1

Great Expectations: A Framework for Assessing and Understanding Key Factors Affecting Student Learning of Foundational Reading Skills

Audrey-Marie Moore, Amber Gove, Karen Tietjen

Abstract

This article addresses the evolution of the underlying theories of change in global education reform efforts between 1990 and 2015, informed by the shift in focus from access to quality and learning. We review recent data regarding how different types of donor interventions (i.e., structural or pedagogical) have contributed to improved reading outcomes and compare effect sizes over a series of intervention studies conducted from 2003 to 2015. Against this background, we present a framework for understanding how the intensity, frequency, and fidelity of the interventions as well as the enabling environments of reform affect the magnitude and rates at which reading and learning outcomes can be expected to improve. In this, we present the context for the articles that follow, identifying the program design characteristics and types of interventions that increase the likelihood of successful expansion of the interventions commonly referred to as "scaling-up," the ability to sustain interventions, and the value (cost effectiveness) of reading programs in low- and middle-income countries. © 2017 Wiley Periodicals, Inc.

The Education for All (EFA) movement is a global commitment to provide high-quality basic education for all children, youth, and adults. At the World Education Forum, which launched EFA in 1999 (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2000), 164 governments pledged to achieve EFA by 2015. Between 2000 and 2011, donors and governments spent an average of $2–3 billion annually to improve educational outcomes for children worldwide (World Bank, 2011). This investment reflects both a widely held belief that basic education is a human right and recognition of the centrality of education to countries' economic, social, and political development. Research supports this belief, demonstrating that education leads to better family health and nutrition, improved capacity of the poor to participate in the political process, and higher lifetime income (Birdsall, Levine, & Ibrahim, 2005; Chapman & Quijada, 2008; Hanushek & Woessman, 2012; Lockheed & Verspoor, 1992; World Bank, 2003).

In recent years, the donor community, including the U.S. Agency for International Development (USAID), has focused on improving reading as a way of ensuring that children develop strong foundational skills and continue their pursuit of higher levels of education and lifelong learning (USAID, 2011). Despite recent international discussions focused on identifying key indicators for monitoring learning (and reading as an essential condition of learning), little discussion has taken place about what *level* of improvement in levels of learning can be expected, *at what scale*, and *under what conditions*. Although many approaches to reading instruction and its support exist, there has been a general consensus in the development community that reading skill acquisition is most effectively brought about by instruction in languages that the student speaks and understands, a phonics-based approach, and a materials-rich environment.

This article addresses the evolution of the underlying theories of change in global education reform efforts between 1990 and 2015. The discussion is informed by the move from focusing on access to focusing on quality and learning. We review recent data regarding how different types of donor interventions (i.e., structural or pedagogical) have contributed to improved reading outcomes and compare effect sizes over a series of intervention studies conducted from 2003 to 2015. Against this background, we present a framework for understanding how the intensity, frequency, and fidelity of the interventions as well as the enabling environments of reform affect the magnitude and rates at which reading and learning outcomes can be expected to improve. This article—and those that follow—identifies the program design characteristics and types of interventions that increase the likelihood of successful expansion of the interventions commonly referred to as "scaling-up." The articles further inform our ability to sustain the interventions and the value (cost effectiveness) of reading programs in lower and middle-income countries.

Moving from Access to Learning: The EFA Era

"Because schools change reforms as much as reforms change schools, judging an innovation's success or failure has been, and is, no easy task."

(Cuban, 1998, p. 453)

EFA has come a long way in 25 years. The number of out-of-school children worldwide has been reduced by half since 1990, and 48% of enrolled children are girls. Fifty-two countries have met the Millennium Development Goal (MDGs) of universal primary education. Despite these achievements, 93 countries are moderately to severely off of the MDG education target, and approximately 57 million children and youth remain out of school (UNESCO, 2015). Student survival to the last grade of primary education scarcely changed between 1999 and 2010, and currently, only approximately 75% of children who enroll in primary education complete the full cycle (UNESCO, 2014). The 2011 global youth literacy rate—90%—masks huge regional differences. For example, in Sub-Saharan Africa, 30% of youth between the ages of 15 and 24 are considered illiterate (UNESCO Institute for Statistics [UIS], 2013), and primary school illiteracy levels are even higher (with more than 60% of grade 3 students in many countries unable to read a single word).

During this same period (1990–2015), the donor community has attempted a variety of approaches and interventions to address issues of access, quality, and equity in education. In 1990–2005, USAID focused on four main objectives: (a) increased educational quality; (b) improved access, retention, and graduation rates; (c) improved equity; and (d) the achievement of improvements so that they are sustainable by the countries they support (Chapman & Quijada, 2008). USAID funding tended to support decentralization, the strengthening of local nongovernmental organization (NGO) capacity, and donor coordination. The most widely used strategies included curriculum revision, materials distribution, teacher training, and materials development (Chapman & Quijada, 2008). At the same time, USAID supported a significant budgetary support program for education that focused on key structural reforms, such as prioritizing primary education, shifting expenditures from higher to lower cycles of schooling, reallocating resources from development to recurrent budgets, and ensuring an equitable distribution between rural and urban schools.

Simultaneously, the World Bank provided structural support to education systems, focusing their investments around school-based management, accountability, and improving both the access to and quality of education (Bruns & Luque, 2015; UNESCO, 2014; World Bank, 2011). Another major funder—the UK Department for International Development (DFID)—embraced and invested strategically in similar target areas: (a) access to education, particularly in conflict-affected zones; (b) the quality of teaching

and learning, with an increasing focus on literacy and numeracy; and (c) skills for transitioning to the workforce (DFID, 2010).

Although expectations were high for achieving EFA and MDGs by 2015, most low-income countries did not meet their targets. By the end of 2015, of those countries with data, only 14 of 122 had ensured that 8 of every 10 children were enrolled in school, and only 13 of 90 achieved completion rates near 97% for primary education (UNESCO, 2014). Ten of those countries are member states of either the Organisation for Economic Co-operation and Development (OECD) or the European Union (EU) (UNESCO, 2014). In terms of quality, the same report estimated that nearly 40% of the world's children (some 250 million) were not acquiring basic skills in reading and math (UNESCO, 2014).

Ushering in a new era for global development goals, in late September 2015, the United Nations (UN) General Assembly adopted new Sustainable Development Goals (SDGs) (UN, 2015) to be achieved by 2030. Goal 4 pledges all countries to "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (UN, 2015, p. 104). As the donor community has increasingly narrowed its focus in education to building foundational reading and numeracy skills, as a development community, we face three major challenges:

1. We need large gains, quickly. Education systems in low-income countries must demonstrate improvements of 1 or 2 standard deviations (SDs) in learning outcomes to help nearly 250 million children acquire foundational reading and mathematics skills (Bruns & Luque, 2015). Recent metareviews have identified few interventions that have achieved such large gains, and even those that do are rarely accomplished at scale (Evans & Popova, 2015; McEwan, 2015). Small-scale programs with effect sizes of this magnitude are usually the result of direct, well-supported, and well-designed pedagogical interventions (Conn, 2014; Crouch & DeStefano, 2015). In the United States, improving one student's achievement by one SD increase costs approximately $25,000 based on an annual expenditure of $8,000/student. International assessment measures (e.g., Trends in International Mathematics and Science Study [TIMSS] and Program for International Student Assessment [PISA]) show that middle- and low-income countries typically experience learning improvements of 0.33 SD in a year (Rodriguez, 2004). This result shows that it will take nearly 3 years to close the gap of one SD given current progress and investments.

2. It will not be easy. Comprehensive system reform is complicated and involves multiple stakeholders with competing interests, including teachers (and teachers' unions), elected and appointed officials, government ministries, international organizations, parents, and children and youth. System reform that aims to change how teachers teach (and

children learn) is the most challenging of all. In his review of 100 years of teaching in the United States, Cuban (1993) questioned why so few instructional reforms are actually implemented in classrooms. Understanding context and the enabling environment (or lack thereof) is critical to advancing system and classroom reform.

3. We need to be committed (and strategic) about how to do it. Nothing about the way education systems are set up is accidental. Economic and political elites tend to support the status quo to ensure they and their allies will continue to benefit from the gatekeeping and resource allocations they enjoy. Thus, how can change be driven in such a system? Elected officials must be pushed by stakeholders through communication and advocacy channels: One example of systemic change happening because of strategic education reform support activities is South Africa during the transition from apartheid. USAID technical assistance to the democratic opposition in the form of data analysis and communications brokering helped to raise the technical level of the debate and move the conversation forward on what a rebuilt South African education system would look like (Crouch & Healey, 1997).

The importance of well-targeted and well-timed technical and policy support in implementing major reforms is critical and dependent on focusing on a limited set of core system functions (Gillies, 2010). Indeed, such focus is what leads to the system-level reforms that improve student learning (Crouch & DeStefano, 2015).

However, when programs extend the reach of education interventions to increasing numbers of communities or "go to scale,"[1] what can be considered "reasonable" gains? What interventions will help close the learning gap, with what intensity, for how long, and in what type of environment? The following section provides some context for answering these questions and sets the stage for our proposed framework for analyzing the impact of education development projects.

What Are Reasonable Gains in Basic Education?

Expanding the reach of interventions nationally, or "scaling up" a classroom-based intervention, is not recreating a recipe or gearing up to produce a line of new cars. Scaling up an intervention to thousands or even millions of classrooms in which teachers may lack proper skills is an entirely different type of challenge (Bold, Kimenyi, Mwabu, Ng'ang'a, & Sandefur, 2013; Glennan, Bodilly, Galegher, & Kerr, 2004; Management Systems International [MSI], 2012; Thompson & Wiliam, 2007). The difficulties of moving from pilot educational interventions to scale include the sheer number of classrooms, the complex system in which these classrooms often function (i.e., different environments even within the same country), the separateness of the classrooms, the private nature of teaching, and the

type of pedagogical support these teachers receive on a regular basis (Bold et al., 2013; Glennan et al., 2004; MSI, 2012; Thompson & Wiliam, 2007). For a scaled-up intervention to show an impact, every individual teacher must understand the change and "do it" right. Interventions can work well and be based on solid theories of action, but the enabling environment may not be in place or may be easier to maintain in a more limited area. Interventions also may be difficult to sustain among large numbers of teachers and students (Bold et al., 2013; Glennan et al., 2004; MSI, 2012; Thompson & Wiliam, 2007). The projects can be well designed, but if the teachers cannot deliver the interventions with high quality at the appropriate intensity, the implementation effort will be wasted (Thompson & Wiliam, 2007). Scaling up is also a call for governments to come to grips with the reality of their resource constraints. If they are committed to changes at scale, they will have to find the resources for its support, which means taking those resources from other programs. The risk is that the political will for scale up is great, but that the same will does not extend to securing adequate budgets. The result is that teachers lack required training, materials, and support to effectively implement the reading program in the classroom. This is the challenge of closing the learning gap.

Given these issues, how can we improve learning and reading outcomes in particular? How do we ensure that all children have the foundational skills they need to succeed in life? Donor and development organizations have grappled with the challenge of improving the quality of teaching and learning for decades. The difficulty of moving from a pilot project that demonstrates good effect sizes and local impact to a project that has nationwide impact is challenging to all. Indeed, improving learning by one or two SDs is not an insignificant challenge and it requires time, commitment, and resources.

Table 1.1 summarizes the results from efforts in 14 countries, and the findings are instructive. Program interventions revealed a wide range of investment in terms of intensity and content delivered. Table 1.2 presents the results from the community-based activities delivered as part of Literacy Boost, a program implemented by Save the Children in multiple countries, with effect sizes ranging from 0.03 to 0.85 SD.

As demonstrated by the results presented in Tables 1.1 and 1.2, the effect sizes of interventions vary greatly and, in some cases, are as low as 0.03. The programs that achieved strong effect sizes focused narrowly on reading in the early grades and provided teachers with substantial support through coaching and supervision. However, the dosage, duration, and enabling environment within which the interventions occurred also made a difference. Similar variations were found across the World Bank policy papers, which reported effect sizes from school accountability interventions that varied from no effect to a maximum of 0.91 (Crouch & DeStefano, 2015). The World Bank projects that tended to focus on direct pedagogical interventions and support rather than classroom- or school-based

Table 1.1. Oral Reading Fluency (ORF) Intervention Effect Sizes

Country (Implementer)	Duration (Months)	Design	Grade(s)	Schools	Students	ORF Effect Size	Scale-up Schools	Scale-up Students	ORF Effect Size
Egypt (RTI)	18	QE	1–3	166	9,000	0.55	25,000	4,200,000	n/a
Guatemala (JA)	10	QE	1–3	114	6,000	0.45**	11,668	3,212,544	n/a
Jordan (RTI)	10	RCCT	1–3	43	12,000	0.46	2,651	400,000	n/a
Kenya (RTI)	22	RCCT	1–2	547	56,000	0.24	22,600	5,400,000	n/a
Liberia (RTI; FHI360/RTI)	18	RCCT	1–3	120	16,500	0.80	1,200	70,000	0.41
Mozambique (WE)	10	RCCT	2–3	120	45,000	n/a	538		
Philippines (EDC)	10	RCCT	1–3				2,500	742,500	0.45
Rwanda (EDC)	10	RCCT	1–2	90	31,792	0.55	2,035	1,338,079	0.19
Senegal (EDC)	18	RCCT	1–6	n/a	6,000	0.80			

Source: Calculations by the authors of each of the articles in this special issue.
**Note: effect size is not ORF but reading comprehension.
Legend: See Table 1.2.

Table 1.2. ORF Effect Sizes for Literacy Boost Community Reading Activities

Country (Implementer)	Duration (Months)	Design	Grade(s)	Students	ORF Effect Size
Bangladesh (SC)	22	RCCT	2	66,512	0.21**
Burundi (WV)	24	RCCT	3	2,264	0.04
Ethiopia 1 (WV)	12–15	RCCT	3	901,801	0.85***
Ethiopia 2 (SC)	18	QE	3	42,580	0.27***
Indonesia 1 (SC)	15	QE	3	2,304	0.10
Indonesia 2 (SC)	22	QE	2	5,399	0.05
Indonesia 3 (SC)	9	QE	1	15,271	0.14*
Malawi 1 (SC)	10	QE	3	68,298	0.22*
Malawi 2 (WV)	30	QE	4	25,089	0.34
Philippines (SC)	18	QE	2	593,024	0.03
Rwanda 1 (WV)	30	RCCT	3	26,558	0.58**
Rwanda 2 (SC)	15–25	RCCT	2	113,625	0.14***

Source: Dowd et al., Article 2 in this issue.
Notes for Tables 1.1 and 1.2: EDC = Education Development Center, JA = Juárez and Associates, RTI = Research Triangle Institute, SC = Save the Children, WE = World Education, WV = World Vision (Partner with Save the Children); RCCT = Randomized-cluster controlled trial; QE = Quasi-experimental.

management tended to achieve higher effect sizes and, thus, exert a larger influence on learning outcomes. Nonetheless, the question remains: As countries scale up interventions and institutionalize the interventions into local systems, can similar targets and effect sizes be achieved? What explains the wide ranges of effect sizes obtained with the same interventions?

Going to Scale: A Framework for Understanding the Impact of Educational Interventions

Learning takes place in the classroom because of the direct interactions between teachers and their students. One could argue that if we, as a community, are going to affect learning outcomes positively, we must intervene in this "black box" of daily classroom instruction (Black & Wiliam, 1998; Fullan, 2007; Hawley, 2007). Only when such an intervention has been successfully accomplished can we understand how to take it to scale.

In 2001, the Basic Education Coalition (BEC) was established to advance the global commitments made at the 2000 World Education Forum in Dakar, Senegal. Composed of members from 24 organizations, the BEC brings together technical expertise and experience in designing, implementing, and measuring the impacts of educational interventions globally. In 2014, the BEC Working Group for Monitoring and Evaluation wanted to understand (a) to what degree reading programs were achieving the ambitious targets envisaged for student reading outcomes; (b) to what extent these would translate at scale; and (c) whether different combinations of

interventions had greater or lesser effects on improving reading outcomes. Using data from several USAID-funded programs, the working group devised a framework for analyzing the effect sizes of these interventions to inform realistic targets, time frames, and resource envelopes and how best to improve reading outcomes, particularly at scale.

This framework, which incorporates the concepts of dosage, duration, and enabling environment as key to understanding program effectiveness, is reflected in varying degrees in the articles in this thematic journal issue. The hope is that the results presented can both inform and guide donors and implementing organizations to continue to improve learning outcomes and entice researchers to collaborate to better inform and enhance ongoing efforts.

The Dosage of Interventions Makes a Difference. Dosage includes the frequency and intensity (or strength) of an intervention. It is critical to implementation and is tightly intertwined with other important factors, such as fidelity, content, quality, and exposure (Daro, 2010; Downer & Yazejian, 2013; Paulsell, Boller, Hallegren, & Esposito, 2010). We distinguish two levels of *dosage*: the implementation level and the intervention level (Wasik, Mattera, Lloyd, & Boller, 2013). Implementation dosage refers to strength of the preparation that stakeholders (e.g., coaches, teachers) receive to deliver an intervention (i.e., pedagogical method, new materials). Examples of implementation dosage in education include the amount of training coaches and teachers receive in preparation to deliver a new pedagogical method, the amount of time coaches spend working with teachers on the delivery of a scripted lesson, or the amount of time teachers spend receiving training on the use of new materials (Wasik et al., 2013). The intervention dosage refers to the frequency with which an intervention is provided to a particular stakeholder (e.g., students, teachers, or community members). For example, is an intervention, such as a training workshop, provided once during the life of a project or on a monthly basis? How many hours per year does a student receive reading instruction? Understanding the intervention dosage—or frequency of delivery that is necessary to achieve sustainable change—is important, affecting not only project outcomes but also costs, staffing, replication, and the expansion efforts that follow.

The existing research on dosage clearly states that one dose of an intervention is usually not enough (Boller et al., 2004; Joyce & Showers, 1980; Winton & McCollum, 2008). For example, Boller et al. (2004) and others (Raikes et al., 2006; Winton & McCollum, 2008) showed that holding 1-day workshops for teachers does not provide the necessary depth of understanding and is usually insufficient (i.e., too infrequent, thus not reinforcing learning) to affect teacher learning or change and improve long-term classroom practices. Professional development interventions must be delivered more intensively and, usually, with a longer duration or higher frequency to make a difference (Halle et al., 2010). However, there is little

evidence that tells us what the right dosage of teacher training might be to move student learning upward.

How the dosage is organized and administered also matters in understanding dosage. Several rigorous studies on children's learning in kindergarten highlight the extent to which dosage can be contextually specific. The results have consistently shown that children who attend full-day kindergarten programs develop twice the literacy skills and stronger numeracy skills than those who attend half-day programs (Cooper, Allen, Patall, & Dent, 2010; Lee, Burkam, Ready, Honigman, & Meisels, 2006). However, research has also shown that this is not true for children attending subpar full-day programs; thus, context and quality matter (Magnuson, Meyers, Ruhm, & Waldfogel, 2004; Robin, Frede, & Barnett, 2006).

Finally, regarding the concept of *dosage*, it is important to understand whether it is necessary to attain a certain dosage level—that is, a threshold—to impact change. Developing such a threshold (or a range) can help organizations monitor and measure change based on the dosage or frequency of the intervention and then adapt as needed when attempting to replicate or scale the intervention (Wasik et al., 2013).

The program interventions involved in the 14 countries' experiences documented in this special issue revealed a wide range of programmatic investments in terms of dosage and supported system-level interventions, such as policy reforms and capacity development (discussed later under "The Enabling Environment"). The degree to which each of these components was emphasized varied in important ways. For example, whereas programs nearly universally relied on teacher training and support delivered through school sites (including periodic coaching visits), some programs provided training through government employees (e.g., district supervisors or central ministry staff through cascade training). In contrast, others relied on project-contracted staff (i.e., coaches hired during project implementation). In Kenya, experimentation with the number of schools assigned to each government coach (10 vs. 15) resulted in modestly better outcomes for the smaller school-to-coach ratio, although the effect did not justify the additional staffing cost. In Liberia, the coach-to-school ratio increased from 4 to 12 between pilot and scale-up; unsurprisingly, the results declined correspondingly (Gove, Korda Poole, & Piper, Article 5).

Materials development also varied, with some efforts relying on existing materials (reprinted with project funds) and others supporting the creation of new supplementary reading materials (e.g., big books and leveled readers), student workbooks, textbooks, and teacher guides. Materials development is complex, is expensive, and requires high levels of expertise that are not customarily housed within the NGO community. Therefore, to fill the gaps in capacity, implementers often partner with publishers or directly contracted illustrators, designers, linguists, and curriculum experts to generate content. Programs in Egypt, Kenya, and Liberia included support for the development of project-provided textbooks distributed to students

at a 1:1 ratio; following development and testing, the materials were made freely available to schools outside the project domain (Gove, Brombacher, & Ward-Brent, Article 6; Gove et al., Article 5). In Rwanda, technical assistance supported the articulation of standards and technical specifications for leveled books in Kinyarwanda (Cristina & Vinogradova, Article 3). In Mozambique, teachers received two copies of each of 18 read-aloud titles, and each student received 18 low-cost decodable books (USD $0.02 each) to keep and use at home. Subsequently, in the impact evaluation, 80% of students in program schools were observed with books, compared to only 10% of children in comparison schools (Burchfield, Hua, Noyes, & van de Waal, Article 7). Significantly, USAID has increased its efforts to ensure that all materials developed with USAID funding are licensed under the auspices of Creative Commons, with freely available rights for adaptation and distribution.

Although all programs included activities that involved communication with and engagement of families and communities, Save the Children's programming relied particularly heavily on out-of-school investments to improve results in foundational reading skills. Programming for learners included attending "reading buddy" meetings, borrowing books from a book bank, attending a reading camp, participating in a "make-and-take" activity to create reading materials to take home from the reading camp, and participating in a read-a-thon. Increased participation in these activities had a positive and significant effect on basic reading skills in all of the project sites (12 sites in 7 countries), with effect sizes between 0.2 and 1.13 SDs (Dowd et al., Article 2 in this issue). Home literacy materials and habits had less of an impact (the highest effect size was 0.3, with most in the 0.1 range), and the Philippines even reported a negative effect. More research is needed to determine why home practices and resources did not have more of an impact on outcomes.
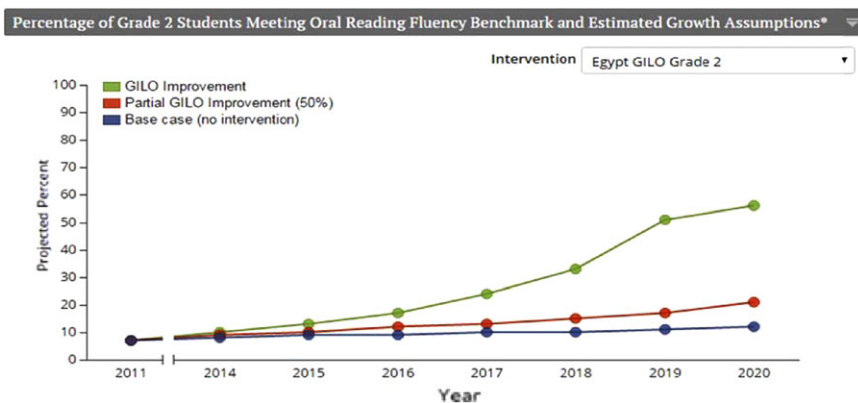
The Duration of an Intervention Also Matters. Duration refers to the length of time that interventions are provided to participants (i.e., students, teachers, and school directors). Duration can be measured by maintaining records of how often a particular intervention is conducted (e.g., how often peer coaches meet with teachers over the life of a project). Ensuring that the content delivery and support are of high quality is critical because a subpar program or training may have no effect on the delivery of pedagogical interventions and, over time, could even worsen the classroom delivery. As shown by the articles in this issue, when support for interventions is provided for a longer time, the chance that those interventions will take hold and demonstrate positive results increases.

The programs presented in this special issue ranged in duration from 10 months (approximately one school year) to 30 months. In several cases, the duration (and consequently, the dosage) was affected by teacher strikes, civil unrest, delays in project implementation and rollout, and/or teacher and student absences. The duration of programming frequently

differed from the time frame between assessments, suggesting that additional coordination among implementers, funders, and external evaluators is needed. For most USAID-funded programs, impact evaluation designs were cross-sectional in nature; for example, two different cohorts of grade 2 student performance were compared at the end of the school year. Longitudinal studies specifically following students who had participated in early grade reading programs from grade 1 through grade 3 or 4 were not available. Although there does not appear to be a relationship between duration and reading outcomes, as measured by effect size, arguably none of the programs—or more specifically their interventions—had been implemented long enough on a routine basis to feasibly show the effect of duration.

The Enabling Environment. An *enabling environment* is a set of interrelated conditions (political, institutional, technical, and cultural) that affect the capacity of actors (teachers, parents, communities, donors, and governments) to engage in development processes in a sustained and effective manner. The challenge often faced in implementing education reform and taking it to scale is that the situation existing in a country is not accidental. Instead, well-entrenched interest groups often exist that want to ensure that the institutional arrangements from which they directly benefit do not change (Conn, 2014; DeStefano & Crouch, 2006). Proposed school-level reforms or pilot programs are frequently seen as affecting only a small part of the system, and thus, these groups do not feel threatened. However, when changes (such as policies or the scale-up of interventions) begin to shift resource allocations, interest groups that may not benefit from the change respond, often forcefully (DeStefano & Crouch, 2006). Moreover, in a government that is subject to considerable political pressures, stakeholders and government officials may respond differently than they would to an NGO-led experiment, leading to weaker implementation during the scale-up process (Conn, 2014). The successful reform and scale-up of interventions requires strong leadership able to cope with the political battles and trade-offs that must be fought or negotiated. In some cases, positive, enabling environments may exist from which projects can benefit; however, many projects must create their own enabling environments because such contexts can directly affect an intervention's impact on its targeted population.

As noted previously, the USAID-funded programs were conducted in partnership with host country governments through bilateral agreements with technical assistance and implementation support provided by U.S.-based NGOs. The degree to which programs were implemented by government staff depended on the capacity and availability of human resources: In some cases, efforts were largely government led, with assistance and support from implementing partners. In other cases, government partners played more of an oversight and review role. Finally, sometimes, implementers worked closely with governments to design and implement programs at

Figure 1.1. Sample ORF Projection

Source: www.earlygradereadingbarometer.org.
Note: GILO is the acronym for the USAID-funded Girls' Improved Learning Outcomes project in Egypt.

a cost that the governments could afford, which is a critical factor for ensuring eventual scale-up.

One key avenue for engagement was shown to be the development and articulation of standards and benchmarks for reading performance. In several countries, researchers worked with government counterparts to analyze the assessment results and generate projections for policy makers to discuss. Based on this analysis, some countries established benchmarks and standards to guide both project implementation and teachers in the classroom. Informed by this process, USAID developed and launched a data visualization tool for interpreting and depicting early grade reading assessment results, which is available at www.earlygradereadingbarometer.org. Figure 1.1 depicts a growth projection that relies on both national survey data and project impact evaluation results. Using this tool, policy makers can see the potential impacts of interventions on the percentage of students meeting grade-level expectations and the effect of setting lower or higher expectations for performance.

Critical Gaps and Opportunities

Sharing the results of these early reading interventions across 14 countries should prove useful for informing future programming for USAID and other donors. Convergence appears to exist on the content of programming choices and there is potential for knowledge sharing across these interventions, particularly regarding the design, dosage, and duration of teacher support mechanisms and the development of materials (two critical and cost-intensive components). For example, all of the programs used

ongoing coaching rather than large-scale, one-time cascade training to support teacher professional development (i.e., extended duration and dosage of the intervention). Performance and impact evaluations should seek to answer questions relating to the dosage and duration of the coaching models under different conditions. Currently, most programs aim to deliver the same level of training to all participants. Instead, identifying those teachers who need the most support and providing them with an extra dose of coaching, at least initially, may be a more effective strategy. Similarly, there is a need for knowledge sharing related to materials development and distribution regarding not just the final product but also lessons learned about the sharing process. If the NGO community rather than the private sector plans to continue to be in the publishing business, how can we best share lessons relating to conceptualizing, designing, producing, and distributing materials?

One key limitation affecting much of this work is that interventions are typically compared to status quo activities, which is, in most cases, an unfair comparison. Kenya, Liberia, and Mozambique were the only study designs that included multiple treatment conditions (Burchfield et al., Article 7; Gove et al., Article 5). Donors and implementers should endeavor to build more comparisons among different approaches within their implementation designs, testing multiple hypotheses to answer key policy and practice questions. Indeed, more work is needed to understand the nuances of implementation (not just treatment vs. control), which will require well-designed monitoring systems to determine who gets what and when. However, the results should prove very useful in finalizing a design or approach for eventual scale-up (Gove et al., Article 5).

Additional research is needed to better understand the program characteristics that are likely to support scaling innovations from "small to big" and "big to better." This analysis did not dig deeply into the particular political and policy conditions that are likely to influence eventual scale-up by host country governments. For example, an optimal time, such as just after a change in government or ministry staff, may exist for the initiation of improvement and reform efforts. Global trends can also influence the likelihood of scale-up. For example, with the increased availability of data on learning outcomes and the adoption of the SDGs, an increased sense of urgency currently exists among both donors and host governments. This urgency may be leading to a rush to rollout before content has been adequately validated and trialed; in some cases, implementers are being asked to design and deploy reading materials in as few as 6 months. We all recognize the need for large, rapid gains, but we risk substantial public failure if we do not take adequate time to validate the approach and content.

What is the right size (and location) of an initial validation/pilot effort? The scale of these efforts ranged from a handful to several hundred schools, and the interventions in Egypt, Kenya, and Rwanda were subsequently expanded to the national scale. Other programs, such as those in Guatemala,

Liberia, and Mozambique, targeted regions that were pre-selected by USAID for poverty or political reasons (Burchfield et al., Article 7; Gove, Korda Poole, & Piper, Article 5; Rubio et al., Article 4). Should programs be trialed in "typical" districts? Or should demonstration efforts instead be conducted in contexts where they are more likely to succeed? Would the dosage and duration of an intervention differ in typical areas versus those that were atypical? Research would suggest that differences would exist, but more work is needed to understand the nuances of the differences to assist us in improving reading outcomes at scale.

One additional methods issue is that all of the programs (and governments) would benefit from improvements in the collection and use of monitoring and implementation data. Data dashboards with accurate and timely information on the numbers of schools and teachers and on the quantities of other resources remain rare, and many countries even lack accurate data on the numbers of schools and teachers within their systems. A better and more nuanced understanding of the factors that increase implementation fidelity is critically needed if we are to continue to make progress. How can we support the development and use of improved, wider scale, and lower cost systems for monitoring that help to inform and improve evaluation and research results? Too often, evaluation is seen as external rather than integral to implementation and improvement efforts, whereas monitoring data are not used in evaluations; instead, the two should be integral to each other.

Last, it is important to note that each of these gaps actually represents an opportunity. Indeed, we have never had a better chance to leverage new developments in data capture, research, implementation design, and evaluation to inform improvement at scale. But building the evidence base is not without its challenges. Despite the availability of research-friendly implementation models such as phased or delayed-entry cohorts (where one group serves as a temporary control group to provide a counterfactual), there is considerable resistance to the idea of research shaping implementation design. There is a need for more focus on formative and impact evaluations that not only can inform implementers on the effects of the dosage and duration of interventions but also can test out and learn from the differing combinations of interventions in differing contexts. We hope that this issue provides insight into how interventions come together to improve reading outcomes—and how the experience of each of these organizations informs our global knowledge of what it takes to go to scale and help every child learn.

Note

1. "Go to scale" or "scale up" is defined as a process of extending community-level change or increasing in the number of participating communities in a given reform or activity.

References

Birdsall, N., Levine, R., & Ibrahim, A. (2005). *Towards universal primary education: Investments, incentives, and institutions*. Sterling, VA: Stylus Publishing.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London, UK: Department of Education and Professional Studies, Kings College.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2013). *Scaling up what works: Experimental evidence on external validity in Kenyan education* (Center for Global Development Working Paper 321). Oxford, UK: University of Oxford.

Boller, K., Uhl, S., Crozier, L., James-Burdumy, S., Novak, T. J., Johnson, A., & Vogel, C. (2004). *Using television as a teaching tool: The impacts of "ready to learn" workshops on parents, educators, and the children in their care*. Princeton, NJ: Mathematica Policy Research.

Bruns, B., & Luque, J. (2015). *Great teachers: How to raise student learning in Latin America and the Caribbean*. Washington, DC: International Bank for Reconstruction and Development/The World Bank.

Chapman, D. W., & Quijada, J. J. (2008). An analysis of U.S. assistance to basic education in the developing world, 1990–2005. *International Journal of Educational Development*, *29*(3), 268–280.

Conn, K. (2014). *Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations*. New York, NY: Columbia University.

Cooper, H., Allen, A. B., Patall, E. A., & Dent, A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research*, *80*(1), 34–70.

Crouch, L., & DeStefano, J. (2015). *A practical approach to in-country systems research*. Paper prepared for the Research on Improving Systems of Education (RISE) Program Launch Conference, Washington, DC, June 18–19, 2015. Retrieved from http://www.riseprogramme.org/sites/www.rise.ox.ac.uk/files/14_Crouch-DeStefano.pdf

Crouch, L., & Healey, F. H. (1997). *Education reform support. Vol. 1: Overview and bibliography* (SD Publication Series, Paper No. 47; Advancing Basic Education and Literacy Technical Paper No. 1). Washington, DC: Office of Sustainable Development, Bureau for Africa, USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/PNACA717.pdf

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890–1990*. New York, NY: Teachers College Press.

Cuban, L. (1998). How schools change reforms: Redefining reform success and failure. *Teachers College Record*, *99*(3), 453–477.

Daro, D. (2010). *Replicating evidence-based home visiting models: A framework for assessing fidelity*. Princeton, NJ: Mathematica Policy Research.

DeStefano, J., & Crouch, L. (2006). *Education reform support today*. Prepared for USAID under the Education Quality Improvement Program (EQUIP2), Cooperative Agreement No. GDG-A-00-03-00008-00. Washington, DC: Academy for Educational Development (AED). Washington, DC: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/pnadq913.pdf

Department for International Development. (2010). *Learning for all: DFID's education strategy 2010–2015*. London, UK: Author. Retrieved from http://www.ungei.org/resources/files/educ-strat.pdf

Downer, J., & Yazejian, N. (2013). *Measuring the quality and quantity of implementation in early childhood interventions* (OPRE Research Brief OPRE 2013-12). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Evans, D. K., & Popova, A. (2015). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews (World Bank Policy Research Working Paper 7203). Washington, DC: World Bank.

Fullan, M. (2007). *The new meaning of educational change*. London, UK: Routledge/Taylor and Francis Group.

Gillies, J. (2010). *The power of persistence: Education system reform and aid effectiveness*. Washington, DC: USAID. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.5957&rep=rep1&type=pdf

Glennan, T. K., Bodilly, S. J., Galegher, J., & Kerr, K. A. (2004). *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG248.html

Halle, T., Zaslow, M., Tout, K., Starr, R., Wessel, J., & McSwiggan, M. (2010). Beyond how much: What we are learning about structuring effective early childhood professional development. In S. B. Neuman & M. Kamil (Eds.), *Preparing teachers for the early childhood classroom: Proven models and key principles* (pp. 175–188). Baltimore, MD: Brookes Publishing.

Hanushek, E., & Woessman, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal for Economic Growth*, *17*(4), 267–321.

Hawley, W. (2007). The keys to effective schools: Educational reform as continuous improvement. Thousand Oaks, CA: Corwin Press.

Joyce, B. R., & Showers, B. (1980). *Power in staff development through research on training*. Alexandria, VA: Association for Supervision and Curriculum Development.

Lee, V., Burkam, D. T., Ready, D. D., Honigman. J., & Meisels. S. J. (2006). Full-day versus half-day kindergarten: In which program do children learn more? *American Journal of Education*, *112*(2), 163–208.

Lockheed, M., & Verspoor, A. (1992). Improving primary education in developing countries. London, UK: Oxford University Press.

Magnuson, K., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool and school readiness. *American Education Research Journal*, *41*(1), 115–157.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353–394.

Management Systems International. (2012). *Scaling up—From vision to large-scale change: A management framework for practitioners* (2nd ed.). Washington, DC: Author. Retrieved from http://www.msiworldwide.com/wp-content/uploads/MSI-Scaling-Up-Framework-2nd-Edition.pdf

Paulsell, D., Boller, K., Hallgren, K., & Mraz Esposito, A. (2010). Assessing home visit quality: Dosage, content, and relationships. *Zero to Three*, *30*(6), 16–21.

Raikes, H. H., Torquaiti, J. C., Hegland, S., Raikes, H. A., Scott, J., & Scott, S. (2006). Studying the culture of quality early education and care: A cumulative approach to measuring characteristics of the workforce and relations to quality in four Midwestern states. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 111–136). Baltimore, MD: Brookes Publishing.

Robin, K. B., Frede, E. C., & Barnett, W. S. (2006). *Is more better? The effects of full-day vs. half-day preschool on early school achievement*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, *17*(1), 1–24.

Thompson, M., & Wiliam, D. (2007). *Tight but loose: A conceptual framework for scaling up school reforms*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Chicago, IL, April 9–13, 2007.

United Nations. (2015). *Sustainable development goals*. Retrieved from http://www.un.org/sustainabledevelopment/sustainable-development-goals/

UNESCO. (2000). *World Education Forum final report*. Dakar, Senegal: Author.

UNESCO. (2014). *Teaching and learning: Achieving quality for all. Education for All Global Monitoring Report (2013/14)*. Paris, France: Author. Retrieved from www.uis.unesco.org/Library/Documents/gmr-2013-14-teaching-and-learning-education-for-all-2014-en.pdf

UNESCO. (2015). *Development goals in an era of demographic change. Global Monitoring Report (2015/16)*. Paris, France: Author. Retrieved from http://pubdocs.worldbank.org/pubdocs/publicdoc/2015/10/503001444058224597/Global-Monitoring-Report-2015.pdf

UNESCO Institute for Statistics. (2013). *Adult and youth literacy*. Fact sheet no. 26. Retrieved from http://www.uis.unesco.org/literacy/Documents/fs26-2013-literacy-en.pdf

United States Agency for International Development (USAID). (2011). *Education: Opportunity through learning. USAID Education Strategy 2011–2015*. Washington, DC: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/PDACQ946.pdf

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated* (OPRE Research Brief OPRE 2013-15). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Winton, P. J., & McCollum, J. (2008). Preparing and supporting high quality early childhood practitioners: Issues and evidence. In P. J. Winton, J. A. McCollum, & C. Catlett (Eds.), *Preparing and supporting effective practitioners: Evidence and applications in early childhood and early intervention* (pp. 1–12). Washington, DC: Zero to Three Press.

World Bank. (2003). *World development report 2004: Making services work for poor people*. Washington, DC: World Bank.

World Bank. (2011). *Improving the odds of achieving the MDGs: Heterogeneity, gaps and challenges*. Washington, DC: International Bank for Reconstruction and Development/World Bank. Retrieved from https://www.imf.org/external/pubs/ft/gmr/2011/eng/gmr.pdf

AUDREY-MARIE MOORE, PhD, is a senior researcher at Mathematica Policy Research.

AMBER GOVE, PhD, is director of research in international education at RTI International.

KAREN TIETJEN, PhD, is technical director, Education for Development, at Creative Associates International.

All three work on literacy, learning, and evaluation projects in low- and middle-income countries.