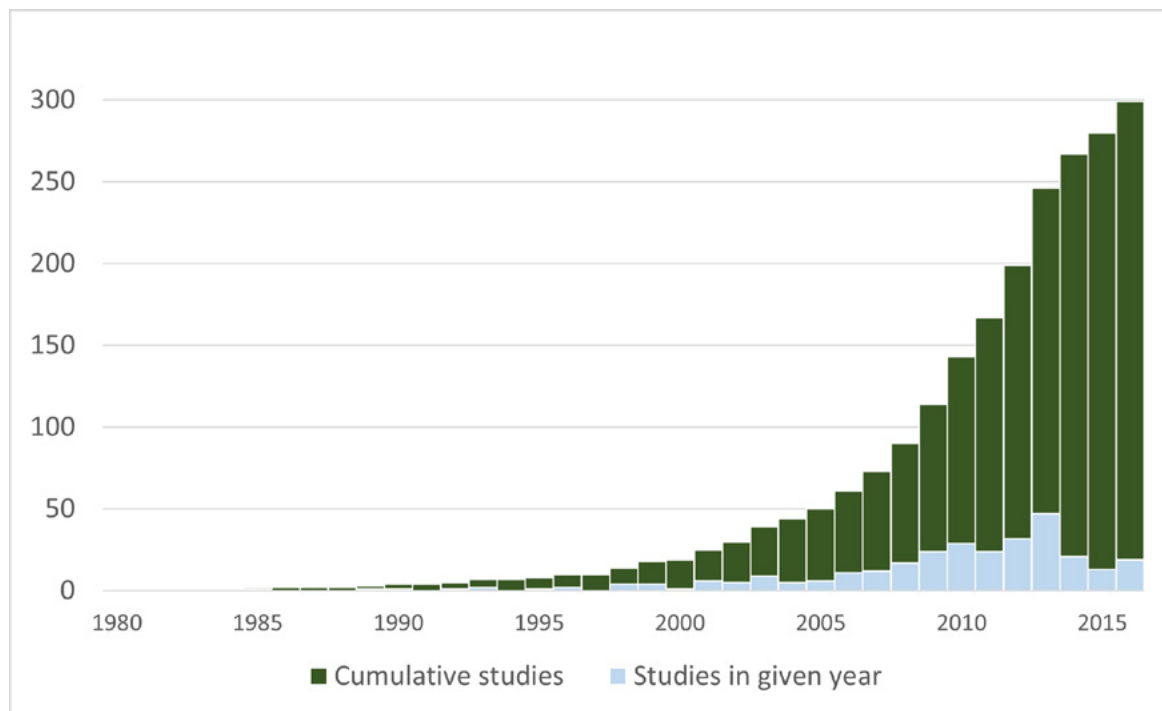
Source: Authors' construction. Adapted from Figure S4.1 from World Bank (2018).

One way to categorize the programs evaluated is based on implementer and scale, yielding four types: (1) pilot interventions implemented by non-government actors, (2) pilot interventions implemented by government actors, (3) large-scale interventions implemented by non-government actors, and (4) large-scale interventions implemented by government actors (Figure 3). For the sake of this paper, interventions that reach at least 10,000 students or were implemented nationwide will be considered "large-scale." (In some countries, even nationwide implementation may not reach 10,000 students.) Obviously that number is ultimately an arbitrary one, but it does allow a distinction between truly small programs and those that require more management infrastructure and resources to implement.

An example of the first type of study—pilot interventions implemented by non-government actors—might be the enrollment of about 2,000 children in alternative schools established by non-government organization (NGO) in Guinea-Bissau (Fazzio et al. 2020). That intervention led to a six-fold increase in girls' test scores. In that context, government provision of education is very limited, and this study demonstrates that a holistic intervention—providing non-government schools with custom-designed teacher training materials and classroom materials for students and teachers, together with monitoring and evaluation of teachers and community outreach—can deliver dramatic gains in learning in an area with historically low learning levels. Another example of such a program evaluation examines the impact of setting up parent-teacher conferences in a study of about 4,000 students in Bangladesh (Islam 2019). The program was implemented by a local NGO, and it more than doubled girls' test scores by the end of two years. Both studies provide promising interventions to increase girls' learning, but neither tells us whether it would be possible to implement such a program at scale. Evaluations of this

style of intervention can be designed to inform scale up, but how well the program will actually work with thousands more students remains a challenge to know with certainty (Banerjee et al. 2017).

Figure 3: What policy makers and partners learn from different types of program evaluations

		Scale	
		Small	**Large**
Implementation mechanism	**Non-government**	Does the theory behind the intervention apply in real life and in this setting?	Is it possible to reach large numbers of children and youth with the program?
	Government	Can existing structures implement this program under careful supervision?	Is it possible to reach large numbers of children and youth through existing structures?

Source: Authors' construction

The second type of study implements an intervention at relatively small scale but using government systems. For example, the Government of South Africa implemented a randomized controlled trial (RCT) in 180 public primary schools, comparing the provision of traditional teacher professional development with the more interactive, on-site coaching. The coaching boosted girls' test scores four times as much as the traditional training (Cilliers et al. 2019).[3] Because this was implemented through government channels, we can be more confident that it is possible to implement using teachers who have been recruited and are remunerated and managed through the government system. This points to a greater confidence that the intervention may scale (Gove et al. 2017).

However, even within government systems, implementing a program on a large scale still poses new challenges (Anderson 2018). First, the quality of implementers may suffer. Another government-implemented pilot impact evaluation in South Africa provided teachers with virtual coaching, i.e., a coach who reached out to teachers via tablet. Because the coach did not travel to schools, one coach was able to provide virtual support to many schools (Cilliers et al. 2020a). But if such a program were scaled up nationally, more coaches would be needed. Would it be possible to find many coaches of similar quality of the pilot coach? In some places, maybe not. Second, the quality of supervision may suffer. Providing careful supervision to a pilot with a dozen or even a hundred schools is a different endeavor than providing careful supervision to thousands of schools. Third, in pilot programs it may be relatively easy to adjust the program as new challenges arise. Large-scale programs lose some of that quick flexibility. Fourth, programs often change models at scale to reduce costs or because of political pressures.

---

[3] The separated boy-girl impacts are not reported in the study, but they were communicated to us by the authors. For girls, the impact of coaching on literacy was 0.15 standard deviations (p=0.10) and the impact of traditional training was 0.04 SDs.

A literacy program in Uganda was highly effective at boosting student learning when implemented by an NGO but was completely ineffective when a reduced cost version was implemented by civil servants (Kerwin and Thornton forthcoming). A program putting teachers on short-term contracts led to learning gains when implemented by an NGO, but its design was compromised when implemented at scale due to political pressures (Duflo et al. 2015; Bold et al. 2018).

Because of these and other challenges, large-scale programs often report smaller impacts than pilot programs, as has been demonstrated both in teacher coaching programs in high-income countries (Kraft and Blazar 2018) and across a variety of education programs in low- and middle-income countries (Evans and Yuan 2020).[4] Thus, rigorous evaluation evidence of programs at scale is uniquely valuable.

Studies of the third or fourth type – implemented at scale, either in government systems or through NGOs that have the capacity to work at large scale – provide the most direct evidence of effective, at-scale interventions. Most of these evaluations are quasi-experimental. For example, an evaluation of a government program providing funds for bicycles to 160,000 secondary school girls in India (Bihar state) compares the girls who received bicycles to girls in a neighboring state and to local boys, who were not eligible for the program, a method called "triple differences" (Muralidharan and Prakash 2017). The aforementioned study of latrine construction in Indian schools compares changes in outcomes among students attending schools where latrines were built through a large-scale government program (Adukia 2017). These evaluations show that it is possible to implement the program at scale and still achieve significant impact.

While it is possible to learn from all these classes of evaluations, the focus of this study will be on studies that fall into the latter two categories, especially those that are implemented by government at scale.[5] Most education is provided by the public sector: across low- and middle-income countries, less than 20 percent of primary education and less than 30 percent of secondary education is provided privately (World Bank 2020c).[6] As a result, interventions at large scale may be most sustainable if implemented through public sector mechanisms. NGOs that achieve results at large scale are also a key source of information for both policy makers and donors.

The methods used for this review

This is a narrative, rapid review of evidence on large-scale efforts to improve girls' education at scale. The research team examined a long list of evaluations and identified those that reports impacts for girls and were implemented at large scale (at least 10,000 beneficiaries or nationwide implementation). The evaluations were drawn from various sources, including previous reviews of evidence on girls' education that included both small and large-scale studies (Evans and Yuan forthcoming; Sperling et al. 2016; Awasom et al. 2020), the Millions Learning initiative (Robinson et al. 2019), the J-PAL

---

[4] DellaVigna and Linos (2021) highlight that this pattern may be driven by "publication bias," in which statistically significant results are more likely to be submitted and accepted for publication in academic journals (see section 4.3.1 of their study). Studies with smaller samples have less statistical power to estimate an effect, so smaller studies will on average require larger effects to show statistical significance.

[5] Previous reviews have examined the full array of evidence. See Evans and Yuan (forthcoming) for one example.

[6] The numbers are 19 percent for primary and 28 percent for secondary. This is the low- and middle-income country aggregate provided for 2019, the most recent year for which data are available. This represents an increase from ten years previously, when the numbers were 15 percent and 23 percent for primary and secondary.

post-primary education initiative, the "Learning @ Scale" initiative (Stern et al. 2020), and reviews on individual education topics (e.g., Read and Atinc 2017). The team included both evaluations that target girls specifically and evaluations that target both boys and girls but that report impacts separately for girls (see Box 2), or studies for which an earlier review identified gender-separated effects that were not reported in the original studies (Evans and Yuan forthcoming).

Box 2: Gender targeted versus nontargeted interventions

Interventions to improve school completion and learning for girls may include either interventions that are targeted specifically to girls (e.g., a girls' scholarship) or general interventions. Previous research demonstrates that in some circumstances, general interventions deliver comparable gains to targeted interventions (Evans and Yuan forthcoming). The choice of targeted or general interventions will depend on the principal obstacles girls face and whether those obstacles are principally felt by girls or not.

The principal outcomes of interest were those nearer the end of the education production process: learning and school completion. Studies with other outcomes, such as enrollment or attendance, were included principally insofar as they were instrumental to learning or to school completion, or in cases where other studies of similar programs established impacts on learning or completion. Subsequent life outcomes—e.g., income, fertility, or employment—were included in the relatively rare cases that they were available. We reviewed this collection of studies and synthesized and summarized the findings. While this review is focused around improving gender equality in education, we include discussion of studies that impressively boost educational outcomes for girls in areas where girls remain disadvantaged in school—particularly Sub-Saharan Africa and South Asia (Evans et al. 2021a)—even in cases where both girls' and boys' education increase. In those cases, similar absolute gains may reduce inequality if girls begin at a lower level: i.e., an increase in primary school completion of ten percentage points represents a higher percentage increase for girls if they begin at a lower level of completion. Even in cases where the percentage gains are similar, sizeable and significant gains to the quality and accessibility of girls' education is likely to benefit girls regardless of impacts on boys.

Our discussion of the effects is largely qualitative. In this rapid review, we do not standardize effect sizes across interventions. Standard deviations in outcomes, while useful, can vary widely due to factors separate from the impact of the intervention such as the underlying distribution of initial value of the outcome in the study population, as well as – in the case of measuring learning – differences in the difficulty of changing the specific skills measured or different test designs (Evans and Yuan 2020).

The solutions

Make school cheaper

Fee elimination and scholarships

Many studies demonstrate that eliminating school fees or providing scholarships can dramatically increase school completion as well as learning outcomes. This applies in both primary and secondary school. While most countries have officially eliminated school fees at the primary level and some have eliminated school fees at the secondary levels, in practice, families are often required to pay some sort of contributions to the school, above and beyond the cost of school materials and transportation and the opportunity cost of sending children and youth to school (Williams et al. 2015).

At the secondary school level, providing scholarships for youth in Ghana who had passed the secondary school entrance exam but who did not have the resources to pay—so keep in mind that this is a select group of students—led to more than a 60 percent increase in senior secondary school completion for girls (an increase of 25 percentage points relative to 40 percentage points in the comparison group). Girls were also much more likely to be enrolled in tertiary, although few were enrolled to begin with: that increased from 8 percent to 12 percent. The scholarships also led to a range of other positive impacts: better test scores in both reading and mathematics, better national political knowledge, media engagement, and a higher likelihood of having a bank account. Girls even had fewer pregnancies (Duflo et al. 2019). Eliminating school fees for secondary school girls in the Gambia increased the number of girls taking the high school exit exam (one proxy for completion) by 55 percent (Blimpo et al. 2019). Likewise, providing vouchers to cover the cost of private secondary school in Colombia – a program that reached 125,000 children – increased both test scores and secondary school completion rates, at comparable rates for both girls and boys (Angrist et al. 2006). Another scholarship program – for more than 100,000 girls in upper primary and lower secondary grades in the Democratic Republic of the Congo – boosted both reading and mathematics scores (Randall and Garcia 2020).

Another program that paid school fees – this time in Tanzania – also covered other, informal costs for tens of thousands of secondary school girls who had been identified by their communities as highly vulnerable. The program also provided other benefits, such as textbooks and life skills training. Beneficiary girls had much higher test scores than girls at comparison schools. Less poor girls who attended beneficiary schools but did not receive financial support also had higher test scores, as did boys, suggesting a positive spillover effect. Girls who had their fees paid were 25 percent less likely to drop out of high school (Sabates et al. 2020). The evaluation of this program matched beneficiary girls with girls at other schools based on observable characteristics, so we are slightly less confident of the causal impact claim—i.e., it is possible that students in beneficiary schools were different than their comparators in ways we do not observe but that affect educational outcomes. Still, the results are consistent with evidence elsewhere that eliminating fees can boost learning and reduce dropout.

Feed children at school

There is a long history of evaluation evidence demonstrating that school meals boost enrollment at school. A large-scale school meals program in Pakistan – reaching hundreds of thousands of girls – reduced one measure of malnutrition (wasting) by almost half and boosted school enrollment by forty percent (Pappas et al. 2008). There is some evidence that children across 32 African countries benefiting from a World Food Program school feeding initiative were enrolled in school at higher rates, with 27 percent higher gains for girls than for boys (Gelli et al. 2007).

A new generation of evidence demonstrates that school meals boost student learning as well. In Ghana, a large-scale school feeding program for which funding is now integrated into the government budget was evaluated via randomized controlled trial. After two years, both math and literacy scores rose for all children on average, but the largest impacts were for girls and for children in poverty. In other words, school feeding boosted learning especially for girls (Aurino et al. 2020). Likewise, a large-scale midday meal program in India led to improved test scores in both math and reading: in that case, girls and boys benefitted equally, as did poorer and less poor children (Chakraborty and Jayaraman 2019).

Cash transfers

Cash transfers are a widely used tool to achieve multiple ends: often the primary goal is that of a social safety net—ensuring that low-income households have money for essential needs—but they are often paired with further objectives related to health and education, either explicitly though conditions that households must fulfill to receive the transfers or more subtly through labeling and encouragement (Benhassine et al. 2015). Many of these programs have been implemented at scale, and there are many variations, including those that combine a transfer with the direct payment of school fees, as in Bangladesh (Schurmann 2009). There have been many evaluations of cash transfer programs on education, and most of those (nearly two-thirds) do report impacts separately for girls (Evans et al. 2020). However, one recent, fairly comprehensive review identified only a handful of studies that reported test scores or grade completion (Baird et al. 2014). That review reports consistent positive impacts on school enrollment—with higher impacts for conditional programs—but small impacts on test scores. Indeed, a recent report rates cash transfers as a "bad buy" if the objective is to boost learning (Global Education Evidence Advisory Panel 2020). That said, ensuring that youth are in school is a likely precondition for learning and completion outcomes, so while cash transfers are not an effective instrument for learning alone, they may be crucial to making sure the most vulnerable children have the opportunity to reap learning gains from other interventions.

Make school more physically accessible

Two classes of interventions to make schools more accessible—constructing schools and providing transportation—have been implemented at scale with success and have shown positive impacts. Indeed, the Global Education Evidence Advisory Panel's Smart

Buys report identifies reducing travel times to schools as one of its "good buys" (its second best rating, after "great buys") for boosting learning in general, citing school construction and school transportation as two tested instruments (Global Education Evidence Advisory Panel 2020). The benefits are experienced disproportionately by girls. Distance to school is a major constraint for many girls, especially at the secondary school level. For example, teenage girls in India who live 15 kilometers from a school have been more than one-third less likely to attend than those who live near a school (Muralidharan and Prakash 2017).

One solution to that challenge is to build schools close to girls. Several interventions that have employed this approach have constructed schools with the needs of girls in mind (i.e., "girl-friendly" schools). A program that constructed primary schools in Burkina Faso benefited many thousands of children (the precise number is unclear): after 2.5 years, enrollment and test scores rose, and both of these effects were greater for girls than for boys (Kazianga et al. 2013). After ten years, impacts on test scores and enrollment remained although they were smaller, which may be unsurprising because many comparison villages also had some sort of school. However, primary completion rates for girls were more than double in beneficiary villages than in comparison villages (23 percent versus 9 percent); they were also much higher for boys (39 percent versus 30 percent). Marriage rates for girls were also much lower in beneficiary villages (33 percent versus 39 percent). The schools all had separate latrines for boys and girls and a water source. Importantly, the program did not merely construct schools: it also provided school meals, books, and an information campaign to parents on the importance of education (Davis et al. 2016). A slightly smaller scale program in Niger boosted student test scores, but only for girls (Bagby et al. 2016).

We know something about the long-run impacts of school construction. In Indonesia, the government implemented a massive school construction program: more than 61,000 schools were constructed between 1973 and 1979. Women were more likely to complete primary school (by 4 percentage points) as a result, and their children were more likely to complete secondary school and even tertiary education, with slightly larger effects on their daughters than on their sons (Akresh et al. 2019).

A related intervention extends existing schools. In several countries of Latin America (Brazil, Colombia, Ecuador, Honduras, and Nicaragua), a program provided alternative lower- and upper-secondary education to youth who otherwise would not have access. The program is "alternative" in the sense that it seeks to integrate academic learning with practice livelihood experience, and it has reached some 300,000 students (Kwauk and Robinson 2016). An evaluation of the program in Honduras demonstrated higher test scores at lower cost than traditional schools (McEwan et al. 2015).

The second involves making transportation to school more accessible. In India, a program provided cash transfers intended for the purchase of bicycles to more than 150,000 girls. School principals then provided receipts demonstrating that the cash had been used to purchase bicycles. The impacts of the program are striking: the gender gap in secondary school fell by 40 percent, and the girls who passed the high-stakes secondary school exam rose by 12 percent (Muralidharan and Prakash 2017). A much smaller version of the program (benefiting several thousand girls) distributed bicycles to younger schoolgirls

(in upper primary school); the evaluation found the program reduced absenteeism, commute time, and teasing but had less dramatic effects otherwise (Fiala et al. 2020).

Teach better

General improvements in the quality of instruction

In Kenya, a multi-faceted literacy and numeracy program was implemented through government systems. It included detailed teachers' guides, training for teachers and head teachers, teacher coaching, and literacy and math books for every student. The pilot program—evaluated via a randomized controlled trial—led to impressive literacy gains in the early years of primary school, with comparable impacts for girls and boys. Impacts for mathematics were more modest but still positive (Piper and Mugenda 2014). When the program was scaled up nationwide (an initiative called Tusome), literacy gains were still positive and sizeable, even though some aspects of implementation were not as well done as they were at the pilot. For example, teachers received some coaching in the scaled up version but less than planned (Piper et al. 2018). Gains were slightly bigger for girls than for boys on several of the literacy measures (Fraudenberger and Davis 2017). A related program, implemented at large scale in Pakistan and evaluated using a quasi-experimental program, also provided detailed teachers' guides and reading materials for students, among other activities (Chemonics International Inc. 2019). Beneficiary children boosted their reading more than non-beneficiary children, with larger gains for girls (Management Systems International 2018).

Target children who fall behind

In India, a remedial reading program implemented by Pratham, a large NGO, reaches hundreds of thousands of students. Hiring young women to teach students who are falling behind in their basic numeracy and literacy skills led to significant gains for both girls and boys (Banerjee et al. 2007). Pratham has implemented several variations on programs to bring learners who are falling behind up to speed, including intensive summer reading camps or using one hour of the school day to group learners by ability rather than official grade level (Banerjee et al. 2017).

In Ghana, the government implemented a related program with more than 80,000 students without NGO-support (beyond a visit to Pratham in India to see their programs in action). Schools hired teaching assistants, many of whom had no previous experience with teaching. In one model, the assistants pulled remedial learners out of class for part of the school day for special attention. In a second model, the assistants provided the same attention but after school hours. In a third, assistants worked with half of a class for part of the school day, effectively just reducing class size and allowing more specialized attention per student. In a fourth model, there were no teaching assistants, but teachers were trained to focus their teaching at whatever learning level students were at, rather than being bound by the curriculum for their grade. All four models led to significant gains in student learning, and three of the four (all except the third) were equally cost-effective. Gains were higher for girls than for boys (Duflo et al. 2020).

What about other teacher policies?

Teacher incentives in India, implemented in schools covering more than 20,000 students, led to sizeable, significant gains in language and especially math scores, with no significant differences for girls and boys (Muralidharan and Sundararaman 2011). In another teacher performance pay intervention in Tanzania that reached more than 120,000 students, students in schools that received a combination of school grants and teacher performance incentives saw the largest test score gains, and the gains for girls were significantly larger than those for boys (Mbiti et al. 2019).[7]

A large-scale teacher professional development in China had no impact on student learning for girls (or for boys), arguably because the training was too theoretical (Loyalka et al. 2019). More broadly, at-scale teacher professional development programs tend not to incorporate the elements that are associated with the best student learning outcomes in smaller programs (Popova et al. forthcoming). A multi-faceted educational program in Tanzania, evaluated with a quasi-experimental design, led to significant gains in literacy and numeracy in the early grades of primary school, with the biggest gains for girls. While attribution across components is difficult, the authors hypothesize that the teacher professional development played a major role, as it was implemented effectively and led to increase use of teaching aids, boosted teacher confidence and motivation (Ruddle and Rawle 2020).

Deploy effective forms of education technology to improve instruction

Many education systems invest extensively in education technology (or ed-tech), although reviews have found decidedly mixed impacts of education technology investments on student learning outcomes (Bulman and Fairlie 2016; Escueta et al. 2020; Evans and Mendez Acosta 2021; Rodriguez-Segura 2020).[8] The heterogeneity of impacts likely derives from the fact that technology plays many roles in education: technology can be used to improve the quality of instruction, to nudge parents or students, or for self-led learning. Some interventions to improve the quality of instruction have been successful. Many of these programs have been implemented at large scale, although the evaluations often study a smaller sample of students. For example, providing children with technology-aided after school instruction in urban India lead to large gains in math and language scores that were comparably sized for boys and girls. The evaluation sample was fewer than one thousand students, but the software has been used by more than hundreds of thousands of students (Muralidharan et al. 2019). However, this presents a conundrum: a small evaluation may have much more controlled conditions than large-scale use of a software. In this case, a larger, in-school version of the program was implemented for one period per day and still delivered significant positive learning impacts, albeit smaller than those in the pilot, after school program (Muralidharan and Singh 2019). Likewise, a large-scale computer-assisted learning platform in Uruguay that has reached at least half of all students in third through sixth grade delivered significant

---

[7] Another class of teacher incentives provides additional financing to teachers who work in rural or otherwise disadvantages schools. While those incentive programs, implemented at scale, tend to be effective in leading teachers to relocate, they did not boost girls' test scores in Zambia (Chelwa et al. 2019), although they did, particularly for higher income girls, in the Gambia (Pugatch and Schroeder 2018).

[8] Some high-performing education systems in high-income countries have actually reduced investments in education technology relative to education systems in other high-income countries (Ripley 2014).

learning gains of comparable size for both girls and boys (Perera and Aboal 2019). Both at-scale computer-assisted learning programs delivered similar results: roughly 0.2 standard deviations of mathematics learning, which amounts to a significant gain relative to a year's learning (Evans and Yuan 2019).

There is less evidence on educational television, although there are several ongoing studies of educational television and radio in the context of the COVID crisis (World Bank 2020a). The existing evidence suggests that high-quality educational television programming can boost cognitive outcomes for girls in particular, although that evidence is from younger children (Mares and Pan 2013). That said, across African countries, even with schools closed during the pandemic, relatively few households have access to ed-tech products (Crawfurd 2020).

It is with some reserve that we include ed-tech among the solutions, since simply providing education technology does not inherently promise any gains. Indeed, several large-scale programs have provided computing equipment and had no impact on student learning or other outcomes. This result is important because scaling hardware programs, while costly, may appear logistically straightforward. But providing computers to more than 6,000 schools in Colombia had no impact on student learning for girls or for boys (Linden and Barrera-Osorio 2009). Providing laptops (more than 1.6 million of them!) to children in Uruguay had no impact on girls' learning outcomes in either the short run or the longer run (Yanguas 2020); a similar program that delivered tens of thousands of laptops in Peru had no impact on girls' learning nor on students' dropout rates (Cristia et al. 2014; 2017). Providing laptops or desktops may be a necessary step in providing technology-assisted learning, but programs that do the first without the second are poised for failure.

Many other technological innovations have been either implemented only at smaller scale, with far fewer than 10,000 students (e.g., Berlinski et al. 2016 or Duflo et al. 2012). In other cases, ed-tech interventions have been implemented at large scale but lack serious evaluation on student outcomes, e.g., several large-scale education technology platforms implemented in India (Dhar et al. 2016; Bajpai et al. 2019).

Gender focused interventions

While many obstacles affect girls and boys differentially, several classes of interventions seek explicitly to address obstacles felt principally by girls. Many of these interventions are not focused on academic outcomes, and so do not report school completion or learning outcomes,[9] although there are exceptions. But a lack of evidence on those outcomes does not mean these are not important investments. It may mean that the primary motivations for investing in some of these programs are not to boost learning outcomes or school completion but rather to protect girls, to boost their overall well-being, or to prepare them for life after school.

---

[9] One example of an evaluation that did not report learning or completion outcomes is the Power to Lead program, which reached tens of thousands of adolescent and pre-adolescent girls across six countries. A mostly qualitative evaluation found positive impacts on measures of leadership skills, leadership action, and self-confidence (Miske Witt & Associates Inc. 2011).

Sanitation and menstrual health products

Toilets for girls often arise in discussions of gender-equitable education. Indeed, a national school latrine construction program in India increased enrollment for girls. For the youngest girls, any latrine boosted enrollment. For girls who had reached puberty (i.e., upper primary), only sex-specific enrollment boosted outcomes. While the study did not measure results on completion, the enrollment results are strong and enduring three years after construction. Students did not perform better on direct tests of their learning, but girls (and boys as well) both sat for and passed their official school exams at higher rates (Adukia 2017).

Two reasons that sex-specific latrines often come up in discussions is because of the fear that girls will either miss school during menstruation or that they might experience verbal or physical harassment. On the former point, absenteeism due to menstruation may not be a major problem: in Nepal, researchers estimated that girls on average missed less than half a day over the course of the school year (Oster and Thornton 2011). Menstruation likewise does not increase the gender gap in education in Malawi (Oster and Thornton 2011), and in Kenya, girls reaching the age of menarche were not more likely to be absent than boys—except for absences due to school transfers (Benshaul-Tolonen et al. 2019). In the Kenya study, providing sanitary pads (which are girls' choice) did reduce absenteeism, although those results did not hold up if one excludes girls who had transferred away from the study schools. In both Kenya and Nepal, providing an alternative, less familiar technology—a menstrual cup—did not affect absenteeism. (The cup is much cheaper, but an intervention cannot be cost-effective if it is not effective!) Regardless, these are both relatively small studies, with a few hundred students in Nepal and a few thousand in Kenya. However, looking beyond absenteeism, providing either pads or a cup in Kenya significantly improved emotional and social well-being (Benshaul-Tolonen et al. 2020b); and in Tanzania, having insufficient menstrual materials was associated with more teasing of girls about their periods (Benshaul-Tolonen et al. 2020a). On net, the evidence for providing menstrual hygiene materials on a purely instrumental basis (getting girls to school completion or higher test scores) is still weak, although the emotional well-being of female students is another important, worthy end in itself.

Gender-sensitization training

When discussing gender equality in education, a class of program that often arises in discussion is some sort of training to increase the sensitivity of teachers, school managers, or students to gender issues. To date, there is limited evidence evaluating such programs at large scale or on outcomes such as school completion or student learning. For example, one program evaluated the impact of 2.5 years of classroom discussions on the topic of gender equality among several thousand sixth and seventh graders in India. (The evaluation sample was about 14,000, with just under half of those students receiving the program.) While that is not a small-scale program, it also isn't at a massive scale, and the outcomes measured are reported views on gender norms and some self-reported behaviors (e.g., boys reported a higher likelihood of doing chores). The program tested neither student learning (beyond on gender attitudes) nor school completion (Dhar et al. 2020).

Another program that functionally acted as a gender sensitization program was the Power to Lead program, which provided training leadership skills for girls across six countries. In practice, more than 30 percent of participants were boys, and these reported improved understanding of gendered social norms (Baric 2013).

The fact that neither of these studies measured school completion or student academic learning is not a critique: those were not the principal objectives of the programs, and reducing sexism is a valuable objective in its own right. But it may also have instrumental value, increasing completion or student learning—for example, if girls are able to focus more on their studies in a more gender sensitive environment (i.e., with less teasing or harassment). Further research will be required to see if this is in fact the case.

Safe spaces for girls

Creating places where female students can engage without boys or men is sometimes proposed as a useful intervention (Megevand and Marchesini 2020). Indeed, evidence on this type of intervention is promising, although the outcomes measure is not school completion or student test scores. For example, a program that formed clubs for more than 50,000 adolescent girls in Uganda and provided vocational training and information about sex, reproduction, and marriage, led to reduced adolescent pregnancy and more engagement in income-generating activities four years later (Bandiera et al. 2020a). However, an adaptation of the same program in Tanzania had no impact on similar outcomes (Buehren et al. 2017).

A program forming girls' clubs in Sierra Leone—with a smaller sample of 150 beneficiary villages—was interrupted by school closures due to the 2014/2015 Ebola epidemic; but five years later, girls in communities where girls' clubs had initially been formed were much less likely to have experienced an adolescent pregnancy and much more likely to have re-enrolled in school after the epidemic (Bandiera et al. 2020b). These are important outcomes beyond education; the impact of these types of programs on purely educational outcomes are less well known at a large scale. A small scale, government implemented program in 20 low-performing secondary schools in Trinidad and Tobago converted co-education schools to single sex schools: girls subsequently performed better on secondary school completion exams, and both boys and girls took more advance coursework (Jackson 2021).

Other interventions

Effective interventions with less evidence behind them

Some interventions have proven effective at scale but have been rigorously evaluated in only one or two settings, so policymakers and partners may feel less confident that their success can be replicated elsewhere. For example, a program that provided eyeglasses to almost 30,000 students in China boosted test scores equivalent to nearly a full year of business-as-usual schooling for students with poor vision, benefiting both girls and boys, although one-third of girls refused eyeglasses while only one-fourth of boys did (Glewwe et al. 2016). Likewise, a program in India – implemented by the NGO Pratham (like some of the literacy programs above) – reached a large sample of mothers to either

provide literacy for mothers, train mothers in how to boost their children's literacy, or both. All three variations of the program had positive (but modest) impacts on children's mathematical performance (Banerji et al. 2017). Of course, providing eyeglasses to children who need them is a worthy and important objective, as is making sure that mothers can support their children at home. But the question is not the worthiness of the objective but rather how well such programs can be implemented at scale and the gains relative to other programs in terms of outcomes of interest.

Areas with more limited, mixed evidence

School accountability, often involving publicizing either resources flowing to schools or student performance in schools, is an important area. In Uganda, using newspapers to publicize the amount of funds that schools would be receiving from the central government boosted student learning outcomes, with apparently larger effects for girls (Reinikka and Svensson 2011). But programs publicizing results – at scale – have been less effective. For example, a program in India that facilitated meetings to discuss education and – in some cases – invited community members to create "report cards" on learning status in the community had no impact on learning outcomes for girls (or boys) (Banerjee et al. 2010). In Tanzania, a nationwide, low-stakes accountability program published school rankings: while it boosted learning outcomes in the poorest performing schools, it also led those schools to exclude students – both girls and boys, to equal degrees – from their last year of schooling, an unfortunate, unintended consequence (Cilliers et al. 2020b). Many school accountability programs reached large numbers of students do not report impacts separately for girls (e.g., Barr et al. 2012 in Uganda, Pandey et al. 2008 in India, Andrabi et al. 2017 in Pakistan, and many others).

Discussion

In this review, we have presented various initiatives that have been implemented at large scale, usually through government channels, resulting in large learning or completion gains, especially for girls. However, the reader may ask: which of these initiatives is the best bet? Unfortunately for anyone seeking a silver bullet, the answer comes down to the economist's favorite answer: it depends. Let us demonstrate with two—hopefully obvious—examples.

When is school construction an effective intervention? In Benin in the 1990s, the government or NGOs constructed more than 1,500 new schools, and enrollment surged by almost 200,000, driven mostly by girls. But the increase in enrollment was concentrated in rural areas, where there were fewer schools before the construction boom (Deschênes and Hotte 2019). The interventions where school construction was effective, in Burkina Faso and Niger and Indonesia, were in locations or times when schools were scarcer. Is school construction a good bet? Yes, if there are few schools.

When is school feeding an effective intervention? In contrast to the examples we highlighted above, a large-scale school feeding program in Chile had no impact on learning outcomes or grade progression (McEwan 2013). A program in Sri Lanka similarly showed no impacts. Why not? Chile had already eliminated extreme malnutrition and educational outcomes were relatively strong. Likewise, Sri Lanka already had high rates of enrollment (Snilstveit et al. 2016). School feeding is a powerful tool, but only in places where children face this particular need. Is school feeding a good bet? Yes, if children are malnourished. Ultimately, the most effective intervention in a given location will depend on the circumstances of that location.

Thus, how should an education system decide which interventions to invest in to boost girls' education at scale? We propose three aspects for consideration: constraints to girls' education, potential solutions, and program costs.

Constraints to girls' education

First, education systems need to identify their key constraints. What are the weaknesses in the education system? What is holding girls' education back? In recent years, a range of diagnostic tools have been deployed that have demonstrated challenges in the education system overall, and some can be deployed towards understanding girls' education. Several tools measure student learning, including citizen-led assessments like ASER and Uwezo, along with school-based assessments included in the Service Delivery Indicators and in national and regional exams (ASER 2021; Uwezo 2021; World Bank 2021). These can help identify which regions face the biggest gaps in learning. Likewise, high quality gender-disaggregated systems data can track student school completion rates. In terms of understanding the reasons behind high student dropout or poor student learning, some household surveys (e.g., some of the Demographic and Health Surveys or the Living Standards Measurement Surveys) ask directly about reasons that girls drop out of school. Furthermore, the Service Delivery Indicators measure the health of the teacher workforce, with a focus on skills and absenteeism. If a survey demonstrates that teacher

absenteeism (or any other issue) is an important problem, additional diagnostics may be needed to understand the reasons behind the issue, in order to design the most effective mechanisms.

Other diagnostic tools seek to measure the quality of the overall education system. They may be helpful in reviewing the inputs, resources, different actors and institutions, and the politics and feedback loops to improve education systems for girls (Savage and Martinez 2019; Faul 2016). A review of country-level efforts to promote policies that are more supportive for girls education, such as the Girl's Education Policy Index (Crawfurd and Hares 2020), can help to identify systemic exclusionary policies. Public expenditure tracking surveys measure how well resources reach schools and can inform education priorities.

Potential solutions

With the rapid increase of evaluation evidence, education systems have many solutions available to them. What evidence should policymakers draw on in selecting the best investments to boost girls' education at scale? If there is high-quality local evidence on effective solutions, then that can be an excellent source.[11] If not, then Bates and Glennerster (2017) propose a four-step framework for deciding if evidence of an effective program in one place will apply in another: (1) understand the theory behind the original program; (2) verify that conditions in the new location hold for the same theory to apply; (3) weigh the strength of the evidence that the same mechanism would work to change behavior in the new location; and (4) determine the likelihood that the program can be effectively implemented in the new location. This process involves drawing on a mix of the most rigorous evidence from anywhere and the best available local evidence.

No policy or program operates in a vacuum, so a key, iterative interplay between constraints and solutions will entail examining proposed solutions in the context of existing policies and how they are likely to interact. Effective overall education system reforms that deliver significant gains to girls as well as boys, like those documented in Finland over several decades (Sahlberg et al. 2021) or the Brazilian state of Ceará over a shorter period (Loureiro et al. 2020), require a collection of solutions.

Like all reviews, this study is limited by those areas that have been evaluated. Education systems should continue to innovate. Some of that innovation may be in adapting within areas already shown to be effective. Reducing education costs for girls, for example, may be accomplished in various ways. Other innovation may be completely new. However, there are areas that have not been effective at scale across multiple settings, like distribution of computer equipment either without plans or capability to integrate it fully into the system or without accompanying investments in the complementary technologies needed to deliver gains in learning. These should be avoided.

---

[11] If there is local evidence, Pritchett and Sandefur (2013; 2015) argue that less rigorous local evidence may be more relevant than more rigorous evidence from another context. This requires at least two caveats, however. First, in many cases, there is no relevant local evidence of impact (even non-rigorous evidence). Second, whether or not this is true will depend fundamentally on the size of the selection bias (which RCTs overcome) relative to the size of the impact of the program or policy. If the selection bias is large (e.g., poor people are more likely to get cash transfers and sickly people are more likely to go to the doctor), then non-experimental methods may be less informative.

Program costs

Every new program and most new policies come with a price tag. Ultimately, we care about both effectiveness in delivering gender equality in education and about cost-effectiveness. By definition, the most cost-effective interventions deliver the biggest gains per dollar spent. But cost-effective interventions that deliver small gains, while often worth doing because of low costs, will not be sufficient to close gender gaps. So an information campaign that costs little may be worth doing because of a high benefit-cost ratio. A school feeding program or a school fee elimination program will cost much more but may—depending on the constraints—deliver larger gains as well. (Programs like cash transfers and school feeding programs may appear less cost effective purely in terms of education gains because many of their benefits extend beyond the education sector.)

A minority of evaluations report costs, but the proportion appears to be growing over time (McEwan 2015; Evans and Mendez Acosta 2021). Yet just like impact estimates, costs for the same program can vary significantly across contexts (Evans and Popova 2016). Just as with the four-step approach for adapting impact estimates across contexts, education teams and those who support them will need to adapt cost estimates across contexts as well.

Lessons for scaling

Earlier in the paper (The Knowledge Challenge), we discussed some of the challenges in taking interventions to scale. First, as programs expand, they often outgrow their monitoring, management, and evaluation systems, especially if the scale up involves new partners or a larger (maybe even different) government agency. In some instances, scaling up might mean creating new institutions or legal entities to manage the expansion effectively (Cull and McKenzie 2020).

Second, political momentum and political champions are critical resources to ensure that the program both acquires and retains legitimacy to survive changes in administrations. In Ceará, Brazil, arguably the single most important factor in a successful education reform that dramatically boosted student learning outcomes was consistent political leadership (Loureiro et al. 2020). In addition, scaled up programs or national reforms also capture constituents' attention more than small pilots. In the case of interventions targeted exclusively to girls, this may generate backlash due to a perception of "female bias" (Subrahmanian 2005). Political champions are key to providing platforms necessary for communicating to local leaders and ensuring compliance or acceptance.

Finally, special measures that deviate from existing schooling provision systems—such as alternative schools, early child education systems, or health and hygiene programs—are more likely to be sustained if they are integrated within the formal system of provision rather than standing alone as isolated efforts with separate management systems (Subrahmanian 2005). Similarly, mixed packages of reforms which include simultaneous and complementary programs will increase the likelihood that the effect of one large-scale program can build on the opportunities created by other, similar reforms.

Conclusion

What this review has sought to do is highlight interventions for which there is evidence from multiple settings that they can be implemented effectively at large scale and deliver positive impacts for girls. We have also provided a discussion of gender-specific interventions and guidance as to how to make sense of the large and growing body of evidence. There is no guarantee that a given impact will replicate in every setting, but this collection of evidence provides a menu for policymakers and partners that comes one step closer to feasible implementation than previous reviews that draw on a higher proportion of small-scale, NGO-implemented interventions. Achieving gender equitable education is an ongoing challenge, but there are proven solutions that work at scale.

References

Adukia, Anjali. 2017. "Sanitation and Education." *American Economic Journal: Applied Economics* 9 (2): 23–59. https://doi.org/10.1257/app.20150083.

Akresh, Richard, Daniel Halim, and Marieke Kleemans. 2019. "Long-Term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia." https://drive.google.com/file/d/1c7bF9so7c4bokYH5-_epR5muNjpj4Pk7/view.

Anderson, Jenny. 2018. "We Know How to Make Programs That Help Poor Kids. The Hard Part Is Scaling Up." Quartz (blog). 2018. https://qz.com/1269399/we-know-how-to-make-programs-that-help-poor-kids-the-hard-part-is-scaling-up/.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review* 107 (6): 1535–63. https://doi.org/10.1257/aer.20140774.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96 (3): 847–62. https://doi.org/10.1257/aer.96.3.847.

ASER. 2021. "ASER Centre: Evidence for Action." 2021. https://www.asercentre.org/.

Aurino, Elisabetta, Aulo Gelli, Clement Adamba, Isaac Osei-Akoto, and Harold Alderman. 2020. "Food for Thought? Experimental Evidence on the Learning Impacts of a Large-Scale School Feeding Program." *Journal of Human Resources*, December, 1019-10515R1. https://doi.org/10.3368/jhr.58.3.1019-10515R1.

Awasom, A, M.E. Mpofu, M. Mutanguha, M. Muyatwa, and P.L. Sihwa. 2020. "Co- Impact Research Proposal- Gender Equitable Educational Outcomes." Global Office Consulting.

Bagby, Emilie, Anca Dumitrescu, Cara Orfield, and Matt Sloan. 2016. "Long-Term Evaluation of the IMAGINE Project in Niger (English Version)." *Mathematica Policy Research*, June. https://www.mathematica.org/our-publications-and-findings/publications/imagine-report-english-june-2016.

Baird, Sarah, Francisco H. G. Ferreira, Berk Özler, and Michael Woolcock. 2014. "Conditional, Unconditional and Everything in between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes." *Journal of Development Effectiveness* 6 (1): 1–43. https://doi.org/10.1080/19439342.2014.890362.

Bajpai, Nirupam, John Biberman, and Anchal Sharma. 2019. "Information and Communications Technology in the Education Sector in India." https://doi.org/10.7916/d8-9anm-mc81.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2020a. "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa." *American Economic Journal: Applied Economics* 12 (1): 210–59. https://doi.org/10.1257/app.20170416.

Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra. 2020b. "Do School Closures During an Epidemic Have Persistent Effects? Evidence from Sierra Leone in the Time of Ebola." World Bank Working Paper. http://www.homepages.ucl.ac.uk/~uctpimr/research/ELA_SL.pdf.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73–102. https://doi.org/10.1257/jep.31.4.73.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, 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. https://doi.org/10.1257/pol.2.1.1.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–64. https://doi.org/10.1162/qjec.122.3.1235.

Banerji, Rukmini, James Berry, and Marc Shotland. 2017. "The Impact of Maternal Literacy and Participation Programs: Evidence from a Randomized Evaluation in India." *American Economic Journal: Applied Economics* 9 (4): 303–37. https://doi.org/10.1257/app.20150390.

Baric, Stephanie. 2013. "Where the Boys Are: Engaging Young Adolescent Boys in Support of Girls' Education and Leadership." *Gender & Development* 21 (1): 147–60. https://doi.org/10.1080/13552074.2013.769777.

Barr, Abigail, Lawrence Bategeka, Madina Guloba, Ibrahim Kasirye, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin. 2012. "Management and Motivation in Ugandan Primary Schools: An Impact Evaluation Report." 2012–14. *Working Papers PIERI.* Working Papers PIERI. PEP-PIERI. https://ideas.repec.org/p/lvl/piercr/2012-14.html.

Bates, Mary Ann, and Rachel Glennerster. 2017. "The Generalizability Puzzle." *Standford Social Innovation Review.* https://ssir.org/articles/entry/the_generalizability_puzzle.

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. https://doi.org/10.1257/pol.20130225.

Benshaul-Tolonen, Anja, Sandra Aguilar-Gomez, Naomi Heller Batzer, Rebecca Cai, and Elias Charles Nyanza. 2020a. "Period Teasing, Stigma, and Knowledge: A Survey of Adolescent Boys and Girls in Northern Tanzania." Working Paper. Center for Development Economics and Policy. https://anjatolonen.files.wordpress.com/2020/09/boys_benshaul-tolonen.pdf.

Benshaul-Tolonen, Anja, Garazi Zulaika, Elizabeth Nyothach, Clifford Oduor, Linda Mason, David Obor, Kelly T Alexander, Kayla F Laserson, and Penelope Phillips-Howard. 2019. "Pupil Absenteeism, Measurement, and Menstruation: Evidence from Western Kenya." CDEP-CGEG Working Paper Series No. 74. https://anjatolonen.files.wordpress.com/2019/03/cdep_absenteeism.pdf.

Benshaul-Tolonen, Anja, Garazi Zulaika, and Penelope Phillips-Howards. 2020b. "Sanitary Products, Absenteeism and Psychosocial Wellbeing: Evidence from a Cluster Randomized Controlled Feasibility Study in Kenya." Working Paper.

Berlinski, Samuel, Matias Busso, Taryn Dinkelman, and Claudia Martinez. 2016. "Reducing Parent-School Information Gaps and Improving Education Outcomes: Evidence from High Frequency Text Messaging in Chile." https://assets.publishing.service.gov.uk/media/58778bfa40f0b60e4c000135/726__Reducing-Parent-School-information-gap_BBDM-Dec2016.pdf.

Blimpo, Moussa P., Ousman Gajigo, and Todd Pugatch. 2019. "Financial Constraints and Girls' Secondary Education: Evidence from School Fee Elimination in The Gambia." *The World Bank Economic Review* 33 (1): 185–208. https://doi.org/10.1093/wber/lhw051.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2018. "Experimental Evidence on Scaling up Education Reforms in Kenya." *Journal of Public Economics* 168 (December): 1–20. https://doi.org/10.1016/j.jpubeco.2018.08.007.

Buehren, Niklas, Markus P. Goldstein, Selim Gulesci, Munshi Sulaiman, and Venus Yam. 2017. "Evaluation of an Adolescent Development Program for Girls in Tanzania." WPS7961. The World Bank. http://documents.worldbank.org/curated/en/245071486474542369/Evaluation-of-an-adolescent-development-program-for-girls-in-Tanzania.

Bulman, George, and Robert W Fairlie. 2016. "Chapter 5: Technology and Education: Computers, Software, and the Internet." In *Handbook of the Economics of Education, Volume 5*. Elsevier B.V. https://people.ucsc.edu/~gbulman/tech_2016.pdf.

Cameron, Drew B., Anjini Mishra, and Annette N. Brown. 2016. "The Growth of Impact Evaluation for International Development: How Much Have We Learned?" *Journal of Development Effectiveness* 8 (1): 1–21. https://doi.org/10.1080/19439342.2015.1034156.

Chakraborty, Tanika, and Rajshri Jayaraman. 2019. "School Feeding and Learning Achievement: Evidence from India's Midday Meal Program." *Journal of Development Economics* 139 (June): 249–65. https://doi.org/10.1016/j.jdeveco.2018.10.011.

Chelwa, Grieve, Miquel Pellicer, and Mashekwa Maboshe. 2019. "Teacher Pay and Educational Outcomes: Evidence from the Rural Hardship Allowance in Zambia." *South African Journal of Economics* 87 (3): 255–82. https://doi.org/10.1111/saje.12227.

Chemonics International Inc. 2019. "Sindh Reading Program: Final Report." https://chemonics.com/wp-content/uploads/2019/08/Pakistan-SRP_FinalReport_2019.pdf.

Cilliers, Jacobus, Brahm Fleisch, Janeli Kotze, Nompumelelo Mohohlwane, Stephen Taylor, and Tshegofatso Thulare. 2020a. "Can Virtual Replace In-Person Coaching? Experimental Evidence on Teacher Professional Development and Student Learning in South Africa." Working Paper. https://www.dropbox.com/s/n0hza4nyrvz4mpt/Can%20Virtual%20Replace%20InPerson%20Coaching%20Working%20Paper.pdf?dl=0.

Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor. 2019. "How to Improve Teaching Practice? An Experimental Comparison of Centralized Training and in-Classroom Coaching." *Journal of Human Resources*, February, 0618-9538R1. https://doi.org/10.3368/jhr.55.3.0618-9538R1.

Cilliers, Jacobus, Isaac M. Mbiti, and Andrew Zeitlin. 2020b. "Can Public Rankings Improve School Performance? Evidence from a Nationwide Reform in Tanzania." *Journal of Human Resources*, March, 0119-9969R1. https://doi.org/10.3368/jhr.56.3.0119-9969R1.

Crawfurd, Lee. 2020. "Why the COVID Crisis Is Not Edtech's Moment in Africa." *Center For Global Development Blog Post* (blog). 2020. https://www.cgdev.org/blog/why-covid-crisis-not-edtechs-moment-africa.

Crawfurd, Lee, and Susannah Hares. 2020. "Introducing the Girl's Education Policy Index." 2020. https://www.cgdev.org/blog/introducing-girls-education-policy-index.

Cristia, Julian, Alejo Czerwonko, and Pablo Garofalo. 2014. "Does Technology in Schools Affect Repetition, Dropout and Enrollment? Evidence from Peru." *Journal of Applied Economics* 17 (1): 89–111. https://doi.org/10.1016/S1514-0326(14)60004-0.

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2017. "Technology and Child Development: Evidence from the One Laptop per Child Program." *American Economic Journal: Applied Economics* 9 (3): 295–320. https://doi.org/10.1257/app.20150385.

Cull, Bob, and David McKenzie. 2020. "Implementing Successful Small Interventions at a Large Scale Is Hard." *World Bank Development Impact Blogs* (blog). March 2020. https://blogs.worldbank.org/developmenttalk/implementing-successful-small-interventions-large-scale-hard.

Davis, Mikael, Nicholas Ingwersen, Harounan Kazianga, Leigh Linden, Arif Mamun, Ali Protik, and Matt Sloan. 2016. "Ten-Year Impacts of Burkina Faso's BRIGHT Program." Mathematica Policy Research. https://www.mathematica.org/our-publications-and-findings/publications/ten-year-impacts-of-burkina-fasos-bright-program.

DellaVigna, Stefano, and Elizabeth Linos. 2021. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." Working Paper. Cambridge, MA. https://doi.org/10.3386/w27594.

Deschênes, Sarah, and Rozenn Hotte. 2019. "Assessing the Effects of an Education Policy on Women's Well-Being: Evidence from Benin." halshs-02179704. PSE Working Papers. HAL. https://ideas.repec.org/p/hal/psewpa/halshs-02179704.html.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2020. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." https://faculty.wcas.northwestern.edu/~sjv340/reshaping_gender_attitudes.pdf.

Dhar, Vasant, Nandan Nilekani, Shankar Maruwada, and Nagaraju Pappu. 2016. "Big Data As an Enabler of Primary Education." *Big Data* 4 (3): 137–40. https://doi.org/10.1089/big.2016.29011.vdc.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas. 2020. "External Validity: Four Models of Improving Student Achievement." Working Paper 27298. National Bureau of Economic Research. https://doi.org/10.3386/w27298.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123 (March): 92–110. https://doi.org/10.1016/j.jpubeco.2014.11.008.

———. 2019. "The Impact of Free Secondary Education: Experimental Evidence from Ghana." *Massachusetts Institute of Technology Working Paper Cambridge*. https://web.stanford.edu/~pdupas/DDK_GhanaScholarships.pdf.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78. https://doi.org/10.1257/aer.102.4.1241.

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos, and Vincent Quan. 2020. "Upgrading Education with Technology: Insights from Experimental Research." *Journal of Economic Literature* 58 (4): 897–996. https://doi.org/10.1257/jel.20191507.

Espinoza, Oscar. 2007. "Solving the Equity–Equality Conceptual Dilemma: A New Model for Analysis of the Educational Process." *Educational Research* 49 (4): 343–63. https://doi.org/10.1080/00131880701717198.

Evans, Alice. 2019. "How Cities Erode Gender Inequality: A New Theory and Evidence from Cambodia." *Gender & Society* 33 (6): 961–84. https://doi.org/10.1177/0891243219865510.

Evans, David, Maryam Akmal, and Pamela Jakiela. 2021a. "Gender Gaps in Education: The Long View." *IZA Journal of Development and Migration* 12 (1). https://doi.org/10.2478/izajodm-2021-0001.

Evans, David, Charles Gale, and Katrina Kosec. 2020. "The Education Impacts of Cash Transfers for Children with Multiple Indicators of Vulnerability." Working Paper. Center for Global Development. https://www.cgdev.org/publication/education-impacts-cash-transfers-children-multiple-indicators-vulnerability.

Evans, David, Susannah Hares, Amina Mendez Acosta, and Christelle Saintis. 2021b. "It's Been a Year Since Schools Started to Close Due to COVID-19." *Center For Global Development Blog Post* (blog). 2021. https://www.cgdev.org/blog/its-been-year-schools-started-close-due-covid-19.

Evans, David, and Amina Mendez Acosta. 2021. "Education in Africa: What Are We Learning?" *Journal of African Economies* 30 (1): 13–54. https://doi.org/10.1093/jae/ejaa009.

Evans, David, and Anna Popova. 2016. "Cost-Effectiveness Analysis in Development: Accounting for Local Costs and Noisy Impacts." *World Development* 77 (January): 262–76. https://doi.org/10.1016/j.worlddev.2015.08.020.

Evans, David, and Fei Yuan. 2019. "Equivalent Years of Schooling : A Metric to Communicate Learning Gains in Concrete Terms." WPS8752. The World Bank. http://documents.worldbank.org/curated/en/123371550594320297/Equivalent-Years-of-Schooling-A-Metric-to-Communicate-Learning-Gains-in-Concrete-Terms.

———. 2020. "How Big Are Effect Sizes in International Education Studies?" Center for Global Development. https://www.cgdev.org/publication/how-big-are-effect-sizes-international-education-studies.

———. forthcoming. "What We Learn about Girls' Education from Interventions That Do Not Focus on Girls." *World Bank Economic Review*. http://documents.worldbank.org/curated/en/243741563805734157/What-We-Learn-about-Girls-Education-from-Interventions-that-Do-Not-Focus-on-Girls.

Faul, Moira. 2016. "Scoping Study: Education Systems Diagnostics Tools." https://www.academia.edu/31407684/Scoping_Study_Education_Systems_Diagnostics_Tools.

Fazzio, Ila, Alex Eble, Robin Lumsdaine, Peter Boone, Baboucarr Bouy, Jenny Hsieh, Chitra Jayanty, Simon Johnson, and Ana Filipa Silva. 2020. "Large Learning Gains in Pockets of Extreme Poverty: Experimental Evidence from Guinea Bissau." Working Paper. http://www.alexeble.com/research.

Fiala, Nathan, Ana Garcia-Hernandez, Kritika Narula, and Nishith Prakash. 2020. "Wheels of Change: Transforming Women's Lives with Education and Empowerment." In . https://www.dropbox.com/s/is9sc8khliipbx3/Slides_Cycles%20%28Dec%202019%29.pdf?dl=0.

Fraudenberger, Elizabeth, and Jeff Davis. 2017. "Tusome External Evaluation Midline Report." https://ierc-publicfiles.s3.amazonaws.com/public/resources/Tusome%20Midline%20evaluation%202017%20final%20report%20from%20DEC.pdf.

Gelli, Aulo, Ute Meir, and Francisco Espejo. 2007. "Does Provision of Food in School Increase Girls' Enrollment? Evidence from Schools in Sub-Saharan Africa." *Food and Nutrition Bulletin* 28 (2): 149–55. https://doi.org/10.1177/156482650702800203.

Glewwe, Paul, Albert Park, and Meng Zhao. 2016. "A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China." *Journal of Development Economics* 122 (September): 170–82. https://doi.org/10.1016/j.jdeveco.2016.05.007.

Global Education Evidence Advisory Panel. 2020. "Cost-Effective Approaches to Improve Global Learning: What Does Recent Evidence Tell Us Are 'Smart Buys' for Improving Learning in Low and Middle Income Countries?" World Bank. http://documents1.worldbank.org/curated/en/719211603835247448/pdf/Cost-Effective-Approaches-to-Improve-Global-Learning-What-Does-Recent-Evidence-Tell-Us-Are-Smart-Buys-for-Improving-Learning-in-Low-and-Middle-Income-Countries.pdf.

Gove, Amber, Medina Korda Poole, and Benjamin Piper. 2017. "Designing for Scale: Reflections on Rolling Out Reading Improvement in Kenya and Liberia." *New Directions for Child and Adolescent Development* 2017 (155): 77–95. https://doi.org/10.1002/cad.20195.

Islam, Asad. 2019. "Parent–Teacher Meetings and Student Outcomes: Evidence from a Developing Country." *European Economic Review* 111 (January): 273–304. https://doi.org/10.1016/j.euroecorev.2018.09.008.

Jackson, C. Kirabo. 2021. "Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?" *Journal of Human Resources* 56 (1): 1–39. https://doi.org/10.3368/jhr.56.1.0618-9558R2.

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. https://doi.org/10.1257/app.5.3.41.

Kerwin, Jason T., and Rebecca L. Thornton. forthcoming. "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures." *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_00911.

Kim, jim Yong. 2016. "World Bank Group President Jim Yong Kim's Remarks at 'Let Girls' Learn' Spring Meetings Event." Text/HTML, Washington DC, United States. https://www.worldbank.org/en/news/speech/2016/04/13/world-bank-group-president-jim-yong-kims-remarks-at-let-girls-learn-spring-meetings-event.

Kraft, Matthew A., and David Blazar. 2018. "Taking Teacher Coaching To Scale." *Education Next* (blog). July 31, 2018. https://www.educationnext.org/taking-teacher-coaching-to-scale-can-personalized-training-become-standard-practice/.

Kwauk, Christina, and Jenny Perlman Robinson. 2016. "Sistema de Aprendizaje Tutorial: Redefining Rural Secondary Education in Latin America." The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2016/07/FINAL-SAT-Case-Study.pdf.

Linden, Leigh, and Felipe Barrera-Osorio. 2009. "The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia." https://openknowledge.worldbank.org/bitstream/handle/10986/4032/WPS4836.pdf?sequence=1.

Loureiro, Andre, Louisee Cruz, Ildo Lautharte, and David Evans. 2020. "The State of Ceara in Brazil Is a Role Model for Reducing Learning Poverty." World Bank. https://openknowledge.worldbank.org/handle/10986/34156.

Loyalka, Prashant, Anna Popova, Guirong Li, and Zhaolei Shi. 2019. "Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program." *American Economic Journal: Applied Economics* 11 (3): 128–54. https://doi.org/10.1257/app.20170226.

Management Systems International. 2018. "2017 EARLY GRADE READING ASSESSMENT SINDH – PAKISTAN READING PROGRAM." https://earlygradereadingbarometer.org/files/Pakistan%20Sindh%20PRP.pdf.

Mares, Marie-Louise, and Zhongdang Pan. 2013. "Effects of Sesame Street: A Meta-Analysis of Children's Learning in 15 Countries." *Journal of Applied Developmental Psychology* 34 (3): 140–51. https://doi.org/10.1016/j.appdev.2013.01.001.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *The Quarterly Journal of Economics* 134 (3): 1627–73. https://doi.org/10.1093/qje/qjz010.

McEwan, Patrick J. 2013. "The Impact of Chile's School Feeding Program on Education Outcomes." *Economics of Education Review* 32 (February): 122–39. https://doi.org/10.1016/j.econedurev.2012.08.006.

———. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research* 85 (3): 353–94. https://doi.org/10.3102/0034654314553127.

McEwan, Patrick J., Erin Murphy-Graham, David Torres Irribarra, Claudia Aguilar, and Renán Rápalo. 2015. "Improving Middle School Quality in Poor Countries: Evidence From the Honduran Sistema de Aprendizaje Tutorial." *Educational Evaluation and Policy Analysis* 37 (1): 113–37. https://doi.org/10.3102/0162373714527786.

Megevand, Melanie, and Laura Marchesini. 2020. "Women and Girls Safe Spaces: A Toolkit for Advancing Women's and Girls' Empowerment in Humanitarian Settings." International Rescue Committee and International Medical Corps. https://gbvaor.net/sites/default/files/2020-02/IRC-WGSS-Toolkit-Eng.pdf.

Mendez Acosta, Amina, and David Evans. 2020. "COVID-19 and Girls' Education: What We Know So Far and What We Expect." *Center For Global Development* (blog). 2020. https://www.cgdev.org/blog/covid-19-and-girls-education-what-we-know-so-far-and-what-we-expect-happen.

Miske Witt & Associates Inc. 2011. "The Power to Lead Alliance (PTLA) Empowering Girls to Learn and Lead: Final Evaluation Report for CARE USA." http://careevaluations.org/wp-content/uploads/PTLA-FINAL-EVALUATION-REPORT_MWAI.pdf.

Muralidharan, Karthik, and Nishith Prakash. 2017. "Cycling to School: Increasing Secondary School Enrollment for Girls in India." American Economic Journal: Applied Economics 9 (3): 321–50. https://doi.org/10.1257/app.20160004.

Muralidharan, Karthik, and Abhijeet Singh. 2019. "Improving Schooling Productivity through Computer-Aided Instruction: Experimental Evidence from Rajasthan." RISE 2019, June.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426–60. https://doi.org/10.1257/aer.20171112.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77. https://doi.org/10.1086/659655.

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. https://doi.org/10.1257/app.3.1.91.

Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman. 2008. "Community Participation in Public Schools : The Impact of Information Campaigns in Three Indian States." https://openknowledge.worldbank.org/handle/10986/6359.

Pappas, Gregory, Ajmal Agha, Ghazala Rafique, Kausar S Khan, Salma Badruddin, and Habib Peermohamed. 2008. "Community-Based Approaches to Combating Malnutrition and Poor Education among Girls in Resource-Poor Settings: Report of a Large Scale Intervention in Pakistan." *Rural and Remote Health* 8 (1): 820.

Perera, M., and Diego Aboal. 2019. "The Impact of a Mathematics Computer-Assisted Learning Platform on Students' Mathematics Test Scores." /paper/The-impact-of-a-mathematics-computer-assisted-on-Perera-Aboal/ae058393d4f9da5fa415acde66fed4a31271a08e.

Piper, Benjamin, Joseph Destefano, Esther M. Kinyanjui, and Salome Ong'ele. 2018. "Scaling up Successfully: Lessons from Kenya's Tusome National Literacy Program." *Journal of Educational Change* 19 (3): 293–321. https://doi.org/10.1007/s10833-018-9325-4.

Piper, Benjamin, and Abel Mugenda. 2014. "The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation." USAID. https://pdf.usaid.gov/pdf_docs/PA00K27S.pdf.

Popova, Anna, David Evans, Mary E. Breeding, and Violeta Arancibia. forthcoming. "Teacher Professional Development around the World: The Gap between Evidence and Practice." *World Bank Research Observer*.

Pritchett, Lant, and Justin Sandefur. 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Do Not Mix." *Journal of Globalization and Development* 4 (2): 161–97. https://doi.org/10.1515/jgd-2014-0004.

———. 2015. "Learning from Experiments When Context Matters." *American Economic Review* 105 (5): 471–75. https://doi.org/10.1257/aer.p20151016.

Pugatch, Todd, and Elizabeth Schroeder. 2018. "Teacher Pay and Student Performance: Evidence from the Gambian Hardship Allowance." *Journal of Development Effectiveness* 10 (2): 249–76. https://doi.org/10.1080/19439342.2018.1452778.

Randall, Jennifer, and Alejandra Garcia. 2020. "Let's Go Girls!: Evaluating the Effectiveness of Tutoring and Scholarships on Primary School Girls' Attendance and Academic Performance in the Democratic Republic of the Congo." *FIRE: Forum for International Research in Education* 6 (3): 19–35.

Read, Lindsay, and Tamar Manuelyan Atinc. 2017. "Information for Accountability: Transparency and Citizen Engagement for Improved Service Delivery in Education Systems." https://www.brookings.edu/research/information-for-accountability-transparency-and-citizen-engagement-for-improved-service-delivery-in-education-systems/.

Reinikka, Ritva, and Jakob Svensson. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 95 (7): 956–66. https://doi.org/10.1016/j.jpubeco.2011.02.006.

Ripley, Amanda. 2014. *The Smartest Kids in the World: And How They Got That Way*. Reprint edition. New York: Simon & Schuster.

Robinson, Jenny Perlman, Patrick Hannahan, and Molly Curtiss. 2019. "Millions Learning." Brookings. February 8, 2019. https://www.brookings.edu/product/millions-learning/.

Rodriguez-Segura, Daniel. 2020. "Educational Technology in Developing Countries: A Systematic Review." University of Virginia EdPolicy Works Working Papers. https://curry.virginia.edu/sites/default/files/uploads/epw/72_Edtech_in_Developing_Countries.pdf.

Ruddle, Nicola, and Georgina Rawle. 2020. "EQUIP-Tanzania Impact Evaluation: Final Endline Report." https://www.opml.co.uk/files/Publications/8383-assessing-equip-t/1589895669_opm-ie-final-endline-report-final.pdf?noredirect=1.

Sabates, Ricardo, Pauline Rose, Benjamin Alcott, and Marcos Delprato. 2020. "Assessing Cost-Effectiveness with Equity of a Programme Targeting Marginalised Girls in Secondary Schools in Tanzania." *Journal of Development Effectiveness* 0 (0): 1–19. https://doi.org/10.1080/19439342.2020.1844782.

Sabet, Shayda Mae, and Annette N. Brown. 2018. "Is Impact Evaluation Still on the Rise? The New Trends in 2010–2015." *Journal of Development Effectiveness* 10 (3): 291–304. https://doi.org/10.1080/19439342.2018.1483414.

Sahlberg, Pasi, Ken Robinson, and Howard Gardner. 2021. F*innish Lessons 3.0: What Can the World Learn from Educational Change in Finland?* 3rd edition. New York: Teachers College Press.

Savage, Laura, and Raphaelle Martinez. 2019. "How Can We Best Diagnose Education Systems to Improve Them." *Global Partnership for Education* (blog). 2019. https://www.globalpartnership.org/blog/how-can-we-best-diagnose-education-systems-improve-them.

Schurmann, Anna T. 2009. "Review of the Bangladesh Female Secondary School Stipend Project Using a Social Exclusion Framework." *Journal of Health, Population and Nutrition* 27 (4): 505–17.

Snilstveit, Birte, Jennifer Stevenson, Radhika Menon, Daniel Philips, Emma Gallagher, Maisie Geelen, Hannah Jobse, Tanja Schimdt, and Emmanuel Jimenez. 2016. "The Impact of Education Programmes on Learning and School Participationin Low- and Middle-Income Countries." 3ie Systematic Review Summary 7. London: International Initiative for Impact Evaluation (3ie). https://www.3ieimpact.org/sites/default/files/2019-05/srs7-education-report.pdf.

Sperling, Gene, Rebecca Winthrop, and Christina Kwauk. 2016. *What Works in Girls' Education: Evidence for the World's Best Investment*. Brookings Institution Press. https://www.jstor.org/stable/10.7864/j.ctt17w8hh8.

Stern, Jonathan, Matthew Jukes, and Benjamin Piper. 2020. "Is It Possible to Improve Learning at Scale? Reflections on the Process of Identifying Large-Scale Successful Education Interventions." *Center For Global Development* (blog). 2020. https://www.cgdev.org/blog/it-possible-improve-learning-scale-reflections-process-identifying-large-scale-successful.

Subrahmanian, Ramya. 2005. "Scaling up Good Practices in Girls' Education." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000142240?posInSet=3&queryId=c1b3bcb2-ef1d-4c3b-baef-49f954dbbbbf.

United Nations. 2015. "Sustainable Development Goal 5: Achieve Gender Equality and Empower All Women and Girls." 2015. https://sdgs.un.org/goals/goal5.

Uwezo. 2021. "Uwezo Kenya." 2021. https://www.uwezo.net/.

Williams, Timothy P., Pamela Abbott, and Alfred Mupenzi. 2015. "'Education at Our School Is Not Free': The Hidden Costs of Fee-Free Schooling in Rwanda." *Compare: A Journal of Comparative and International Education* 45 (6): 931–52. https://doi.org/10.1080/03057925.2014.938611.

World Bank. 2018. "World Development Report 2018: Learning to Realize Education's Promise." Washington, DC. https://www.worldbank.org/en/publication/wdr2018.

———. 2020a. "Strategic Impact Evaluation Fund: Presenting the Evaluations of the COVID-19 Emergency Window."

———. 2020b. "The Human Capital Index 2020 Update : Human Capital in the Time of COVID-19." https://openknowledge.worldbank.org/handle/10986/34432.

———. 2020c. "World Development Indicators." 2020. https://databank.worldbank.org/source/world-development-indicators#.

———. 2021. "Service Delivery Indicators." 2021. https://www.sdindicators.org/#:~:text=Indicators,a%20school%20or%20health%20facility.

Yanguas, Maria Lucia. 2020. "Technology and Educational Choices: Evidence from a One-Laptop-per-Child Program." *Economics of Education Review* 76 (June): 101984. https://doi.org/10.1016/j.econedurev.2020.101984.