
Classroom-up Policy Change: early reading and math assessments at work

AMBER GOVE, SAMIR HABIB,
BENJAMIN PIPER & WENDI RALAINGITA
RTI International, Triangle Park, NC, USA

ABSTRACT This article reviews the development of the Early Grade Reading and Mathematics Assessments (EGRA and EGMA), which are locally tailored, timely assessments designed to directly inform policy and instruction for learning improvement, particularly for countries on the lower end of the income spectrum. The history of the design and implementation of the tools, as well as case studies of their use in Egypt and Kenya, are a useful counterbalance to the experience of the more traditional international large-scale assessments (ILSAs) documented in this special issue – in particular for understanding the needs of countries struggling to transform ‘education for all’ into ‘learning for all’.

Measuring Learning from the Bottom Up

The objectives of Education For All (EFA) and Millennium Development Goals (MDGs), while effective at advancing access to and completion of primary school, have left much to be desired about the measurement and improvement of learning within those schools. As argued elsewhere (Chabbott, 2003; Crouch & Gove, 2011; Wagner et al, 2012), the absence of clear, measurable goals related to learning outcomes meant that governments and the international community alike focused nearly all of their attention on the access goals throughout the 1990s and early 2000s. Midway through the ‘oughts’, the international community began to recognize in a more vocal fashion that learning had somehow slipped off the agenda, particularly in low-income countries.

The World Bank’s Deon Filmer and colleagues first called for a Millennium Learning Goal (MLG) in 2006, citing evidence that many countries which had reached the MDG for primary completion were failing to ensure adequate learning (Filmer et al, 2006). Focusing on learning achievement for all, the authors argue, would eliminate the false dichotomy between access and quality, as reaching an MLG would depend on both. Around the same time, portfolio reviews of the key donors in education revealed that decades of investment had yielded little in the way of measured learning outcomes, leaving key players in the global education space with no evidence of actual learning gains (even if any had occurred). The World Bank’s Independent Evaluation Group reviewed 30 years of education loans and grants (Nielsen, 2006), finding that few, if any, had adequately addressed issues of measurement and improvement of learning outcomes. Chapman and Quijada’s (2008) review of 28 US Agency for International Development (USAID) projects whose stated goal included raising educational quality revealed just five projects with meaningful increases in learning that could reasonably be attributed to the project.

In this context, a number of players began to recognize the lack of learning assessments to meet the needs of low-income countries. Expansion of universal primary education, coupled with stagnating primary completion rates, was of increasing concern. Development professionals and donors alike called for locally tailored, timely assessments that could directly inform policy and instruction for learning improvement, particularly for countries on the lower end of the income spectrum (Wagner, 2003; Abadzi, 2006; Center for Global Development, 2006; Chabbott, 2006;

Nielsen, 2006; Chapman & Quijada, 2008). This article reviews the development of two such tools, designed to assess reading and mathematics in the early grades, and their implementation, beginning at the classroom level, in Egypt and Kenya. The history of the design and implementation of the Early Grade Reading and Mathematics Assessments (EGRA and EGMA), coupled with their implementation to improve learning at the classroom level, is a useful counterbalance to the experience of the more traditional international large-scale assessments (ILSAs) documented in this special issue – in particular for understanding the needs of countries struggling to transform ‘education for all’ into ‘learning for all’.

The Early Grade Reading Assessment (EGRA)

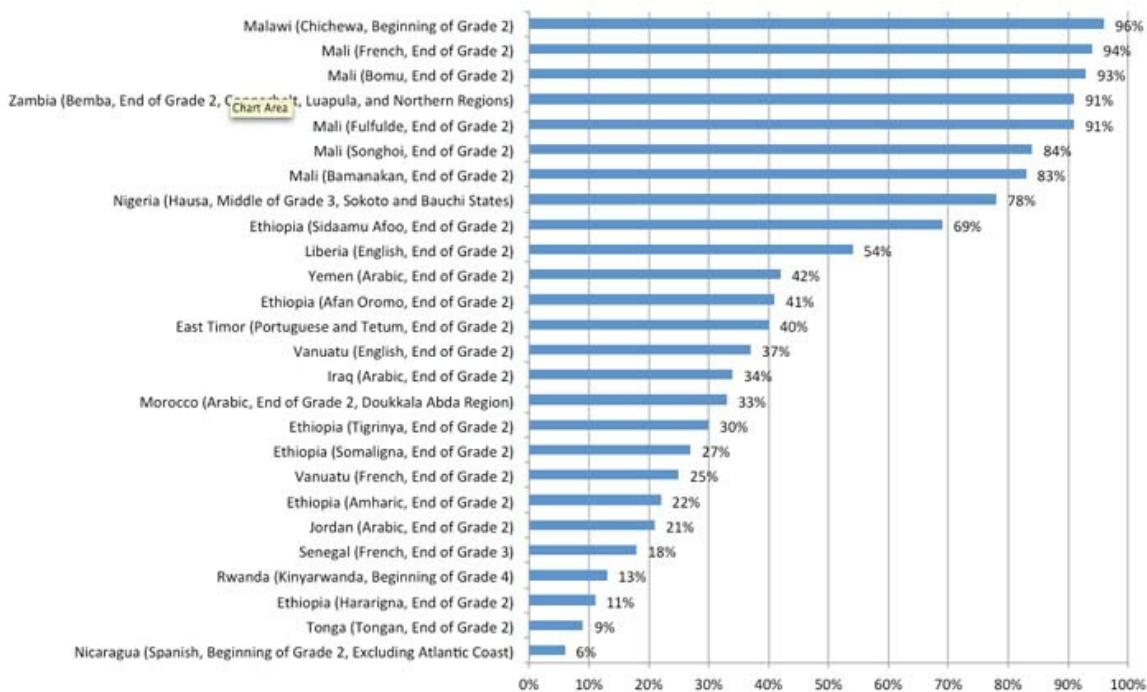
Development of the EGRA, initiated in 2006 by RTI International with support from USAID and the World Bank, was motivated by three key gaps in the landscape of the traditional ILSAs. The first was the absence of data on learner performance in the early grades of primary school, a critical phase of student development. The ILSAs available at the time focused on students enrolled in fourth grade (Progress in International Reading Literacy Study [PIRLS]), fourth and eighth grades (Trends in International Mathematics and Science Study [TIMSS]), or 15-year-olds (Program for International Student Assessment [PISA]). Second, the ILSAs did not seem to be adequately serving those countries most in need of information about how to improve learning. To date, relatively few low- or lower-middle-income countries have participated in PISA, PIRLS or TIMSS (Crouch & Gove, 2011). Third, ILSAs’ 3- to 4-year time cycles, while effective in allowing time for consensus and comparability and useful for plotting long-term trajectories, have not been particularly helpful to developing countries struggling to identify and correct what was suspected to be a growing learning crisis.

EGRA began with development of a conceptual framework to close these information and policy gaps. Following a review of research on early literacy learning and assessment, RTI International developed a draft testing protocol and convened a panel of cognitive scientists, early-grade reading instruction experts, research methodologists and assessment experts to review the proposed key components of the instrument (RTI International, 2009). The panel provided suggestions for improving the draft protocol in 2006; pilots were conducted in Senegal (French and Wolof) and the Gambia (English) in 2007, followed by further demonstration efforts in Nicaragua, South Africa and Kenya.

In 2009, the World Bank supported development of a toolkit to document the research foundations of the protocol and provide guidance to test developers and policy makers for its use (RTI International, 2009). While EGRA and its supporting documentation provide a common framework, the subtasks of the assessment have varied in accordance with the needs and interests of each implementing country or context.[1] Supporting documentation, training and reports have helped take EGRA to a global scale. The *EGRA Toolkit* (RTI International, 2009) and subsequent *Guidance Notes for Planning and Implementing EGRA* (RTI International & IRC, 2011) have informed development of early reading assessments in more than 60 countries and nearly 100 languages (as of January 2013). While USAID, as a key funder, has provided strong support, more than half of the countries that have conducted early reading assessments have done so with financing from other donors or through direct government or nongovernmental organization (NGO) efforts (RTI International, 2013). Efforts range from multi-country assessments conducted by the World Bank to those of small NGOs supporting a handful of schools. Many users download the instruments directly from the EdData II website <http://www.eddataglobal.org> (e.g. applying Hausa instruments developed in Nigeria to another sample of schools in the same country), while others use the resources and documentation provided to develop their own assessments (e.g. Juarez and Associates’ work in Tajikistan and the Kyrgyz Republic). Many of these instruments and resulting reports have been posted online for others to build on and draw from, a practice that is slow but growing as specialists come to realize the benefits of collaboration and open-source/creative commons-type approaches to development work. An online community of practice, residing at <http://www.earlygradelearning.org>, supports the continued expansion and use of the assessment to inform policy and instruction. A key product of this collaboration was the development of a monograph titled *Early Reading: igniting Education For All* (Gove & Cvelich, 2011), which reported

Classroom-Up Policy Change

on the dire results collected to date. Figure 1 is a sample snapshot of zero scores, or the proportion of grade 2 students unable to read a single word of a grade-level paragraph.



Note: A score of zero is recorded for those students who are unable to correctly read any words on the first line of a one-minute oral reading fluency assessment using a graded reading passage. Results are based on national samples of pupils unless otherwise noted. For complete reports and further documentation, please see <http://www.eddataglobal.org>

Figure 1. Percentage of zero scores for EGRA Oral Reading Fluency Subtask, 2008-2012

Source: Assessments conducted by Ministries of Education, in collaboration with the World Bank (East Timor, Tonga and Vanuatu) or RTI International (all others). Compilation and calculations by RTI International.

Building on the community of practice, in 2010 RTI International began developing an open-source electronic data capture tool, Tangerine. The software is designed for use on mobile computers, including netbooks, tablet computers and smartphones. While EGRA benefited from post-assessment electronic data entry from its inception, the potential advantages of field-based data capture include improved data quality, reduced time, and cost savings. Tangerine is currently being used and improved upon by organizations such as Room to Read, Save the Children and Catholic Relief Services.

Expansion of EGRA has not been without its drawbacks. One unintended consequence of an open-source approach is that content is available for all to use in whatever way they wish. Despite guidance to the contrary, it is possible that EGRA has been used inappropriately (e.g. as an assessment tool for impact evaluations that have no reading improvement component). However, the authors believe that the benefits of sharing outweigh the costs. Critics have argued that EGRA has contributed to a ‘reading mania’ on the part of development agencies (Klees, 2012) or to a narrowing of the curriculum, prioritizing one set of results and approaches over another (Hoffman, 2012). Yet, compared with the access-only agenda that persisted through the 2000s, the renewed focus on learning is a welcome change.

The Early Grade Mathematics Assessment (EGMA)

While early-grade reading is a critical pillar of education, it is insufficient. Students need mathematics in their everyday and professional lives. Mathematics underpins all the sciences, and is foundational to both science and technology. For both developed and developing countries,

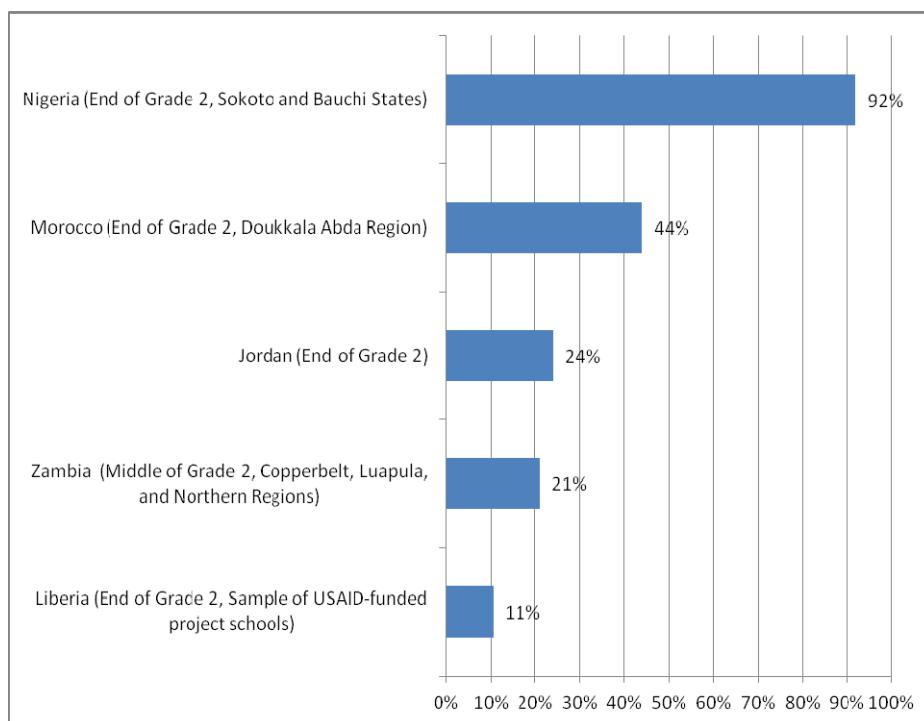
science, mathematics and technology are central to national socioeconomic development. As with reading, proficiency in mathematics starts with mastering the most fundamental competencies in the early grades. Given the need to address early-grade mathematics, and the dearth of information regarding learning outcomes at this level, the Early Grade Mathematics Assessment instrument was designed to provide information about basic competencies which should be mastered in the very early grades, and without which students will struggle, or potentially drop out. Development of the instrument, again supported by USAID and carried out by RTI International, began with a systematic and extensive review of literature and research on early-grade mathematics learning and assessment, and with consultation with leading researchers in the field. The conceptual framework underlying the instrument is grounded in extensive research over the past 60 years (e.g. Chard et al, 2005; Baroody et al, 2006; Clements & Sarama, 2007). To construct the EGMA protocol, developers systematically sampled early numeracy skills, particularly those underlying number sense. These abilities and skills are key in the progression toward more advanced problem-solving and the acquisition of advanced mathematics skills (Baroody et al, 2006; Clements & Sarama, 2007; Foegen et al, 2007).

Several key characteristics define the instrument: it is an individual, oral assessment, which allows for assessing certain skills, such as simple number identification, that cannot be tapped through a written, group-administered test. An oral assessment also allows for analysis of mathematics skills without confounding factors, such as the possibility of low reading skills, or underdeveloped writing skills that might make it difficult to separate out students' ability in mathematics. The instrument is fairly quick to administer, averaging 15-20 minutes per child, and is easy to score. These characteristics facilitate test administration in developing-country contexts, where access to highly trained, experienced assessors may be limited. Additionally to any statistical validity, the face validity is important as practitioners and policy makers at various levels must be able to easily recognize how the skills tested relate to mathematics and mathematics learning.

In addition, criteria were established for the subtasks to be included in the instrument, to provide stakeholders – from ministries of education to aid agencies to local education officials – with information to improve teacher education and support, curriculum development, and implementation. The skills tested should represent what the curricula in both developing and developed countries have determined should be acquired in early grades. That is, while early-grade mathematics curricula vary from country to country, the skills included in the EGMA instrument should be universal, or nearly so. Also, the subtasks should reflect skills that are most predictive of future performance according to available research and scientific advice. Taken together, they represent a progression of skills that lead toward proficiency in mathematics. Looking across subtasks, they target both conceptual understanding and procedural fluency, both of which are essential to mastering mathematics. Since the ultimate goal is to assess and track early-grade mathematics learning outcomes to inform policy or instructional change, it is important that the subtasks measure skills that can be improved through better instruction.

The initial EGMA instrument was developed in 2008–2009. Including 10 subtasks, it was reviewed by a panel of experts in January 2009, piloted in Kenya in June 2009, and revised. Over the next year and a half, the assessment was adapted and tested in additional countries, and a 'core' instrument was established, retaining those subtasks with the best performance, including clearly interpretable results.[2]

As of January 2013, the complete version of EGMA has been used in at least nine countries, and an abbreviated version in four additional countries. While results vary, general trends indicate that students are not meeting expectations laid out in local curricula. While students in most countries tested appear to be mastering basic operations, they are doing so later than expected, which is likely to impede their progress in mathematics as they move on to more complicated and more abstract mathematics in higher grades. Thus, while the 'news' for early-grade mathematics may not be as dire as it has been in some countries for reading, results thus far indicate a tendency for students to master basic competencies later than expected, if at all – an unfortunate situation for countries looking to build expertise in science, mathematics and technology. Figure 2 provides EGMA results from a selection of countries, showing the percentage of students with zero correct responses in the subtraction subtask, which consists of 20 single-digit subtraction problems.



Note: A score of zero is recorded for those students who are unable to correctly respond to at least one of 20 single-digit subtraction problems within one minute. Results are based on national samples of pupils unless otherwise noted. For complete reports and further documentation, please see <http://www.eddataglobal.org>

Figure 2. Percentage of zero scores for EGMA subtraction subtask, 2011-2012

Source: Assessments conducted by Ministries of Education, in collaboration with RTI International. Compilation and calculations by RTI International.

Case Studies

We now turn to two case studies illustrating the implementation of each instrument, rich experiences that may prove useful to policy makers and their advisors in using assessments to inform the development of learning improvement programs.

EGRA in Egypt: a reading revolution

Egypt's legal and policy commitments to education abound: primary education has been compulsory since 1923 (UNESCO, 2000) and combating illiteracy is a 'national duty for which all the people's energies should be mobilized' (Constitution of the Arab Republic of Egypt, 1980, amended in 2007; see Arab Republic of Egypt, 2007). Despite these commitments, Egypt, a middle-income country which has long enjoyed near-universal access to education for the vast majority of urban children, has struggled to provide high-quality education to marginalized populations and governorates, particularly girls and those residing in Upper (southern) Egypt. UNESCO's World Inequality Database on Education (WIDE) reveals the opportunity gap for young people in Egypt, and girls in particular (UNESCO, 2013a). While prospects for schooling have advanced considerably for girls (from 1992 to 2008, Egypt cut by two thirds the share of primary-aged girls never enrolled in school), inequalities still persist. In 2008, 45% of young women aged 17–22 in the poorest quintile in rural Egypt had less than 2 years of schooling, compared with 17% of young men, a gap of 28 percentage points (but an improvement over 1992, when the same gender gap was 53 percentage points).

In addition to these equity and demographic challenges, teaching and learning Arabic present a number of challenges, principal of which is the difference between colloquial spoken and classical

literary or Modern Standard Arabic. Arabic-speaking children grow up in environments where they use the colloquial form of Arabic at home and learn the classical form at school. Throughout the Arabic-speaking world, there is a belief that literary or Modern Standard Arabic is too difficult for children and that they should not be exposed to it before grade 1 (what exposure they do have is often limited to memorization of Koranic text) (Ayari, 1996). Consequently, children encounter literary Arabic in the first grade of primary school almost as a second language. Recent research reveals that children exposed to literary Arabic early on have better reading comprehension than those children exposed only to spoken Arabic (Abu-Rabia & Taha, 2005). Teacher-guided read-aloud tasks and testing from an early age are therefore critical to improving reading in later grades. In Egypt, however, as in many other contexts, most teachers do not regularly assess children's oral reading in the early grades.

In 2008, RTI International began the Girls' Improved Learning Outcomes (GILO) project to support the Egyptian Ministry of Education in improving learning achievement among girls in kindergarten through ninth grade. Funded by USAID/Egypt, GILO directly supported 166 schools in three Upper Egypt governorates. One of the project's main objectives was to improve girls' learning outcomes by adopting a comprehensive and integrated school improvement approach known as school-based reform.

EGRA Sparks a Revolution

Beginning in May 2008, GILO provided technical assistance to the Ministry of Education (MOE) to develop, pilot and implement an EGRA as a practical tool to better understand learner performance in the early grades. EGRA was the first assessment tool available for use at the primary level in Egypt. Collaboration with Arabic linguistics and Egyptian experts in instruction and assessment resulted in a draft instrument that was pretested in a slum area outside Cairo and refined afterward in response to an item analysis. The EGRA used Modern Standard Arabic, thus increasing the possibility of using the tool in other Arabic-speaking countries. GILO's use of the Arabic version of EGRA in a baseline assessment was approved by the Ministry for a secondary purpose as well: GILO and the MOE, working together, would use the findings to design an Early Grade Reading Program to strengthen student reading skills, particularly for girls.

Implementation started with the training of assessors, to which significant attention was paid, along with close supervision to ensure the reliability of assessments and the accuracy of scores. EGRA was administered to 2900 pupils – half girls, half boys – who were randomly selected from grades 2, 3 and 4 in 30 treatment schools and 30 control schools from three governorates in rural Upper Egypt. The baseline survey results revealed significant gaps in instruction, principal among which was that the few students who knew letter sounds were also able to read words. (Improved performance in knowing letter sounds is essential in moving students from basic to intermediate levels of literacy.) Greater attention to classroom instruction in letter sounds – and not letter names alone – was strongly recommended. Overall scores were quite low; average oral reading fluency for grade 2 students was just 11 correct words per minute, whereas 45 correct words per minute is considered the minimum reading speed for Arabic to ensure comprehension. This general pattern of low proficiency, particularly in word and passage reading, held across all grades and mirrored the pattern observed in other countries that had applied EGRA (refer to Figure 1).

GILO presented the baseline findings and recommendations to the Minister of Education and other key stakeholders. The findings were well received, if of significant concern. The Ministry expressed considerable interest in expanding the use of EGRA beyond program schools and using results to strengthen early-grade reading instruction in public schools. The MOE requested GILO's collaboration in assembling a working group to develop an early-grade literacy program and supplemental instructional materials to complement the existing curriculum. GILO accordingly revised the project's strategy for teacher professional development to include a more intensive focus on teacher training and instructional materials development for early-grade reading acquisition in Arabic, drawing on the recommendations from the assessment findings. EGRA in effect opened a previously closed door to introducing modifications to the teaching of reading in Arabic.

Developing a Response: the Early Grade Reading Program (EGRP)

The working group began designing an 8-day training program for teachers in grades 1 through 3, Arabic-language teachers, and supervisors. Training covered social marketing of early-grade reading, phonemic awareness, phonics and reading comprehension. The training package included a trainer's and teacher's manual comprising strategies for explicit literacy instruction, an implementation plan, integrated supportive text for literacy instruction, and specific early literacy phonics routines and lesson plans for teacher professional development activities. The working group also produced recommendations on ways to integrate the routines within the existing MOE curriculum, as well as guides and examples of skill-mastery testing, monitoring and assessment of student performance, and definitions of components of reading.

GILO analyzed early-grade textbooks, finding an immense number of grade 1 vocabulary items, many of which were used only once; words with multiple syllables, up to 6; abstract concepts unrelated to the life experiences of grade 1 pupils; and words introduced 'in isolation' without rhyming support. In addition, many of the textbook pictures did not match the associated words. In July 2010, the project began training teachers in the Early Grade Reading Program in 166 schools. As part of the program, CDs and print versions of EGRP materials were developed and delivered to these schools. That year, the MOE's Director of Basic Education, Shanaz El Dessouki, issued a decree that teachers of grade 1 dedicate 25 minutes each day for EGRP instruction (S. El Dessouki, personal communication, 2011).

In April 2011, GILO administered EGRA for a second time in the project's intervention schools, using independent assessors, aiming to measure post-treatment reading performance of grade 2 students. The objective of this endline data collection was to examine the change that could be attributed to GILO's EGRP training of teachers and supervisors. Data collectors returned to the same experimental and control schools where EGRA data were collected in 2009. Results revealed that students in the EGRP pilot schools significantly outperformed their control group counterparts on all EGRA dimensions. For example, EGRP pilot school students identified 19 more letter sounds per minute at the end of the intervention period, an increase of 194% over baseline. Meanwhile, students in the control group gained just two letter sounds per minute, an increase of only 21% over baseline. In reading fluency, EGRP pilot school students could read, on average, a total of 10 more words per minute – an increase of 82% over baseline – whereas control school students read only three more words per minute, a 38% increase. The overall effect size for the program was 0.48 standard deviations, considered large for education programs.

From Pilot to Scale-up

The positive results of the 2011 endline evaluation prompted the government to consider strategies for expanding EGRP. In cooperation with the MOE, project staff developed a plan for a two-level cascade learning model using existing education staff as trainers. Expansion of the program was jointly funded, with the MOE assuming responsibility for most training and expansion costs. Government ownership of the program and political commitment to its expansion were strong. The First Deputy Minister of Education declared: 'EGRP is a national program. I expect every grade one student to learn to read through EGRP. This is your responsibility. I am checking and observing. Tell me what you need and I will make it happen' (R. Abouserie, personal communication, 2011).

The training package developed for national scale-up included model lessons focused on student-centered learning, classroom management, library use and management, and effective supervision and coaching, as well as the core components of the EGRP: review of phonics and introduction of new phonetic aspects, vocabulary development, and reading comprehension. In order for teachers to use data in making instructional decisions to improve students' learning outcomes, teachers of grades 1 and 2 were trained through the cascade model on mastery monitoring checks, a form of classroom-based assessment conducted by teachers to verify that the majority of students had mastered the content imparted. If less than 80% of the students passed the assessment, the teachers were advised to remediate for the entire class. If more than 80% passed, remediation would be conducted for the remainder who did not meet assessment objectives.

EGRP continues to expand; by September 2013 it will have reached more than 3.5 million students in grades 1 through 3. Although there is no national baseline data set for the grade 1 and 2 program, in early 2013 USAID supported a baseline assessment of a national sample of grade 3 students, prior to government rollout to that grade level. Preliminary results indicated substantial need for support and improvement efforts. Hope for the continued expansion of what stakeholders perceive to be a successful program is high: In a blog post, Isobel Coleman, Senior Fellow at the Council on Foreign Relations, described teacher and parent perceptions of EGRA as follows:

I have never seen teachers, administrators, and parents more excited about a new approach to learning, anywhere. Teachers tell me they don't need to see the next set of test results to know that their students are making leaps in literacy. Already, they can see that the grade one and two students are reading better than those in grades three and four. (Coleman, 2012)

EGMA in Kenya: informing and tracking improvement

Similar to Egypt's experience in assessing and improving reading, the case of Kenya's experience with mathematics improvement is informative to those embarking on classroom-initiated policy and systems change. Like many African countries, Kenya has made great strides in improving access to education, particularly in primary school. After the government began implementing free primary education, gross enrollment rose from around 87% in 1999 to over 100% in 2003 (Glennerster et al, 2011). While recognizing some remaining access-related issues, such as lingering disparities by gender and region, the Kenyan government and donors have begun focusing more attention on the quality of education (Glennerster et al, 2011). Low primary completion rates, such as 60% in 2002 (Achoka et al, 2007), pointed to a problem with school quality. Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ) results suggest that many students are struggling, with 67% of students at a beginning numeracy level or below in mathematics, and 37% of students reading below a level where they would be able to interpret or make inferences about texts (Onsumu et al, 2005).[3] At the same time, the government of Kenya recognized the importance of science and technology to the country's development. Vision 2030, the Kenyan government's long-term development plan, identifies science, technology and innovation as foundations for socioeconomic transformation, stressing the importance of science and technology throughout the education program (Republic of Kenya, 2007; Department of Education, Republic of Kenya, 2012).

Developing an Early Grade Mathematics Program

The standard EGMA instrument was piloted in Malindi district, Kenya, in 2009 by RTI International and East African Development Consultants (EADEC), a local Kenyan research firm with experience working in early-grade assessment. The goals of this application were twofold: to test the instrument in the field, and to provide information about students' math performance to local officials and practitioners. The initial instrument was subsequently reviewed and revised during an adaptation workshop, with broad participation from stakeholders and experts. Subsequently, assessors underwent intensive training, including practice in local schools, and interrater reliability (IRR) testing, with a final IRR score of 95% (Reubens & Kline, 2009).

The EGMA was administered to a sample of 480 students in grades 1 through 3 from 20 schools in Malindi. The schools were randomly selected from across three Malindi divisions. Results from this pilot indicated that students from the divisions sampled were not meeting the expectations laid out in the Kenyan curriculum. Students had particular difficulty with missing numbers and word problems, with students even in grade 3 responding correctly to only 40% of missing number items and 47% of word problems. Results also indicated that students were mastering basic addition later than would be ideal, with almost half of the grade 1 students able to complete at most two basic addition problems. The difficulties with addition and subtraction were twofold: not only were children inaccurate in their attempts to solve these problems, they were also quite slow when they did answer them correctly. Accordingly, the Ministry of Education requested that USAID design a program of interventions in both subjects. The Ministry recognized the importance of ensuring that efforts made to improve mathematics or reading, must be based on

Classroom-Up Policy Change

evidence of what would work in the Kenyan context. RTI International, with support from USAID, worked with the Kenyan Ministry to develop an intervention program with a randomized controlled trial (RCT) design to assess effectiveness in the Kenyan environment. Information also would be collected on cost-effectiveness and ways the intervention could be structured so that if the interventions succeeded, the Ministry could scale them up in a cost-effective, sustainable manner.

This intervention (2011–2014) is based on the same theoretical foundations underlying the EGMA. It combines a focus on conceptual understanding, through development of mathematical concepts and models, procedural fluency and flexibility, and mathematical vocabulary. The program's scope and sequence were developed in close collaboration with the Kenyan Ministry of Education and with local mathematics education experts, plus contributions from experts at the Kenya Institute of Education, to ensure alignment with the local curriculum and appropriateness to the local context. The intervention includes teacher training and follow-up support for teachers, as well as supporting materials for both teachers and students, all designed for Ministry future use.

As noted, the program has an experimental design, which includes three treatment phases: one cohort of schools that began in 2012; another cohort that started in January 2013; and the final cohort, which will serve as the control group but will receive the intervention beginning in 2014, after the October 2013 endline assessment. Layered onto the RCT design is a longitudinal study, in which 1000 individual students will be followed through the first three grades of school. The research design has been developed to explore five research questions: the size of the impact of the intervention in terms of improvement in students' learning outcomes; whether the maturation of the program makes a difference – that is, whether the program has more impact after 3 years than it does after 1 or 2 years; the effects of coach-to-school ratios – 10 schools per coach versus 15 schools per coach; comparisons of the impact in urban and rural schools; and comparisons of the impact in formal schools and non-formal education institutions located in Kenya's slum areas. The longitudinal study is designed to investigate how students' early outcomes in mathematics predict later educational success, how these results relate to family background and socioeconomic status, and how language skills and language of instruction influence student outcomes.

EGMA will be used throughout the experimental program to provide data on mathematics learning outcomes in target grades (1 and 2). Pupils in the longitudinal subsample will be followed through grade 3 as well. A baseline evaluation took place in October 2011, and a midterm evaluation was held in October 2012. As noted above, the endline assessment for the first two cohorts is planned for October 2013. The baseline and midterm assessments already have yielded information to help guide the intervention, while data from all three assessments will be used to track progress and compare results from treatment schools against control schools, in order to answer the research questions. The EGMA was coupled with classroom observations and interview protocols, allowing for an analysis of teachers' fidelity to the new approach, which can provide information both on the program implementability (i.e. the degree to which the intervention approach is successful in inducing schools and teachers to implement with fidelity), and on the ways in which differences in fidelity might impact results. By the end of the program, all of these results will be combined with cost data in order to carry out the cost-effectiveness analysis.

Looking Forward

The intervention program and the EGMA data collection and analysis were designed to allow for eventual scale-up by the Kenyan Ministry, in two important ways. First, the program was designed with existing governmental structures in mind, beginning with a study of the institutional framework – identifying key positions in the administrative system that would be able to support efforts to improve early-grade mathematics instruction, job expectations for those positions, the incentives that influence how those jobs are performed, existing constraints (financial, administrative or legal), and other dimensions of the institutional education environment. Personnel in such key positions will be integrated during the life of the experimental program, in ways such as building capacity of the tutors from the Teachers' Advisory Centres (TACs) to act as teacher trainers. Similarly, Ministry personnel are involved in EGMA data collection.

In addition to explicitly integrating government and Ministry officials into implementation, the program, with its experimental orientation, allows for the testing of several hypotheses about which conditions – e.g. human resources, training, materials and ongoing support – contribute to cost-effective improvements in student achievement. As part of this process, data are being collected concerning the conditions necessary for this reading and math intervention to be successfully implemented. Examples are: the amount of training given to TAC tutors, instructional coaches, teachers and head teachers; the number and content of school support visits; and the results of teacher observations. This information, when linked to EGMA data showing which schools experienced gains in student achievement, will help the Ministry understand factors that contributed to program success.

As noted, those differential impact data will be coupled with evaluation of the human and financial costs of the approach. This evaluation will include the time spent by the full range of actors involved in implementation – such as the teachers' time for training and support activities. It will also include rigorous tracking of project expenditures. Calculating the number of hours of training and support for teachers and comparing them to student achievement impacts will allow the Ministry, the government and NGOs to better understand the human resources necessary to improve student learning. This information will be reviewed with the Ministry to provide recommendations for how to increase budgetary allocations for early learning, across several departments.

While there is already interest in scale-up within the Ministry, waiting until this impact and cost-effectiveness evidence from the program is available will better ensure successful scaling up of the program. At the same time, during the course of program implementation, the Ministry will gain skills in using EGMA to evaluate early-grade student achievement and in using EGMA data in conjunction with other assessments to track progress and make policy and practice decisions. Coupled with donor support to initiate scale-up, this experimental project, if successful, has the potential to lead to sustained improvements in mathematics in the early grades and beyond.

From 'Education For All' to 'Learning For All'

The United Nations (UN) and an array of agencies are involved in consultations on the next round of Millennium Development Goals. Of particular note in the run up to 2015 is a dramatic increase in the attention given to learning. From the UN's 'Education First' initiative, launched in September 2012, to the Brookings Institution's Center for Universal Education Learning Metrics Task Force, nearly every multi- and bilateral agency working at the global level in education has issued a statement on the importance of improving learning outcomes. The education strategy documents of the three largest donors to basic education – the United Kingdom's Department for International Development (DFID, 2010), USAID (USAID, 2011) and the World Bank (World Bank, 2011) – have all identified improving learning as a key or central focus.

This change seems to be occurring even at the development community's most august, and slow-to-change, institutions. After naming former UK Prime Minister Gordon Brown to the post of Special Envoy for Global Education in July 2012, UN Secretary General Ban Ki-moon launched the Education First initiative in September 2012, outlining three key priorities: (1) put every child in school; (2) improve the quality of education; and (3) foster global citizenship (UNESCO, 2013b). In April 2013, as part of the World Bank–International Monetary Fund Spring meetings, the 'Learning for All Ministerial' was held in Washington, DC, bringing together representatives from eight countries (Bangladesh, Democratic Republic of Congo, Ethiopia, Haiti, India, Nigeria, Yemen and South Sudan) that collectively account for more than half of the world's out-of-school children. According to World Bank President Jim Yong Kim, who, together with Mr Ban and Mr Brown, hosted the meeting, the event marked the first time that ministers of finance and ministers of education came together to discuss

what we need to do to make sure that every child can go to school and that every child is learning. We all agreed on the need to take urgent action to address the global learning crisis, which poses a serious threat to our mission to end poverty and build shared prosperity. (Yong Kim, 2013)

With learning increasingly positioned at the center of the global development agenda, tools such as EGRA and EGMA, which seek to respond quickly to government and donor needs to understand the status of student performance on basic literacy and numeracy tasks, may well come into increased usage. Alternative offerings, including expansion of both PIRLS and PISA to ‘prePIRLS’ and ‘prePISA’, are also under development to increase the range of assessment tools available to countries on the lower end of the income and learning spectrums (Wagner, 2011). As long as assessments, whether large in scale or individually tailored to a country’s context, keep the goal of improving learning – rather than just measuring it – at the heart of the work, there is great potential for achieving the system-level improvements that so many governments seek.

Notes

- [1] Despite differences, nearly all efforts have included at a minimum the following core components: letter or syllable identification (either names or sounds); familiar word reading; non-word reading; oral reading fluency with comprehension; and listening comprehension. Additional tasks have included phonemic awareness (measured by phoneme segmentation, initial and final sound identification or phoneme discrimination), as well as dictation, vocabulary and maze/cloze subtasks. Instruments and reports on the results are available at the USAID-sponsored Education Data for Decision Making (EdData II) project website, <http://www.eddataglobal.org>, including instruments and reports conducted by RTI as well as other implementing partners that have agreed to make these resources publicly available.
- [2] The subtasks of the current core instrument include: number identification, which assesses the student’s knowledge and ability to identify written symbols; quantity comparison, which assesses the student’s ability to make judgments about differences by comparing quantities, represented by numbers; missing number (number patterns), which assesses the student’s ability to discern and complete number patterns; addition and subtraction, level 1, which assesses the student’s procedural competency in the basic operations of simple addition and subtraction (addends under 10); addition and subtraction, level 2, which assesses the student’s procedural competency in slightly higher-level addition and subtraction, including 1- and 2-digit numbers; and word problems, which assess the student’s ability to understand and solve problems that have been given orally, using basic operations. The number identification and addition/subtraction level 1 sections are timed at one minute in order to analyze both accuracy and automaticity.
- [3] In 2009, the Early Grade Reading Assessment found that Kenyan students in grade 2 were able to answer on average only 8% of comprehension questions in English and 17% in Kiswahili (Piper, 2010). In 2009, a study by the NGO Uwezo found comparable results, with only 28% of grade 3 students reading at least at a grade 2 level in English, and 36% reading at a grade 2 level in Kiswahili. The math results were similar, with only 36% of students reaching even grade 2-level performance by grade 3 (Uwezo, 2011). These studies contributed to recognition of low performance in early-grade reading, which led to the first call for investment in improving the quality of education in the early grades.

References

- Abadzi, H. (2006) *Efficient Learning for the Poor*. Washington, DC: World Bank.
- Abu-Rabia, S. & Taha, H. (2005) Reading in Arabic Orthography: characteristics, research findings, and assessment, in R. Malatesha Joshi & P.G. Aaron (Eds) *Handbook of Orthography and Literacy*, pp. 321–338. Mahwah, NJ: Lawrence Erlbaum Associates.
- Achoka, J.S.K., Odebero, S.O., Maiyo, J.K. & Mualuko, N.J. (2007) Access to Basic Education in Kenya: inherent concerns, *Educational Research and Review*, 2(10), 275–284.
<http://www.academicjournals.org/ERR/PDF/pdf%202007/Oct/Achoka%20et%20al.pdf>
- Arab Republic of Egypt (2007) *The Constitution of the Arab Republic of Egypt 1980, amended 2007*.
<http://www.right-to-education.org/country-node/333/country-constitutional>
- Ayari, S. (1996) Diglossia and Illiteracy in the Arab World, *Language, Culture and Curriculum*, 9(3), 243–253.
<http://dx.doi.org/10.1080/07908319609525233>
- Baroody, A., Lai, M.-L. & Mix, K.S. (2006) The Development of Young Children’s Early Number and Operation Sense and its Implications for Early Childhood Education, in B. Spodek & O. Saracho (Eds)

- Handbook of Research on the Education of Young Children*, pp. 187–221. Mahwah, NJ: Lawrence Erlbaum Associates.
- Center for Global Development (2006) *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Washington, DC: Evaluation Gap Working Group, Center for Global Development.
<http://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation>
- Chabbott, C. (2003) *Constructing Education for Development: international organizations and Education for All*. New York: RoutledgeFalmer.
- Chabbott, C. (2006) *Accelerating Early Grades Reading in High Priority EFA Countries: a desk review*. Produced by American Institutes for Research under the USAID Educational Quality Improvement Program (EQUIP1) Leader with Associates Award (LWA). <http://www.equip123.net/docs/E1-EGRinEFCountriesDeskStudy.pdf>
- Chapman, D.W. & Quijada, J.J. (2008) An Analysis of US Assistance to Basic Education in the Developing World, 1990–2005, *International Journal of Educational Development*, 29(3), 268–280.
<http://dx.doi.org/10.1016/j.ijedudev.2008.08.005>
- Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D. & Katz, R. (2005) Using Measures of Number Sense to Screen for Difficulties in Mathematics: preliminary findings, *Assessment for Effective Intervention*, 30(2), 3–14. <http://dx.doi.org/10.1177/073724770503000202>
- Clements, D. & Sarama, J. (2007) Early Childhood Mathematics Learning, in F.K. Lester (Ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, pp. 461–555. Charlotte, NC: Information Age Publications.
- Coleman, I. (2012) Democracy in Development: Egypt's reading revolution [Web log post, 15 November], US Council on Foreign Relations. <http://blogs.cfr.org/coleman/2012/11/15/egypts-reading-revolution>
- Crouch, L. & Gove, A. (2011) Leaps or One Step at Time: skirting or helping engage the debate? The Case of Reading, in W.J. Jacob & J.N. Hawkins (Eds) *Policy Debates in Comparative, International and Development Education*, pp. 120–151. Basingstoke: Palgrave Macmillan.
- Department of Education, Republic of Kenya (2012) *A Policy Framework for Education: aligning education and training to the constitution of Kenya (2010) and Kenya Vision 2030 and beyond*. Nairobi: Department of Education.
- Department for International Development (DFID), United Kingdom (2010) *Learning for All: DFID's education strategy 2010–2015*. London: DFID.
<http://consultation.dfid.gov.uk/education2010/files/2010/04/learning-for-all-strategy.pdf>
- Filmer, D., Hasan, A. & Pritchett, L. (2006) A Millennium Learning Goal: measuring real progress in education. Working Paper No. 97. Washington, DC: Center for Global Development.
http://www.cgdev.org/files/9815_file_WP97.pdf
- Foegen, A., Jiban, C. & Deno, S. (2007) Progress Monitoring Measures in Mathematics: a review of the literature, *Journal of Special Education*, 41(2), 121–139. <http://dx.doi.org/10.1177/00224669070410020101>
- Glennerster, R., Kremer, M., Mbiti, I. & Takavarasha, K. (2011) Access and Quality in the Kenyan Education System: a review of the progress, challenges and potential solutions. Report prepared for the Office of the Prime Minister of Kenya. <http://www.povertyactionlab.org/publication/access-and-quality-kenyan-education-system>
- Gove, A. & Cvelich, P. (2011) *Early Reading: igniting Education For All. A Report by the Early Grade Learning Community of Practice*, rev. edn. Research Triangle Park, NC: RTI International.
<http://www.rti.org/pubs/early-readingreport-revised.pdf>
- Hoffman, J.V. (2012) Why EGRA – a Clone of DIBELS – will fail to improve literacy in Africa, *Research in the Teaching of English*, 46(4), 340–357.
- Klees, S.J. (2012) Reading Mania [Web log post], Education in Crisis: monitoring, analysis and resources for education activists. <http://www.educationincrisis.net/blog/item/552-reading-mania>
- Nielsen, H.D. (2006) *From Schooling Access to Learning Outcomes: an unfinished agenda*. Washington, DC: Independent Evaluation Group, World Bank.
[http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/DB332E8D5FF2968E852571A800797D10/\\$file/primary_education_evaluation.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/DB332E8D5FF2968E852571A800797D10/$file/primary_education_evaluation.pdf)
- Onsumu, E., Nzomo, J. & Obiero, C. (2005) *The SACMEQ II Project in Kenya: a study of the conditions of schooling and the quality of education*. Harare, Zimbabwe: Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ).

Classroom-Up Policy Change

- Piper, B. (2010) *Kenya Early Grade Reading Assessment Findings Report*. Prepared for the William and Flora Hewlett Foundation under the Monitoring of Learning Outcomes in Sub-Saharan Africa project. Research Triangle Park, NC: RTI International and East African Development Consultants. <https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=275>
- Republic of Kenya (2007) *Kenya Vision 2030: a globally competitive and prosperous Kenya*. <http://www.vision2030.go.ke/index.php>
- Reubens, A. & Kline, T. (2009) *Pilot of the Early Grade Mathematics Assessment: final report*. Prepared for the USAID Bureau for Economic Growth, Agriculture and Trade (EGAT/ED) under the Education Data for Decision Making (EdData II) Project. Research Triangle Park, NC: RTI International. <https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=194>
- RTI International (2009) *Early Grade Reading Assessment Toolkit*. Prepared with funding from the World Bank and the USAID EdData II Project. <https://www.eddataglobal.org/reading/index.cfm?fuseaction=pubDetail&ID=149>
- RTI International (2013) *EGRA Tracker* [Excel-based summary of worldwide administrations of the Early Grade Reading Assessment]. Research Triangle Park, NC: RTI International. <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=188>
- RTI International and International Rescue Committee (IRC) (2011) *Guidance Notes for Planning and Implementing EGRA*. Washington, DC: RTI International and IRC. <https://www.eddataglobal.org/reading/index.cfm?fuseaction=pubDetail&ID=318>
- United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2000) *The EFA 2000 Assessment: country reports (Egypt)*. http://www.unesco.org/education/wef/countryreports/egypt/rapport_1.htm
- United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2013a) *World Inequality Database on Education* [Data file]. <http://www.education-inequalities.org/>
- United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2013b) *Global Education First Initiative: the UN Secretary General's global initiative on education*. <http://www.globaleducationfirst.org>
- United States Agency for International Development (USAID) (2011) *Education Strategy: opportunity through learning. USAID Education Strategy 2011–2015*. Washington, DC: USAID. http://transition.usaid.gov/our_work/education_and_universities/documents/USAID_ED_Strategy_feb2011.pdf
- Uwezo (2011) *Are Our Children Learning? Numeracy and Literacy Across East Africa*. <http://twaweza.org/go/are-our-children-learning-numeracy-and-literacy-across-east-africa-2011>
- Wagner, D.A. (2003) Smaller, Quicker, Cheaper: alternative strategies for literacy assessment in the UN Literacy Decade, *International Journal of Educational Research*, 39, 293-309. <http://dx.doi.org/10.1016/j.ijer.2004.04.009>
- Wagner, D. (2011) *Smaller, Quicker, Cheaper: improving learning assessments for developing countries*. Paris: International Institute for Educational Planning. <http://unesdoc.unesco.org/images/0021/002136/213663e.pdf>
- Wagner, D.A., Lockheed, M., Mullis, I., Martin, M.O., Kanjee, A., Gove, A. & Dowd, A.J. (2012) The Debate on Learning Assessments in Developing Countries, *Compare*, 42(3), 509–545. <http://dx.doi.org/10.1080/03057925.2012.670480>
- World Bank (2011) *Learning for All: investing in people's knowledge and skills to promote development. World Bank Group Education Strategy 2020*. Washington, DC: International Bank for Reconstruction and Development/World Bank. http://siteresources.worldbank.org/EDUCATION/Resources/ESSU/Education_Strategy_4_12_2011.pdf
- Yong Kim, Jim (2013) Girl Rising: a rally for girls and women – webcast and live blog [video, 9:07–13:03]. Speech presented at Girl Rising Rally, 18 April, World Bank, Washington DC. <http://live.worldbank.org/girl-rising-rally-girls-and-women>

AMBER GOVE is Director, Teaching and Learning, International Education Division, RTI International. Her research interests are: improvement and measurement of student learning; education finance; conditional cash transfer programs; factors affecting achievement, enrollment and attendance; program and impact evaluation research. Correspondence: agoive@rti.org

SAMIR HABIB is Chief of Party, USAID/Egypt Girls' Improved Learning Outcomes (GILO) Project, RTI International. His research interests are: Arabic reading acquisition; improvement and measurement of student learning; education policy, systems and governance. *Correspondence:* sshafik@rti.org

BENJAMIN PIPER is Chief of Party, USAID/Kenya Primary Math and Reading (PRIMR) Initiative, RTI International. His research interests are: transitions between primary and secondary education; evaluation of education quality; quantitative research, including longitudinal data analysis, randomized controlled trial designs and mixed-methods designs. *Correspondence:* bpiper@rti.org.

WENDI RALAINGITA is Senior Reading Advisor, USAID/Ethiopia Reading for Ethiopia's Achievement Developed (READ) Technical Assistance Project, RTI International. Her research interests are: school- and classroom-based research; teacher professional development; mathematics, reading and language instruction; program management and evaluation. *Correspondence:* wralaingita@rti.org