Compare: A Journal of Comparative and International Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ccom20

The debate on learning assessments in developing countries

Daniel A. Wagner [a], Marlaine Lockheed [b], Ina Mullis [c], Michael O. Martin [c], Anil Kanjee [d], Amber Gove [e] & Amy Jo Dowd [f]

[a] University of Pennsylvania, USA

[b] World Bank (retired), NJ, USA

[c] Boston College, USA

[d] Tshwane University of Technology

[e] RTI International

[f] Education Research at Save the Children

Version of record first published: 30 Mar 2012

To cite this article: Daniel A. Wagner, Marlaine Lockheed, Ina Mullis, Michael O. Martin, Anil Kanjee, Amber Gove & Amy Jo Dowd (2012): The debate on learning assessments in developing countries, Compare: A Journal of Comparative and International Education, 42:3, 509-545

To link to this article: http://dx.doi.org/10.1080/03057925.2012.670480

Compare
Vol. 42, No. 3, May 2012, 509–545

COMPARE FORUM

The debate on learning assessments in developing countries

Daniel A. Wagner
University of Pennsylvania, USA

Over the past decade, international and national education agencies have begun to emphasize the improvement of the quality (rather than quantity) of education in developing countries. This trend has been paralleled by a significant increase in the use of educational assessments as a way to measure gains and losses in quality of learning. As interest in assessment has grown, low-income countries have begun to adopt and adapt international and other assessments for a variety of uses, including the comparability of national quality with other countries, improved ways of measuring reading achievement, and further attempts to reach marginalized populations within a country. The present group of papers provides multiple perspectives on the debate currently underway about the best approaches to create and use learning assessments in low-income countries.

Papers

(These papers were originally presented at the Annual Meetings of the Comparative and International Education Society, Montreal, May 2011.)

Daniel A. Wagner, What should be learned from learning assessments?

Marlaine Lockheed, Policies, performance and panaceas: what international large-scale assessments in developing countries

Ina Mullis and Michael O. Martin, PIRLS/prePIRLS: new options for developing countries to monitor children's progress in reading

Anil Kanjee, Options for developing integrated assessment systems to improve learning and teaching

Amber Gove, Think global, act local: how early reading assessments can improve learning for all

Amy Jo Dowd, An NGO perspective on assessment choice: from practice to research to practice

ISSN 0305-7925 print/ISSN 1469-3623 online

http://dx.doi.org/10.1080/03057925.2012.670480
http://www.tandfonline.com

What should be learned from learning assessments?

Daniel A. Wagner
University of Pennsylvania, USA

Learning assessments have grown increasingly important as policy-makers and other educational consumers (agencies, schools, communities, parents, individuals, etc.) have sought to understand what is (and isn't) learned as a function of information inputs. Increasingly, ministers of education are no longer satisfied with *who* is attending school, but also *how well* what is being taught is learned, and perhaps how well the minister is doing relative to ministers in other countries. Even at the individual level, in both wealthy and poor countries, both parents and children want to know *whether or not* they will succeed in school, or in learning a second language, or be able to get a job with the skills that they have or might have. In sum, learning assessments have been around as long as parents have been trying to teach their children, and institutions have been trying to determine who is intellectually fit for a particular job. And, they are increasingly used as the globalized economy puts a bigger and bigger premium on what skills individuals possess.

Yet, learning assessments are essentially tools. And, like any tool, a learning assessment can be *very* right for the job (a flat-headed, hand-sized hammer hitting a nail, for example), *partially* right (a round-headed hammer that sometimes drives in the nail, but often doesn't), and *not good at all* (a sledge-hammer trying to hit a small nail). Of course, such metaphors are only partially helpful. And that is also a problem with learning assessments. No tool is perfect, and no learning assessment is either. Indeed, the research literature on learning assessments is quite robust, especially for use in relatively wealthy OECD countries that have a long history of substantive empirical and statistical work on such issues as validity, reliability, sampling and scaling methodologies. The strength of this history of empirical work has, for many in the scientific community as well as in the press and in politics, conveyed a powerful sense of credibility to learning assessments. It is not surprising, therefore, that there has developed some certainty that we know what we need to know about designing adequate learning assessments, whether across diverse populations in rich countries, or in poor countries.

Such certainty is inappropriate, however. For it is not just a matter of how to agree on the technical parameters of a 'good' assessment. Sample sizes, alpha coefficients, test-retest reliability and predictive validity are all areas where there is substantial agreement among test-makers the world over. By contrast, *who* gets tested, *what* gets tested, *when* tests occurs, *how* and *why* a test takes place are all contested issues. And, rightly, they are particularly becoming contested in low-income countries where the growth in assessments is most rapidly expanding and where the empirical base is least developed.

As argued elsewhere (Wagner 2010, 2011), there are many factors to consider besides the kinds of technical parameters mentioned above, and these can directly affect and bias assessment outcomes. Let us take, as a hypothetical example, a typical primary school in rural Senegal. If the policy goal is to promote literacy (universal basic education) by the end of primary school (which is part of Senegal's Education For All plans), then learning assessments can play an important role by tracking progress over time toward that goal. But what kinds of learning assessments would be most useful? Then, the five questions – *who*, *what*, *when*, *how*, and *why* – become of central importance. *Who* should be included in the assessment population: those that speak small minority African languages; refugees from the Sierra Leone and Liberia civil wars; 'over-age' children who are repeating first and second grades for two or three times? The answers to these (and other) questions will not only affect progress toward learning achievement, but also the type of content (the *what* question) that must be included in the assessment, and in the school curriculum and teacher training as well. These factors, in other words, have real costs – political and budgetary (see Kanjee; Wagner, Babson, and Murphy 2011).

Similarly, knowing *when* an assessment should take place is non-trivial. As pointed out by Gove, when children are assessed for reading at a younger age, there is more opportunity to remediate at the individual level, teacher level and policy level. There is simply more time to make adjustments that can improve both policy and practice in learning and instruction. The *how* of learning assessments has often been thought of as the contrast between individualized versus group testing, but it can also be thought of as the kinds of item adjustments that need to be made when samples of learners simply do not fit on the same scales as other students from more disadvantaged environments (Mullis and Martin).

The *why* question is less obvious, since there are often multiple reasons why learning assessments are put into place. For large-scale assessments, Lockheed suggests convincingly that one of the key rationales includes the transparency of education system outcomes that may be compared across national contexts and that supports analytic work based on solid data sets. Dowd, by contrast, argues that NGOs have a clear and present need to continuous track the implementation of its programmes, often on a small scale, and that learning assessments are one of the ways to achieve this goal. Other purposes for learning assessments include support of teacher professional development, improved instructional design, reducing learning inequities, and more.

What should we learn from learning assessments? Clearly, there is much that we need to know, some of which have been cited above in the five questions. Most important is the need to demand purpose-appropriate assessments that enable change to take place that tracks the social goals that are accepted and supported by the international community. We must accept

that learning assessments are a work in progress, and that no one has the perfect tool for the many goals and contexts that need to be addressed.

In 1865, Lewis Carroll famously wrote: 'If you don't know where you are going, any road will get you there'. In order to improve the quality of education, we need to know some of the multiple roads that might take us there, along with the milestones to know we are going in the right direction.

References

Carroll, L. 1865/2006. *Alice in wonderland*. New York: Firefly Books.

Dowd, Amy Jo. 2012. An NGO perspective on assessment choice: from practice to research to practice. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Gove, Amber. 2012. Think global, act local: how early reading assessments can improve learning for all. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Kanjee, Anil. 2012. Options for developing integrated assessment systems to improve learning and teaching. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Lockheed, Marlaine. 2012. Policies, performance and panaceas: what international large-scale assessments in developing countries. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Mullis, Ina V.S., and Michael O. Martin. 2012. PIRLS/prePIRLS: new options for developing countries to monitor children's progress in reading. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Wagner, D.A. 2010. Quality of education, comparability, and assessment choice in developing countries. *COMPARE: A Journal of Comparative and International Education* 40, no. 6: 741–60.

Wagner, D.A. 2011. *Smaller, quicker, cheaper: Improving learning assessments for developing countries*. Washington, DC/Paris: FTI/UNESCO-IIEP.

Wagner, D.A., A. Babson, and K.M. Murphy. 2011. How much is learning measurement worth? Assessment costs in low-income countries. *Current Issues in Comparative Education* 14: 3–21.

Policies, performance and panaceas: the role of international large-scale assessments in developing countries

Marlaine Lockheed
World Bank (retired)

The participation of low and middle-income countries in international large-scale assessments has increased dramatically over the past decades, from only a handful in the 1970s to nearly half of all developing countries by 2010 (Figure 1). But has this increase improved education policy and performance in these countries, and, if so, what is the evidence?

The term 'international large-scale assessments' refers to assessments having three characteristics; they: (1) involve multiple countries; (2) utilize tests that are uniform and standardized across countries in terms of content,

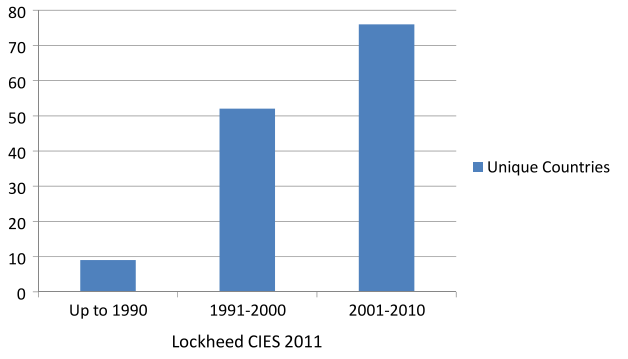

Figure 1. Developing countries in international large-scale assessments (ILSAs).

administration process, timing and scoring; and (3) involve large samples of test-takers scientifically selected from comparable populations across countries. Thus, they differ sharply from national assessments and examinations in scale, scope and purpose; for a discussion of these differences, see Lockheed (2008).

Major international large-scale assessments that meet the above three criteria are regularly carried out by the International Association for the Evaluation of Education Achievement, or IEA (e.g. TIMSS and PIRLS), the OECD (e.g. PISA), the Laboratoria Latinoamericano de Evaluacion de la Calidad de la Education or LLECE (e.g. TERCE), and the Southern and Eastern Africa Consortium for Monitoring Education Quality, or SACMEQ; other examples include ETS's International Assessment of Educational Progress, International Adult Literacy Survey and Programme of International Assessment of Adult Competencies.

There is little doubt that international large-scale assessments are a public good, in that they provide transparency regarding the outputs of educational systems and meaningful measures of human capital development (Hanushek and Woessmann 2009). Their results are widely reported at professional conferences, cited in the academic literature and summarized in UNESCO's EFA Global Monitoring Reports (e.g. UNESCO 2009). Countless secondary analyses of the data from these assessments shed light on both the *effects* of student achievement (for example, cross-country economic growth studies) and its *determinants* (for example cross-country multilevel models of the effects of institutional, school, classroom and family background on achievement).

But what evidence do we have that international large-scale assessments actually benefit participating countries? Have they led to better educational policies (and if so, which ones)? Have they improved the performance of students, teachers, schools or institutions? Is international large-scale

assessment the remedy for all educational diseases – that is, an educational panacea – that some claim it to be for developing countries?

First, international assessments have affected selected education policies

Education policies drive activities related to these policies. Wyckoff (2009) describes these five activities as: (1) redistribution activities (moving money from one group to another, as in the case of Social Security); (2) investment activities (directing money into human or physical capital, as in the case of highways); (3) service activities (providing goods or services, as in the case of lighthouses); (4) regulatory (establishing rules for social and economic activity, as in the case of courts); and (5) behavioural (trying to change the behaviour of groups of individuals, as in the case of teen pregnancy).

While the results of international large-scale assessments could influence all five types of government policy activities, their main influence comes from cross-country comparisons related to only three: investment; regulatory; and behavioural activities. First, international assessments can document the poor performance of a country relative to other countries at similar levels of economic development, which could motivate a country to alter its investments in human capital development through education. Second, a detailed analysis of the strengths and weaknesses of a nation's curriculum as compared to that of other countries could motivate regulatory reform regarding the content and methods of instruction. And third, results related to differences in student achievement associated with cross-country differences in teaching strategies could lead to efforts to change the behaviour of teachers through programmes of pre-service and in-service professional development.[1]

Concrete evidence that international assessments actually have had any policy impact in developing countries is relatively slim, however, and largely based on non-scientific surveys and interviews. However, several systematic evaluations have found that international assessments have indeed influenced regulatory activities around curriculum content and performance standards, and behavioural activities around classroom instruction and teacher professional development. These were the conclusions of four evaluations of IEA's assessments across low- and middle-income countries (Elley 2002; Gilmore 2005; Aggarwala 2004; Lockheed 2010) as well as three evaluations of SAQMEC in sub-Saharan Africa (Murimba 2005; Ercikan et al. 2008; Nzomo and Makuwa 2006).

Second, international assessments have strengthened the performance of testing and assessment organizations

International large-scale assessments are poorly suited for improving the individual performance of students or the individual effectiveness of teachers or schools; this is because they typically are sample based – for content domains as well as by students, teachers and schools.

A country's participation in international large-scale assessments reinforces national technical and managerial capacity for assessment, however, through both training and hands-on experience. It exposes participants to international quality standards in testing and measurement; it provides participants' experience with the technical fields of sampling, test development, questionnaire development, data management and quality control; it builds participants' management capacity for undertaking large research endeavours; it helps education officials prepare reports for policy-makers. These are the conclusions of all five of the evaluations of the IEA international assessments, based on surveys and interviews with the national research coordinators of the various assessments (Elley 2002; Gilmore 2005; Aggarwala 2004; Lockheed 2010). They are also the conclusions of the evaluations of SACMEQ and SERCE (Ercikan et al. 2008; Solano-Flores and Bonk 2008).

In part, international large-scale assessments build technical skills *because* they are international, and require cross-country cooperation. Researchers meet regularly and 'learn by doing' as part of country teams and international teams. Outside of meetings, communication through the Internet enables technical experts in one country to assist others, in real time. This type of learning is a powerful tool for enhancing performance.

Whether participating in international large-scale assessments actually increases a country's probability of undertaking a national large-scale assessment is moot; it appears that the two move hand-in-hand. Over time, the number of countries with no assessments has declined, while the number of countries with both international and national large-scale assessments has increased (Lockheed 2011). Participation in one international large-scale assessment is strongly related to participation in a subsequent international large-scale assessment, and conducting a national large-scale assessment is strongly related to conducting a subsequent national large-scale assessment, but only small cross-over effects from one type of assessment to the other can be observed (Lockheed 2011).[2]

Third, international assessments are not panaceas – remedies for all diseases – for education problems

International large-scale assessments provide transparent information regarding the outputs of educational systems; gaps in achievement within countries; and correlates of cognitive skills. And they provide this useful descriptive information and these rich data sets – suitable for secondary analysis – largely because they are voluntary, low-stakes enterprises.[3] Since they rarely provide information about *all* students, teachers or schools in a country, they are *not* good for holding schools or teachers accountable or for creating incentives to reward school or teacher performance (Lockheed 2008; Barlevy and Neal 2010). While they *might* be used to hold entire edu-

cation systems accountable or to create performance incentives for an entire education system, to date there are no international bodies charged with this responsibility, and – in any case – the assessments would be corrupted by such accountability (Braun and Kanjee 2006).

Existing international bodies with some responsibility for education, such as UNESCO and the World Bank, emphasize the descriptive role of large-scale international assessments. The 2009 and 2010 UNESCO *EFA global monitoring reports* (GMRs) emphasize measurement over accountability and include significant coverage of the results from international assessments (UNESCO 2008, 2009). They articulate the utility of international large-scale assessment in providing evidence regarding both levels and gaps in learning achievement. For example, the 2010 GMR argues that 'participating in global and regional assessment exercises … make it possible to measure disparities between countries in terms of the skills students attain after a given period of learning' (UNESCO 2009, 105). The World Bank's newest strategy paper, *Learning for all: Investing in people's knowledge and skills to promote development*, also pays significant attention to international large-scale assessments and their role in measuring actual learning outcomes worldwide (World Bank 2011).

By comparison, some Washington DC think tanks have called for assessment-based education accountability and performance incentives. For example, a recent Brookings Institution paper calls for establishing systems of national learning assessments and national education accounts so that education outcomes can be better linked to education spending – clearly a call for accountability (van der Gaag and Adams 2010).[4]

Should these calls extend to international large-scale assessments? My view is that they should not. Changing international large-scale assessments from low-stakes to high-stakes enterprises – where actual resources would be provided to countries on the basis of student test performance – would in all likelihood introduce distortions in the assessments and compromise their ability to provide both valid and reliable cross-national measures of human capital and valuable data for secondary analysis. Such distortions include a narrowing of the curriculum to focus on the subjects or competencies tested in the assessments, an undue emphasis on test preparation, and even outright cheating by students, teachers and schools participating in the assessment (Braun and Kanjee 2006; Lockheed 2008).

Summary

Thus, to summarize, I conclude that large-scale international assessments should be valued – and supported – for what they are good for. First, they motivate regulatory and behavioural policy reforms around the content of teaching and learning, Second, they create a learning environment in which assessment specialists can improve their technical skills and related

performance. Third, they increase transparency regarding education system outcomes and human capital development in a cross national context and support analytic work enabled by the data sets.

But they should not be viewed as a panacea for education, particularly in the areas of accountability and incentives. Invoking Campbell's Law,[5] I think that using international large-scale assessments for such purposes would surely kill the goose that lays a very golden egg (Campbell 1975).

Notes

1. Recent increased attention to Cuba's education system, Singapore's mathematics curricula and Finland's teacher policies suggests these effects.
2. Details are available from the author.
3. Of course, some countries encourage their students to work hard and try to excel on international assessments, and some ministries of education have responded negatively to the results of these assessments, that international large-scale assessments have few direct consequences for students, teachers or schools that participate in them.
4. By comparison, the Center for Global Development's book, *Cash on delivery: A new approach to foreign aid*, proposes that education aid should be delivered to countries on the basis of the number of students who take a standardized test at the end of primary school, but recognizes the pitfalls of conditionality based on student performance on the test (Birdsall and Savedoff 2010).
5. 'The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures, and the more apt it will be to distort and corrupt the social processes it is intended to monitor' (Campbell 1975).

References

Aggarwala, N. 2004. *Quality assessment of primary and middle education in mathematics and science (TIMSS)*. Takoma Park, MD: Eaton-Batson International.

Barlevy, G. and D. Neal. 2010. Pay for percentile. NBER Working Paper 17194. Cambridge, MA: National Bureau for Economic Research.

Birdsall, N., and W. Savedoff. 2010. *Cash on delivery aid: A new approach to foreign aid*. Washington, DC: Center for Global Development.

Braun, H., and A. Kanjee. 2006. Using assessment to improve education in developing nations. In *Educating all children: A global agenda*, ed. J. Cohen, D. Bloom, and M. Malin, 303–54. Cambridge, MA: MIT Press.

Campbell, D. 1975. Assessing the impact of planned social change. In *Social research and public policies: The Dartmouth/OECD conference*, ed. G. Lyons. Hanover, NH: Public Affairs Center, Dartmouth College.

Elley, W. 2002. *Evaluating the impact of TIMSS-R (1999) in low- and middle-income countries: An independent report on the value of World Bank support for an international survey of achievement in mathematics and science*. New Zealand.

Ercikan, K., R. Arim, M. Oliveri, and D. Sandilands. 2008. *Evaluation of dimensions of the work of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and of its programme of cooperation with the International Institute for Educational Planning (IIEP)*. Document IOS/EVS/PI/93. Paris: UNESCO Internal Oversight Service Evaluation Section.

Gilmore, A. 2005. *The impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries: an evaluation of the value of World Bank support for international surveys of reading literacy (PIRLS) and mathematics and science (TIMSS)*. New Zealand.
Hanushek, E., and L. Woessmann. 2009. Do better schools lead to more growth? Cognitive skills, economic outcomes and causation. NBER Working Paper 14633. Cambridge, MA: National Bureau for Economic Research.
Lockheed, M. 2008. Measuring progress with tests of learning: Pros and cons for progress-based aid in education. CGD Working Paper 147, Center for Global Development, Washington, DC.
Lockheed, M. 2010. *The craft of education assessment: Does participating in international and regional assessments build assessment capacity in developing countries?* NJ: Princeton.
Lockheed, M. 2011. Policies, performance and panaceas: What international assessments are good for (and not good for) in developing countries. Paper presented at the Annual Meeting of the Comparative and International Education Society, May 1–5, in Montreal, Canada.
Murimba, S. 2005. The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). *Prospects* 35, no. 1: 91–108.
Nzomo, J., and D. Makuwa. 1975. How can countries move from cross-national research results to dissemination, and then to policy reform? In *Cross-national studies of the quality of education: Planning their design and managing their impact*, ed. K.N. Ross and I. Jurgens-Genevois. Paris: IIEP (UNESCO).
Solano-Flores, G., and W. Bonk. 2008. *Evaluation of the Latin American Laboratory for the Evaluation of Education Quality (LLECE)*. Document IOS/EVS/PI/91. Paris: UNESCO Internal Oversight Service Evaluation Section.
UNESCO (United Nations Educational, Scientific and Cultural Organization). 2008. *EFA global monitoring report 2009: Overcoming inequality: Why governance matters*. Paris: UNESCO.
UNESCO (United Nations Educational, Scientific and Cultural Organization). 2009. *EFA global monitoring report 2010: Reaching and teaching the most marginalized*. Paris: UNESCO.
van der Gaag, J., and A. Adams. 2010. *Where is the learning? Measuring Schooling efforts in developing countries*. Washington, DC: The Brookings Institution Center for Universal Education.
World Bank. 2011. *Learning for all: Investing in people's knowledge and skills to promote development*. Washington, DC: The World Bank.
Wyckoff, P. 2009. *Policy and evidence in a partisan age: The great disconnect*. Washington, DC: Urban Institute Press.

PIRLS/prePIRLS: new options for developing countries to monitor children's progress in reading

Ina Mullis and Michael O. Martin
Boston College, USA

Background

PIRLS (Progress in Reading Literacy Study) is IEA's international assessment of trends in reading comprehension for fourth grade students.[6] First

assessed in 2001, PIRLS was developed beginning in 1998 in collaboration with 40 countries, most of them industrialized, to assess students who had learned to read and were now reading to learn (Mulllis et al. 2009). Designed to be a state of the art replacement for IEA's 1991 Reading Literacy Study, PIRLS includes long texts from a range of reading materials (up to 1,000 words), various question types (half the questions ask students to write out their answers) and devotes substantial time to higher-order reading skills.[7] PIRLS has been monitoring trends in student achievement at five-year intervals since 2001, with nearly 50 countries participating in PIRLS 2011 and plans underway for 2016.

Considerable effort is expended to ensure that PIRLS is administered in students' language of instruction (e.g. South Africa translated the PIRLS assessment into 11 languages used in primary school instruction). However, even with this attention to students' languages, the results from PIRLS 2006 made it clear that PIRLS was too difficult for many of the fourth grade students in several of the developing countries, including Kuwait, Morocco, Qatar and South Africa (Mullis et al. 2007). For a variety of reasons, the curriculum in some countries is such that children in the fourth grade are still developing fundamental reading skills.

To meet the needs of the increasing number of developing countries wanting to participate in PIRLS 2011, IEA developed a less difficult assessment to bridge to PIRLS, called prePIRLS. The prePIRLS assessment has fewer and shorter reading texts, and places less emphasis on higher-order reading skills (Mulllis et al. 2009). Depending on a country's educational development, prePIRLS can be given at the fourth, fifth or sixth grade. As another new initiative, PIRLS also can be given to students in the fifth or sixth grade in countries where the assessment might be too difficult for fourth grade students. With the two new initiatives, incorporated in 2011, PIRLS and prePIRLS together now meet the needs of a broader range of countries, providing new options for developing countries to assess reading at the end of the primary school cycle.

Considering the PIRLS/prePIRLS assessment options

Using Lockheed's definition of international large-scale assessments involving (1) multiple countries, (2) uniform tests and standardized procedures, and (3) large nationally representative samples of students, PIRLS/prePIRLS have some characteristics in common with other international large-scale reading assessments(Lockheed). However, some of these assessments are for older students (PISA for 15-year olds and adult literacy). Of those assessments for younger students, some are regional assessments (e.g. SACMEQ), whereas PIRLS/prePIRLS are global assessments including countries from

all around the world. In fact, PIRLS/prePIRLS are the only international reading assessments that:

- Assess reading comprehension at fourth, fifth and sixth grades (end of primary cycle);
- Use uniform tests across countries, standardized in terms of content, administration process, timing and scoring;
- Include diverse countries from Africa, Asia, Australasia, Europe and North and South America;
- Monitor trends at regular intervals over time;
- Collect extensive information about national curricula and curriculum implementation;
- Use scientific sampling methods and matrix sampling to reduce student burden; and
- Train countries in data analysis so they can use their PIRLS data to study national concerns.

Although both PIRLS's extension into sixth grade and, in particular, the prePIRLS assessment provide new options for developing countries, effective country participation in PIRLS/prePIRLS depends on children having at least a minimal level of basic reading skills. If as many as 30% of the students have not developed basic reading skills by the sixth grade, then a country most likely is not educationally ready to participate in PIRLS. For countries where students at the end of primary school are still developing basic skills, participation in prePIRLS would be a more beneficial and viable option.

Considering both the grade level and the reading ability of their students, countries took advantage of the two new PIRLS/prePIRLS initiatives in several ways in 2011. For example, Honduras conducted PIRLS at the sixth grade, and Libya and Egypt were planning to assess prePIRLS at the fourth grade. Participating in PIRLS/prePIRLS at two grades can provide countries with comprehensive data about how well students are learning to read at different points in their educational systems. As such, Morocco conducted PIRLS at both the fourth and sixth grades, Botswana conducted PIRLS at the sixth grade and prePIRLS at the fourth grade, and South Africa conducted prePIRLS at the fourth grade with a special PIRLS study at the fifth grade. Most interesting, Colombia conducted both PIRLS and prePIRLS at the fourth grade with the same students (following a counterbalanced design developed by the TIMSS and PIRLS International Study Center). The Colombian data enable a link in reading achievement between PIRLS and prePIRLS.

Also, there are a number of approaches to gathering school, regional and national achievement results about children's reading skills and competencies that can be implemented prior to (or in conjunction with) participation in

PIRLS/prePIRLS. For example, if aiming to assess the level of beginning reading in the early years of schooling, a less-developed country can consider using the Early Grade Reading Assessment (EGRA) described by Amber Gove. Typically administered to children during first to third grades and adapted to local contexts, EGRA focuses on such cognitive pre-reading skills as phonemic awareness and could be considered a 'pre' prePIRLS assessment to gain information about whether students are developing the basic comprehension skills necessary for participation in prePIRLS or PIRLS.

Assessing reading purposes and process

The PIRLS and prePIRLS assessments are based on the PIRLS assessment framework, collaboratively updated each assessment cycle by participating countries. Both developed and developing countries updated the framework for 2011 to improve PIRLS and include prePIRLS (Mullis et al. 2009). The prePIRLS assessment is based on the same conception of reading literacy as PIRLS, but is intended to measure the reading comprehension skills of students still learning to read. As shown in Figure 2, purposes for reading and processes of comprehension are the foundation for the PIRLS and prePIRLS assessments of reading comprehension.

The PIRLS and prePIRLS assessments focus on the two overarching reading purposes that account for most of the reading done by young students in and out of school – reading for literary experience and to acquire and use information. Both PIRLS and prePIRLS employ the same assess-

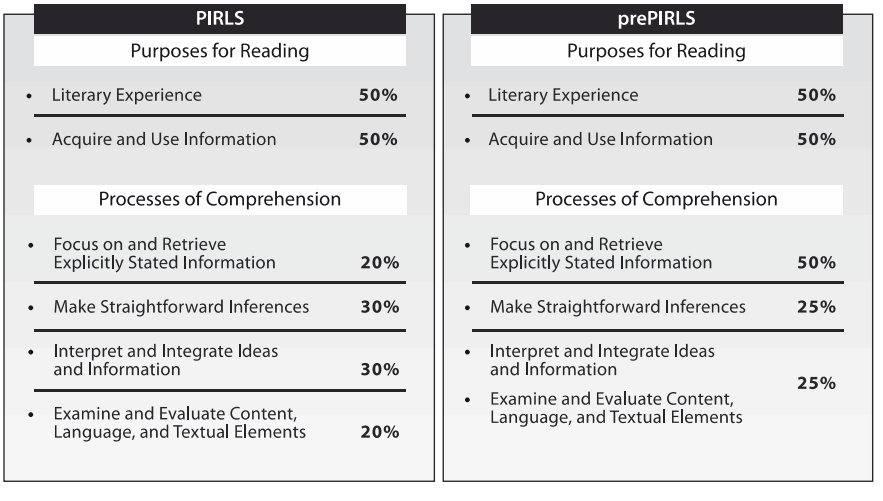

Figure 2. Percentages of the PIRLS and prePIRLS reading assessments devoted to reading purposes and processes (reproduced from *PIRLS 2011 Assessment Framework*).

ment approach whereby students are given reading passages (texts) and asked 13–16 questions about each passage. PIRLS and prePIRLS each contain approximately 130 questions, with approximately half being multiple-choice questions and half being in a constructed-response format where students write their answers.

In both PIRLS and prePIRLS the passages are accompanied with colourful illustrations to help engage student interest, with half literary passages (fictional stories) and half informational passages (content-oriented articles with non-continuous text features such as text boxes, illustrations or diagrams). However, prePIRLS has slightly fewer passages than PIRLS (eight compared to ten) and the passages themselves are shorter (about 400–500 words compared to 800–1,000 words). To further reduce reading difficulty, the prePIRLS passages have easier vocabulary as well as simpler grammar and syntax. Also, the format intersperses questions throughout the passages so that students can read short portions of text and then answer questions, then read a little more and answer more questions, with several questions about the entire passage at the end.

Both PIRLS and prePIRLS assess reading comprehension processes across the two purposes for reading, but prePIRLS places more emphasis on children being able to comprehend and retrieve information from text. The focus of the prePIRLS assessment is on:

- Recognizing words and phrases;
- Understanding sentences and simple paragraphs;
- Retrieving explicitly stated information;
- Making straightforward inferences; and
- Comprehending the overall message.

Context questionnaire data

In addition to data on student achievement, a major purpose of IEA studies is to provide important information for educational improvement. Per the PIRLS framework, the PIRLS/prePIRLS assessments are accompanied by a series of questionnaires about a participating country's national, school, classroom and home contexts in which students learn to read. Figure 3 reflects the contents of the PIRLS/prePIRLS contextual questionnaire data.

The *PIRLS 2011 Encyclopedia* (Mullis, Martin, Drucker, and Minnich 2012) and the *PIRLS 2011 International Results in Reading* (Mullis, Martin, Foy, and Drucker 2012) contain a wealth of information about the educational systems and practices in the participating countries. In the *PIRLS 2011 Encyclopedia*, each participating country describes its reading curriculum and outlines the policies and practices that guide reading instruction and teacher education. The *PIRLS 2011 International Results in Reading* contains the reading achievement results for PIRLS/prePIRLS and presents the context questionnaire data in relation to reading achievement.

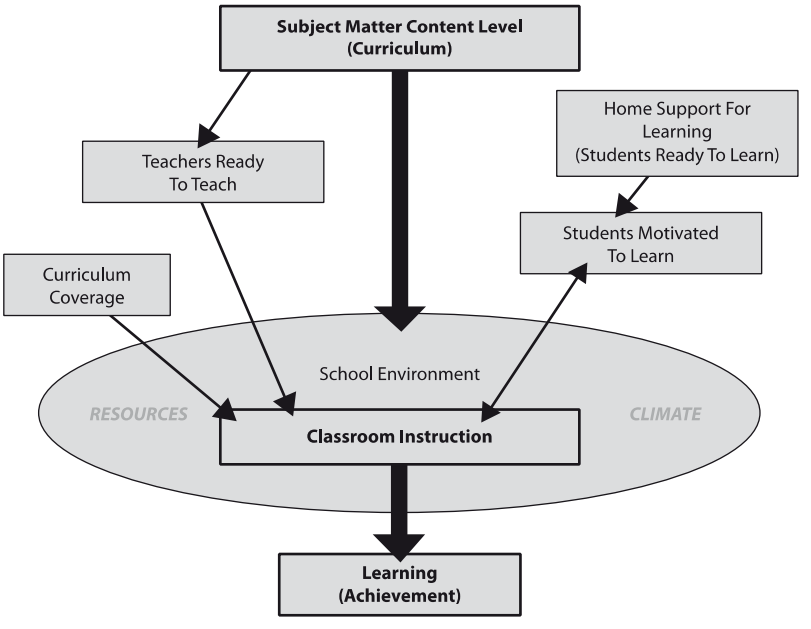

Figure 3. Frame for the PIRLS/prePIRLS 2011 contextual questionnaire data (reproduced from *PIRLS 2011 International Results in Reading*).

Using PIRLS/prePIRLS to improve teaching and learning

Countries that participate in PIRLS/prePIRLS obtain high-quality information about how well their children can read and the factors that positively influence learning to read, thereby acquiring substantial grist for working on educational improvement. The results about achievement and trends in achievement are summarized for reading overall and according to four international benchmarks (advanced, high, medium and low). Studying the item-by-item results can provide important insight into relative strengths and weaknesses within countries that can lead to improvement in curriculum and instruction.

As described by Lockheed, participating countries use international assessment data for system-level monitoring in a global context, curriculum reform and building capacity in assessment methods, and PIRLS/prePIRLS present a good example of this. Because PIRLS/prePIRLS regularly provide high-quality reading achievement data that are comparable county-to-country, they are especially useful for system-level monitoring; that are, countries can assess their educational health in a global environment. Also, countries can set goals, initiate reforms and then see if the next assessment results indicate any improvement. Also, a range of developed and developing countries have used the PIRLS frameworks, achievement results and assessment items to implement curricular reform in reading (Mullis and Martin 2012). For example, developing countries have added informational

reading to their curriculum (e.g. Iran), increased emphasis on higher-order comprehension strategies (e.g. Botswana), and analysed their data to study equity issues (e.g. South Africa).

Regarding capacity building, PIRLS/prePIRLS are organized to encourage participation by subject matter specialists from the participating countries at the various stages of implementation (framework review, item development, data analysis, etc.). Also, countries receive special training in sampling, test administration and scoring procedures. After each assessment, the TIMSS and PIRLS International Study Center provides complete technical documentation[8] and prepares a well-documented database. IEA's Data Processing and Research Center provides training in how to analyse the data to study national issues.

In summary, prePIRLS was specifically designed to provide developing countries an assessment option that could be used for system-wide monitoring in a global context. The prePIRLS assessment was built on the same internationally designed framework for reading comprehension as PIRLS, but is less difficult. In addition, for countries where students have gained basic reading skills by the end of the primary cycle, PIRLS may be given at Grade 5 or Grade 6. These are good assessment choices for developing countries that are interested in establishing a baseline achievement measure and then monitoring trends in that achievement over time. The information about achievement trends in reading together with the data about contexts for teaching and learning are fundamental for evidence-based decision-making.

Notes

1. The International Association for the Evaluation of Educational Achievement (IEA) is an independent international cooperative of national research institutions and government agencies. IEA pioneered the idea of conducting studies of cross-national achievement so that countries could learn from each other, and has been conducting international comparative studies in wide range of subjects since 1959.
2. The IEA General Assembly approved PIRLS as an essential component of IEA's regular cycle of core international assessments, which also includes the TIMSS assessments of mathematics and science. PIRLS/prePIRLS and TIMSS/TIMSS Advanced are directed by the TIMSS and PIRLS International Study Center at Boston College in the United States.
3. Beginning in 2011, the PIRLS technical documentation is being provided on an ongoing basis via the 'Methods and Procedures' section of the website (http://pirls.bc.edu).

References

Gove, Amber. 2012. Think global, act local: how early reading assessments can improve learning for all. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.

Lockheed, M. 2012. Policies, performance and panaceas: The role of international large-scale assessments in developing countries. *COMPARE: A Journal of Comparative and International Education* 42, no. 3.
Mullis, I.V.S., and M.O. Martin. 2012. Using TIMSS and PIRLS to improve teaching and learning. *Revue Recherche en Education* 14.
Mullis, I.V.S., M.O. Martin, K.L. Drucker, and C.A. Minnich., eds. 2012. *PIRLS 2011 encyclopedia: Education policy and curriculum in reading, volumes 1 and 2*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Mullis, I.V.S., M.O. Martin, P. Foy, and K.L. Drucker. 2012. *PIRLS 2011 international results in reading*: *Student learning in relation to curriculum, school environment, instruction, and home support*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Mullis, I.V.S., M.O. Martin, A.M. Kennedy, and P. Foy. 2007. *PIRLS 2006 international report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Mullis, I.V.S., M.O. Martin, A.M. Kennedy, K.L. Trong, and M. Sainsbury, eds. 2009. *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Options for developing integrated assessment systems to improve learning and teaching

Anil Kanjee
Tshwane University of Technology

Introduction

The increased emphasis placed on improving the quality of education, in particular learning outcomes, has resulted in a growing recognition within many education ministries to develop or enhance their assessment systems (Wagner 2011). In practice, assessment systems comprise four components: classroom assessments; examinations; assessment surveys; and school evaluations. However, the development of these components, and their effective application, vary widely among countries. With strong will and support from ministry officials and school staff, it is possible to significantly improve any assessment system. Given the context of limited resources and financial support, inadequate capacity as well as relatively short time-periods to institute change, the key challenge for policy-makers is to identify which component (s) of the assessment system to focus on to maximise the returns on any investments made.

This article presents a number of options for policy-makers to consider when developing integrated education assessment systems. These options are derived from grounded professional and academic experiences, consultations with international experts and ministry officials, from a review of the literature, and from recent work conducted for the Russian Aid for Educa-

tion Development (READ) Project (READ. 2009, 2010). The options presented are not exhaustive and should be viewed within the specific development trajectory of an education system, and the different contexts that define how learning and teaching take place within that system.

Assessment system framework

An integrated assessment system can be defined as a group of interrelated or interdependent policies, practices, structures and processes implemented by stakeholders at the different levels of the education system to obtain and apply evidence about learner performance for certifying or improving learning (Kanjee 2008). Most assessment systems comprise the four components mentioned above, and shown in Figure 4, each of which are interrelated and function within the social, political, cultural and economic context that are generally reflected by national policies, institutions as well as systems and structures.

Within this framework, classroom assessment is used to obtain evidence on the knowledge, skills and attitudes of learners for use by teachers to improve learning and teaching. Examinations are undertaken to determine individual candidates' mastery of specific knowledge and skills for the primary purpose of selection, and comprise internal and external examinations. Assessment surveys refer to an external process undertaken to obtain evidence on the performance of an education system (or part thereof) by assessing learners, teachers, school heads and education officials as well as evaluating the functioning of structures and programmes within that system. School evaluations refer to an internal or external process of monitoring and

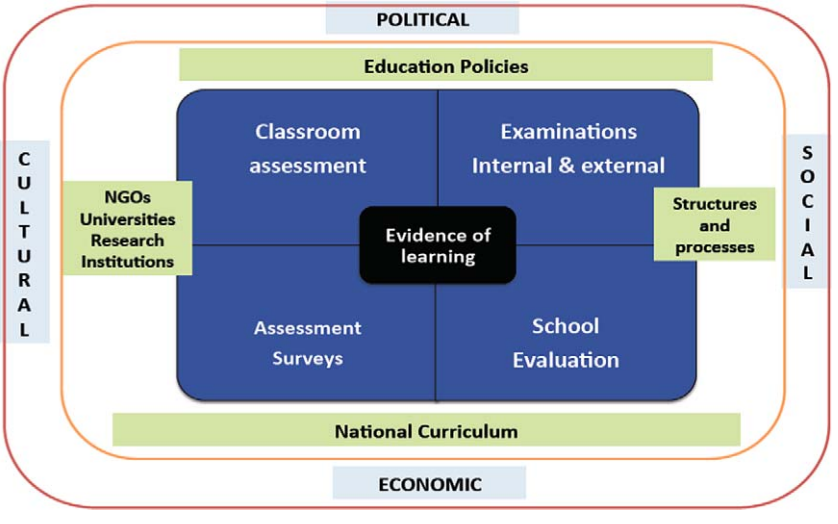

Figure 4. Components of an integrated assessment system.

reviewing the effectiveness and efficiency of a specific school in order to address challenges identified and to improve the performance of the school. While school evaluations do not by themselves constitute an assessment activity, in many instances learner (and teacher) assessments are specifically conducted to evaluate the performance of schools. More importantly, school evaluations provide the avenue within which results from classroom assessments, examinations and assessment surveys can be used to improve the quality of learning and teaching.

Developing effective assessment systems

Developing effective assessment systems that address the information needs of different stakeholders within the education sector is a complex and long-term undertaking that often requires high investments in human and financial resources. Braun and Kanjee (2007) and Crooks (2002) note that this task is best undertaken in stages, especially in countries with limited financial resources and limited technical and practical assessment expertise. These authors note that policy makers should focus on a specific component of the assessment system or on specific aspects that need to be addressed across different components; for example, information technology or staff development. The question, however, is how to identify the key issue or component on which to focus.

To address this question, policy-makers need to review the effectiveness of relevant policies, institutions and structures that support the assessment system, determine the available expertise and experience of staff, calculate current costs and expenditures and identify key challenges and gaps that need to be addressed. To this end, the World Bank has developed a self-diagnostic tool to assist policy-makers identify key areas for intervention to improve the assessment system (Clarke 2010).[1] Over the last two years, this tool has been effectively applied in a number of countries (READ 2010).

For example, based on their self-diagnostic review, the ministry of education in Zambia has instituted capacity development programmes across three components to: enhance staff capacity within the examinations division for item writing and test development; train staff to conduct advance statistical and psychometric analysis on large-scale assessment data; and develop teacher classroom assessment skills to improve the teaching and learning process (READ 2009). Similarly, the ministry of education in Ethiopia prioritised the development of the school evaluation component by implementing a new policy framework, introducing a new structure within the ministry and hiring dedicated staff (READ 2010).

Selected options for consideration

Having identified the specific component/issue for intervention, how should policy-makers proceed, and what are the next steps? A general response that addresses the unique needs of countries is not possible given the range of contexts within which education systems operate and the different and unique roles and uses of assessment systems. Instead, a number of options relevant to the different components of the assessment system are presented to provide a sense of how issues have been approached in various countries.

Classroom assessments

In Vietnam, the ministry has prioritised the development of classroom assessment and participation in international assessments. For the former, the first step was to determine current teacher practices in assessment as well as to identify whether and how teachers were trained in the effective use of assessment. This they did by conducting extensive qualitative studies and reviewing current teacher training programmes (READ 2010).

Similarly, efforts by the department of education (DoE) in South Africa to improve classroom assessment practices focussed on providing teachers with appropriate tools to reduce teacher workload related to test development, item analysis, interpretation, recording and reporting of learner performance (Kanjee 2009). The DoE targeted low performing schools and supplied all teachers with Assessment Resource Banks (ARB), booklets that comprise of a series of 'testlets' based on the national curriculum. These testlets were used to monitor learners' performance and determine whether learners had mastered the content taught. Given the overwhelming success of the ARBs (Kanjee 2009), the ministry has recently launched a computerised assessment system to support teachers improve their classroom assessment practices.

Examinations

Conducting additional analysis on examination results along with effective dissemination and provision of relevant support can lead to improved performance in examinations. For this to occur, Wasanga (2010) argues that examination results should be reported and disseminated to spur discussion and action, not to blame or accuse. In Kenya effective systems of analysis and feedback have been developed to report information to district officers, to schools, to teachers and to the public (Kellaghan and Greaney 2004). Wasnaga notes that reports are tailored to specific audiences and are intended to provide information for use in improving learning and teaching. These reports provide: (1) detailed information on specific content areas where candidates performed poorly; (2) exemplars of actual responses to specific questions taken from high- and low-performing learners; and (3) concrete

advice on how to address difficulties that most candidates exhibited in their responses (Wasanga 2010).

In addition, ministry officials also provided workshops and seminars and visited schools, particularly those with poor results, to support teachers and principals improve learning. Furthermore, Kellaghan and Greaney (2004, citing Eisemon 1990) note that when guidance was provided to teachers in the preparation of students for examinations, and teachers understood the demands of examinations, the inclusion of tasks requiring higher-order skills in examinations led to greater emphasis on teaching these skills in classrooms.

Assessment surveys

In Uruguay, the ministry of education effectively used assessment information from their census surveys to improve learner performance, with largest gains reported in schools serving low socio-economic learners (Ravela 2005). The ministry instituted a programme to improve learning by combining the use of assessment information with teacher professional development and school-based interventions. The programm was conducted over a number of months, and comprised: (1) the use of learner scores obtained from a census assessment survey to identify poor performing schools; (2) facilitating bi-weekly teacher meetings, held during weekends, to interrogate school results and provide teachers with the necessary training for implementing interventions that address specific needs of their school; and (3) conducting follow up sessions to monitor and review progress in the implementation of interventions (Ravela 2005).

In an effort to improve the capacity and skills of staff, Chulu (2009) notes that the ministry of education in Malawi embarked on a capacity development programme by supporting large numbers of staff to attain masters and doctoral qualifications in assessment studies at select overseas universities. The participation of ministry officials as well as academics from teacher training institutions resulted in strengthening of the teacher training programmes and improved the assessment skills based within the ministry (Chulu 2009). Uganda, on other hand, used donor support to enhance the capacity of staff within the existing Examinations Council in the management, reporting and utilisation of data from large-scale assessment surveys (Kanjee and Acana 2010). Today, an independent unit, comprising only Uganda nationals, conducts all large-scale assessment studies in the country.

School evaluations

In Vietnam, all schools are required to undergo an intensive self-evaluation process at least once every four to five years to identify areas in need of improvement and to develop intervention strategies to improve learning and teaching in the school (MOET. 2009). These evaluations are conducted by

school staff themselves (i.e. management and teachers), and comprise a critical review of all aspects of the functioning of the school using assessment results and other relevant information. The process is generally conducted over a six-month period and culminates in a detail report that is presented to district or provincial officials for review and approval. Support for conducting these evaluations is provided by school evaluation directorates that are located in all ministry of education and training provincial and district offices. While school heads and teachers are tasked with implementation of interventions, the school evaluation directorate is responsible for monitoring progress made by schools (MOET 2009). In addition, schools can also volunteer to participate in the national evaluation process that is conducted by an external team. School that meet the stringent national criteria accrue significant benefits that include additional funding, infrastructure development and national prestige.

Conclusion

It is widely accepted that assessment systems can provide valuable information for use in improving learning and teaching. A challenge that still exists, however, is how to develop and sustain effective assessment systems to address the specific needs of teachers, policy-makers and parents, all of whom are striving to improve the quality of learning and teaching. Given the large variation in the contexts within which teaching and learning take place, this challenge is best addressed by taking into account the specific circumstances that impact on any education system, in particular the stage of development of any assessment system, the specific goals of the country, as well as the availability of relevant human and financial resources.

There are, however, a number of assessment issues that must be addressed for the development of effective assessment systems in any country. Foremost among these include: to obtain, report and disseminate reliable and valid assessment information to key education role-players in a timely manner; to ensure the effective use of assessment information to address key challenges of equity within the education system; to regularly review the costs and expenditures for maintaining the assessment system; and to determine return on investments (Wagner 2011), and to monitor all aspects of the assessment system to ensure assessments conducted are effectively utilised to attain the goals of improving learning for all.

Note

1. The tool is available from http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTEDUCATION/0,contentMDK:22808581~menuPK:282391~pagePK:148956~piPK:216618~theSitePK:282386,00.html.

References

Braun, H., and A. Kanjee. 2007. Using assessment to improve education in developing nations. In *Educating all children: A global agenda*, ed. J.E. Cohen, D.E. Bloom, and M.B. Malin, 303–53. Cambridge, MA: MIT Press.

Chulu, B.W. 2009. Institutionalisation of assessment capacity in developing nations: The case of Malawi. Paper presented at the 3rd global meeting of the Russian Aid for Education Development programme, 30 September–2 October, in Kenya.

Clarke, M. 2010. Roadmap for building an effective assessment system. Unpublished draft. Washington, DC: World Bank.

Crooks, T. 2002. Educational assessment in New Zealand schools. *Assessment in Education* 9, no. 2: 217–37.

Kanjee, A. 2008. A framework for developing national education assessment systems: Challenges and prospects. Paper presented at the National Policy dialogue forum on developing effective assessment systems, November 19–21, in Pretoria.

Kanjee, A. 2009. Enhancing teacher assessment practices in South African schools: Evaluation of the assessment resource banks. *Education and Change* 13, no. 1: 67–83.

Kanjee, A., and S. Acana. 2010. *Creating an enabling environment for an effective assessment system in Uganda*. Paper presented at the 2nd global meeting of the Russian Aid for Education Development programme, October 4–6, in Kenya.

Kellaghan, T., and V. Greaney. 2004. *Assessing student learning in Africa*. Washington, DC: World Bank.

MOET. 2009. Vietnam self-diagnosis report: Ministry of education and training. Unpublished report. Hanoi: World Bank.

Ravela, P. 2005. A formative approach to national assessments: The case of Uruguay. *Prospects* 35, no. 1: 21–43.

READ. 2009. *Annual report of the Russian Aid for Education Development Trust Fund: 2009*. Washington, DC: World Bank.

READ. 2010. *Annual report of the Russian Aid for Education Development Trust Fund: 2010*. Washington, DC: World Bank.

Wagner, D.A. 2011. *Smaller, quicker, cheaper: Improving learning assessments for developing countries*. Paris: UNESCO-IIEP; Washington, DC: FTI.

Wasanga, P. 2010. *The role of examinations results as a feedback tool: The Kenyan experience*. Paper presented at the 2nd Global meeting of the Russian Aid for Education Development programme, October 4–6, in Kenya.

Think global, act local: how early grade reading assessments can improve learning for all

Amber Gove
RTI International

Introduction

Since the global education community set forth the Education for All (EFA) goals in 1990, efforts have centred on the most measureable and concrete of the six targets: improving access to and completion of primary education.

Universal primary enrolment became the rallying cry of the last two decades, with considerable success. Some 40 million more children have entered school since 1999, yet a total of 67 million children remain out of school. By any measure, this is excellent progress: the enrolment gap between low-income and high-income countries has nearly closed, as shown in Figure 5. More elusive, however, is the goal of universal primary completion – at the current rate, closing the gap between rich and poor countries would take an additional 30 years, long past the target date of 2015 (see Figure 6).

With all of these additional children in school, why has primary completion lagged behind? The missing link is learning. Primary school learning,

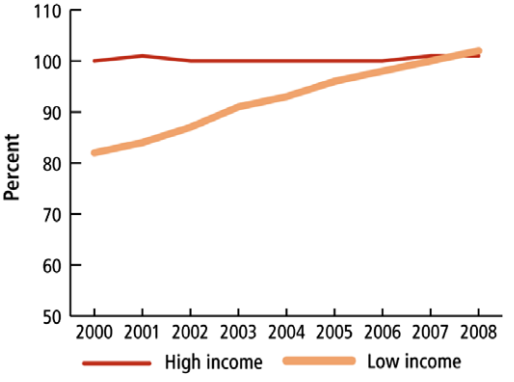

Figure 5. Gross primary enrolment in high-income and low-income countries, 2000–2008.
Source: World Bank. (2010).

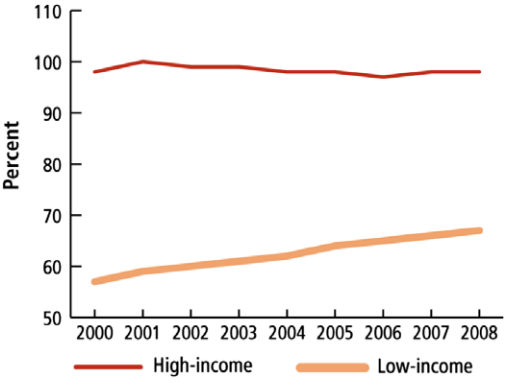

Figure 6. Primary school completion in high-income and low-income countries, 2000–2008.
Source: World Bank (2010).

vaguely defined in EFA Goal 6 as 'improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills' has received relatively little attention. One key reason for the relative lack of emphasis on learning at the global level has been the absence of clear measureable targets for learning in the primary grades, in particular for the poorest countries. Existing international assessments, the main source of global level learning targets, typically measure learning at the end of the primary cycle or later – fourth and eighth grade in the case of TIMSS and PIRLS, 15-year-olds in the case of PISA. For many low-income countries, these assessments occur too infrequently (most are on a five-year cycle) and too late for most of their students, the vast majority of whom are enrolled in the early grades and repeat or drop out before reaching upper primary. In the poorest countries, including those with the highest number of out of school children, grade repetition is frequently the norm – with many low performing students caught in a vicious cycle of school failure and eventual drop out (Gove and Cvelich 2011). High-stakes national examinations are themselves a source of repetition as failing students frequently opt for another chance at passing (Eisemon et al. 1993). Those students who do pass are not guaranteed to have the required skills to succeed; most national examinations in developing countries demand a high level of factual recall, but little in the way of important initial competencies or critical thinking skills (Somerset 2011).

Measurement matters: moving from access to learning

In an effort to fill this gap in measurement, in 2006, the United States Agency for International Development and the World Bank supported the development of a framework for assessing reading skills in the early primary grades, defined as grades 1 through 3. RTI International, a non-profit research think tank, was charged with drafting and testing a protocol for a simple oral assessment that could capture a range of students' foundation skills in reading across a diversity of alphabetic languages. Accompanying parent, student, teacher and school context questionnaires and observation tools probe language and instructional practices, including effective use of classroom time, both in and outside the classroom. An independent international panel of reading specialists, testing experts, teacher educators and linguists reviewed and proposed changes to the protocol; pilot testing in several languages and subsequent review of results took place in early 2007.

The resulting approach, called the Early Grade Reading Assessment (EGRA) differs markedly from other international assessments. Comprised of eight to ten modular subtasks based on principles common to alphabetic languages, the numerous adaptations of EGRA adhere to a (reasonably) global framework for assessment rather than following a strictly standardized

assessment model (e.g. there is no attempt to translate a single set of questions or items). EGRA is therefore not fully comparable in the traditional psychometric sense across different languages or tests, as each one is locally tailored and developed. Instead, any cross-national or language comparisons must derive from data indicating the proportion of students meeting locally established goals, with reading for understanding as the critical objective for all students. For example, if the Gambia set a target of 45 correct words per minute with 80% comprehension, 5.7% of grade 2 students would have met that target in 2011. With additional calibrated assessments applied every year, stakeholders interested in tracking progress would be able to compare the share of students reaching the instructional target.

This 'think global, act local' approach aligns with ongoing efforts to establish global-level learning indicators that would require countries to measure the percentage of children meeting locally set targets. In 2006, Filmer, Hasan and Prichett called for a 'millennium learning indicator', while a 2005 World Bank report, *From schooling access to learning outcomes: The unfinished agenda*, decries decades of input-focused investments without attention to learning (World Bank Independent Evaluation Group 2006).

With the mindset that measurement matters, in 2009 the Fast Track Initiative (now the Global Partnership for Education) set forth draft indicators for reading outcomes. Revised in 2011, the indicators ask countries to track and report on:

- The proportion of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text;
- The proportion of students who, by the end of the primary or basic education cycle, are able to read and demonstrate understanding as defined by the national curriculum or agreed upon by national experts.

A key message behind these targets is that countries should measure and establish their own transparent learning targets in accordance with local curricular goals, but with global level monitoring and support for measurement. The emphasis on local ownership of both goals and metrics, coupled with timely application of appropriate tools for measurement and improvement, are promising from both a political and practical standpoint – countries invested in the process of setting their own goals are more likely to measure and track them, as well as make efforts to improve, with the eventual goal that all children meet the locally established benchmarks. Notably absent is any accommodation for instruction in a second language; countries that use mother tongue instruction for initial literacy acquisition are held to the same standard as those who do not, implying that good instruction, regardless of language, can and should be the norm for all students.

How are countries measuring up?

As of October, 2011 early grade reading assessments have been conducted in more than 50 countries and some 70 languages for a variety of different purposes. In a recently published volume (Gove and Wetterberg 2011), the author and colleagues describe the various approaches and uses of EGRA – from teacher-administered classroom monitoring tools of individual student progress to sample-based national- or regional-level assessments. Each approach varies in its construction and sampling framework and the open-source nature of EGRA has led to a proliferation of uses and applications. Expansion of EGRA has been facilitated by the guidance regarding the development and appropriate use of the tool (RTI 2009; RTI and IRC 2011) as well as an on-line community of practice offering resources and advice for potential users (http://www.earlygradelearning.org).

At the national or regional level, results from 14 countries are presented in Figure 7 (to date, RTI has conducted EGRA in 20 countries, the remaining countries have either not yet released their data for public use [Morocco, Rwanda, Yemen, Zambia] or were small, project-level samples [Egypt, South Africa]). The figure presents the percentage of students unable to read a single word of text on the paragraph reading portion of the assessment. In each case the child was tested in the language in which they were being taught (only one language per country is presented here – additional language results are available for Ethiopia, Haiti, Kenya, Mali, Nicaragua, Nigeria, and Uganda at http://www.eddataglobal.org). In the majority of countries, the language of instruction (and assessment) is not the child's first language.

As can be seen in Figure 7, in 6 of the 14 participating countries, more than half of students tested in one of the early grades were unable to read a

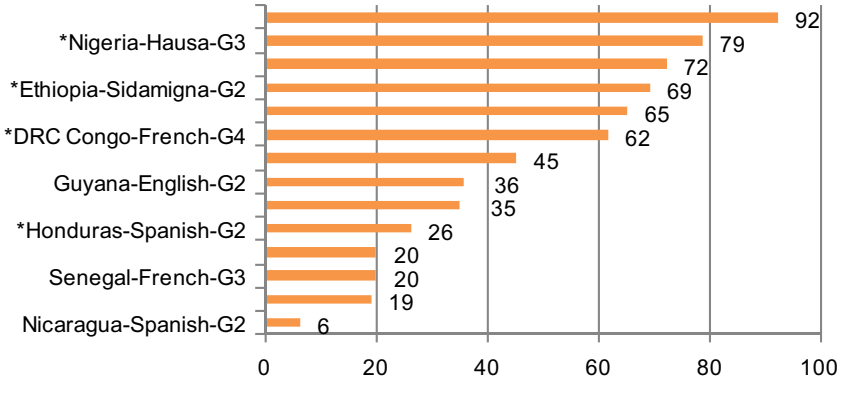

Figure 7. Percentage of non-readers, by grade and language.
Source: EGRA reports 2008–2011, available by country on http://www.eddataglobal.org (see references).
Note: *Results are not nationally representative.

single word of a simple paragraph. While language structure differences between the assessments complicate direct comparisons of fluency (correct words per minute) across countries and languages, zero scores can be reasonably compared as the expectation is that the inability to read at all would not be substantially affected by test differences. Reporting on these zero scores, or 'non-readers', is also a simple way to communicate the data in the absence of country-level standards required for indicators such as those above. Moving forward, countries will ideally establish standards and allow for reporting on the percentage of students reaching locally established goals.

In addition to reporting on the share of students unable to read at all, cross-national comparisons have also been made based on the percentage of grade 2 students correctly answering at least 80% of the comprehension questions in the assessment. Unfortunately very few countries report promising results on this indicator. Of the 13 countries included in Figure 7, only Nicaragua had more than half of its grade 2 students reading with understanding at this level.

What are countries (and donors) doing about these results?

The silver lining behind these mostly dismal results is that many countries are moving quickly to improve. The following rigorously evaluated experiences, including pre- and post-test results and control and treatment groups, are from countries that have undertaken EGRA or EGRA-like approaches and have acted on the results to improve student learning.

- In the Gambia, national-level assessment undertaken by the government in 2007, 2009 and 2011 show considerable progress toward reducing the number of grade 2 non-readers from 57% in 2007 to 42% in 2011. Similar gains were made in nearly all sub-tasks and grades. Following their first assessment in 2007, the government embarked on a significant effort to improve reading, including training more than 3,000 teachers and revising their early reading curriculum and adoption of a tightly structured instructional programme, including the Jolly Phonics curriculum. In the preface to their teacher training manual, the authors state that the dismal EGRA results spurred Gambian education officials to 'embark on an immediate intervention to remedy the situation' (Department of State for Basic and Secondary Education 2007, 5). According to the Permanent Secretary, EGRA was unique in clarifying the scope of the challenge. As he reported to the author, he knew that quality was an issue but it was not until he asked students to read aloud that he came to understand the severity of the problem (Bouy 2007). On the basis of these results, the Gambia is currently developing a

mother tongue reading programme designed to build student literacy in their first language before transitioning to English.

- In Liberia, results of a national-level EGRA in 2008 led to the design and implementation of a randomized control experiment. Two levels of intervention – a 'light treatment' and a 'full treatment' – were applied in schools and then compared against a control group that followed the standard reading instruction approach in Liberia, with each group consisting of 60 schools. In the light treatment group, designed to test an 'accountability' hypothesis, reading levels were tested and schools were informed of the results and shown how to share them with the community through report cards. In the full treatment group, reading levels were assessed; parents and communities were informed and involved in improving student literacy through home-based activities; teachers were trained on how to continually assess student performance; and teachers were provided frequent school-based teaching support, specified lessons plans, resource materials and books for students to use in class and take home. Students in the full treatment group outperformed their peers in all reading skills, nearly tripling the gains made by the control group in oral reading fluency and reading comprehension. The full treatment group also increased non-word fluency sevenfold, indicating a particularly large impact on improving children's decoding – the ability to break new words into sounds and link them together – which is a key intermediate step to unlocking fluency and comprehension. The overall effect size of the intervention was 0.79 standard deviations (SD), considered very large for social science efforts (Piper and Korda 2011). The programme is currently under expansion to some 1,300 schools with USAID and ministry support.
- In South Africa, the Systematic Method for Reading Success (SMRS) programme implemented by RTI in collaboration with the Molteno Institute of Language and Literacy and South Africa's Department of Education focused on effective use of classroom time, mother-tongue instruction following a clearly specified scope and sequence, simplified materials and use of step-by-step daily lesson templates. Materials were developed through two development workshops with ongoing follow-up throughout the programme, resulting in a teachers' manual (covering 25 lessons), learner stories, a programme progression chart, vowel and alphabet charts and mastery tests. Teachers were provided two four- to five-day training workshops and bimonthly school-based observation and support visits. The format of the first training workshop was demonstration and then practice of the programme's seven steps: (1) re-reading of familiar story; (2) phonemic awareness; (3) phonics; (4) decoding/blending of sounds into words; (5) sight words; (6) teacher reads aloud; and (7) independent reading. The second workshop focused on reflection of mastery test data and development

of action plans to strengthen instructional weaknesses. Implemented for less than a year, this activity showed that even in such a short period, children in treatment schools could acquire reading skills, including comprehension, two to three times faster than children in control schools. Effect sizes showed an impact of 0.8 SD, a very promising result considering it was produced in less than one year. The programme is currently being expanded by Molteno; expansion by the department of education is anticipated.

- In Egypt, the Girls' Improved Learning Outcomes (GILO) project conducted an early grade reading assessment in Arabic with students in grades 2, 3 and 4 in 2009. The assessment revealed that half of grade 2 students could not identify a single letter sound, a critical skill for decoding and reading words. To address this, RTI and partners designed a package of lesson plans and materials and trained teachers to teach phonics using cognitive engagement techniques. The project conducted a follow-up EGRA at the end of the school year to measure the impact on student learning relative to a set of matched control schools. Results indicated that the enhanced reading instruction resulted in significant improvement for the GILO-supported schools: on average, students identified 19 more letter sounds per minute at the end of the school year, 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. The impact on students' reading fluency was also dramatic: an average of 10 more words read per minute – an increase of 82% over baseline – compared to three more words read per minute among the control group – an increase of 38% over baseline. Based on the results of the experiment, the Egyptian ministry is currently planning a national roll out and modification of the curriculum to include phonics-based support.

Moving forward

The cases above are just a few examples of how EGRA and EGRA-like assessments, including household-based assessment such as those conducted by Pratham and Uwezo, have provoked education stakeholders and decision-makers into action. Government and civil society efforts to improve early reading in Ethiopia, Guyana, India, Kenya, Mali, Mexico, Nicaragua and Peru have also been significant following similar assessment activities. Perhaps this is a growing trend – one that, more than 20 years on from the historic agreements made in Jomtien, signifies convergence (and a sense of urgency) around the need for global-level learning indicators, with reading as a possible first target area. The *Global Compact on Learning* by the Center for Universal Education at Brookings, makes a similar call for global

level learning targets and improvement and the urgency behind the need to shift to an emphasis on learning:

> The case for a new Global Compact on Learning is both urgent and compelling…. We cannot afford to wait another generation to ensure that all girls and boys are learning and equipped with the skills, knowledge, and competencies needed to live healthy, safe, and productive lives. (Center for Universal Education. 2011, 4)

As described above, the progress made by the EGRA tools to date show that it is possible to generate considerable interest around learning improvement in a relatively short amount of time. EGRA and EGRA-like tools are a viable complement to the standard international assessments, though their apparent simplicity requires caution as results can be interpreted too narrowly. For this reason, the centrality of reading with understanding should always be stressed. With so many countries at different points on the path to development, expansion of the available assessment options can only help to serve the needs of the most disadvantaged. Assessments that generate action, whatever their design, are urgently needed to ensure the true promise of education: that all children learn.

References

Bouy, Baboucarr. 2007. Personal communication.

Center for Universal Education. 2011. A global compact on learning: Taking action on education in developing countries. Washington, DC: The Brookings Institute. http://www.brookings.edu/~/media/Files/rc/reports/2011/0609_global_compact/0609_global_compact.pdf (accessed 6 October 2010).

Crouch, L., and Korda, M. 2008. EGRA Liberia: Baseline Assessment of Reading Levels and Associated Factors. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubdetail&ID-158 (accessed 8 March 2012).

Department of State for Basic and Secondary Education. 2007. *Handbook on early grade reading abilities*. Banjul: Department of State for Basic and Secondary Education, The Gambia.

Eisemon, T., et al. 1993. Providing quality education when resources are scarce. Strategies for increasing primary school effectiveness in Burundi. In *Effective schools in developing countries*, ed. H. Levin and M. Lockheed, 130–57. London: Falmer Press.

Filmer, D., A. Hasan, and L. Pritchett. 2006. A millennium learning goal: Measuring real progress in education. Washington, DC: The World Bank. http://www.cgdev.org/content/publications/detail/9815 (accessed 3 October 2010).

Gove, A., and P. Cvelich. 2011. Early reading: Igniting education for all. A report by the Early Grade Learning Community of Practice. Revised edition. Research Triangle Park, NC: Research Triangle Institute. http://www.rti.org/pubs/early-reading-report-revised.pdf (accessed 5 October 2010).

Gove, A., and A. Wetterberg, A., eds. 2011. The Early Grade Reading Assessment: Applications and interventions to improve basic literacy. RTI Press Publication No. BK-0007-1109. Research Triangle Park, NC: RTI Press. http://www.rti.org/publications/rtipress.cfm?pub=17752 (accessed 25 September 2010).

Global Partnership for Education. 2009. Learning outcomes. Washington, DC: Global Partnership for Education. http://www.educationfasttrack.org/themes/learning-outcomes (accessed 27 September 2010).

Piper, B., and M. Korda. 2011. EGRA Plus: Liberia program evaluation report. Prepared by RTI for USAID. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=283 (accessed 13 September 2010).

RTI International. 2009. Early Grade Reading Assessment toolkit. Prepared for the World Bank, Office of Human Development, under Contract No. 7141961. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149 (accessed 23 September 2010).

RTI International and International Rescue Committee. 2011. Guidance notes for implementing EGRA. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=318 (accessed 23 September 2010).

Somerset, A. 2011. Strengthening educational quality in developing countries: The role of national examinations and international assessment systems. *Compare: A Journal of Comparative and International Education* 41, no. 1: 141–4.

World Bank, Independent Evaluation Group. 2006. From schooling access to learning outcomes–an unfinished agenda: An evaluation of World Bank support to primary education. Washington, DC: The World Bank. http://www-wds.worldbank.org/servlet/WDS_IBank_Servlet?pcont=details&eid=000160016_20060911115958 (accessed 29 September 2010).

World Bank. 2010. EdStats query. http://go.worldbank.org/85XM5TBQA0 (accessed 11 August 2010).

RTI International. 2011. *Democratic Republic of Congo PAQUED Early Grade Reading Assessment baseline report*. Research Triangle Park, NC: RTI International.

RTI International. 2010. Ethiopia Early Grade Reading Assessment regional findings. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=291 (accessed 1 November 2011).

RTI International. 2008. The Gambia final Early Grade Reading Assessment report. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=116 (accessed 1 November 2011).

RTI International. 2008. Guyana Early Grade Reading Assessment final report. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=267 (accessed 1 November 2011).

RTI International. 2008. Haiti Early Grade Reading Assessment final report. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=267 (accessed 1 November 2011).

CIASES and RTI International. 2009. Honduras diagnóstoco de capacidades de lectura: Informe sobre resultados del estudio en escuelas PROHECO. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=327 (accessed 1 November 2011).

RTI International. 2010. Kenya Early Grade Reading Assessment findings snapshot. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=276 (accessed 1 November 2011).

RTI International. 2010. *Malawi Early Grade Reading Assessment report: National baseline*. Research Triangle Park, NC: RTI International.

RTI International. 2009. Mali Early Grade Reading Assessment report, French and Arabic. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=313 (accessed 1 November 2011).
CIASES and RTI International. 2008. Report on the results of the Early Grade Reading Assessment 2008 pilot assessment in Nicaragua. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=198 (accessed 1 November 2011).
RTI International. 2011. Nigeria Northern Education Initiative (NEI): Results of the Early Grade Reading Assessment (EGRA) in Hausa. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=340 (accessed 1 November 2011).
RTI International. 2009. Senegal Early Grade Reading Assessment 2009 final report. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=250 (accessed 1 November 2011).
RTI International. 2010. Uganda Early Grade Reading Assessment findings report: Literacy acquisition and mother tongue. Research Triangle Park, NC: RTI International. https://www.eddataglobal.org/data/index.cfm?fuseaction=pubDetail&ID=293 (accessed 1 November 2011).
UNESCO. 1990. World Declaration on Education for All and Framework for Action to Meet Basic Learning Needs. UNESCO. New York. http://www.unesco.org/education/pdf/JOMTIE_E.PDF (accessed 8 March 2012).
UNESCO. 2000. The Dakar Framework for Action. Education for All: Meeting our Collective Commitments. UNESCO: Paris, France. http://unesdoc.unesco.org/images/0012/001211/121147e.pdf (accessed 8 March 2012).

An NGO perspective on assessment choice: from practice to research to practice

Amy Jo Dowd
Education Research at Save the Children

Like many international actors involved in realizing Education for All, Save the Children (SC) has taken a closer look at the skills of primary school children in schools it supports. That assessment evidence has transformed the basic education programming of SC, offering evidence to drive a refocus upon learning. This article explores how assessment evidence changed 'business as usual' for SC basic education programmes by instigating shifts in its approach to technical guidance, national implementation, advocacy and equitable impact.

SC conducted four early reading/maths assessments in 2007 and 2008 in partnership with USAID's EQUIP 2 programme. The results from Haiti, Nepal, Guatemala and Ethiopia made it clear that, as in many primary school systems across the globe, children in SC-supported schools were struggling to master basic skills. This evidence fuelled the development of the Literacy Boost Toolkit – a set of tools that guides country teams, their government and local NGO partners to refocus on ensuring children learn to

read with comprehension in the first years of schooling. Literacy Boost builds on four key research-based principles:

(1) Reading development is essential in the early grades of primary school and entails the development of letter knowledge, phonological awareness, fluency, vocabulary and comprehension skills (Snow, Burns, and Griffin, 1998).
(2) Reading is complex. It is a cognitive, social and cultural activity. Its development leads to literacy, which is at once an individual competence, a social act and a cultural tool (Wagner 2010; New 2001).
(3) Ensuring children's active participation in classroom sessions and via practice in the home is essential to learning to read. The real, predictive power of motivation to read must be built and sustained using child-centred and active learning approaches that ensure progress to and success in higher levels of education (Pang et al. 2003; Snow Burns, and Griffin 1998).
(4) For optimally effective programming, parents and teachers must collaborate both inside and outside the school walls. Children's literacy development happens in schools and homes, and requires materials. It depends on both teachers and parents and on finding the means to enrich the type and amount of reading materials in children's lives (Goldenberg 2001; Hood, Conlon and Andrews 2008).

SC's Literacy Boost interventions target both teachers and communities, putting these principles into practice in nine monthly, interactive teacher training sessions, community activities that support children's reading skills development – and concrete roles for both literate and illiterate parents in supporting their children's learning – and the development of ample, appropriate, quality reading materials. Thus, the first impact of assessment evidence on SC was improved technical guidance for more effective investment and practice in basic education. Further, over subsequent years, SC has reorganized its basic education portfolio to focus all basic education programmes on learning outcomes.

In each country of Literacy Boost operation, SC implements a proof of concept phase (in 10–20 schools) wherein the baseline assessment data inform the adaptation of Literacy Boost teacher and community interventions – such as book banks, reading camps, reading buddies and parent workshops – to the observed needs in each programme site. After a period of intervention, a further assessment allows SC and partners to review progress, shift priorities, retarget interventions and generate advocacy messages for dialogue with government and donors. Using this cycle, Literacy Boost has demonstrated significant impact on: letter knowledge and single word reading in three months' intervention in Ethiopia; across multiple reading skills and numeracy in one year's intervention in Nepal; across reading

skills in two languages in one year's intervention in Pakistan; and on the average Portuguese reading skill levels of incoming third graders with two years' intervention in Mozambique (Cao et al. 2011; Pinto 2010; Mithani et al. 2011; Mungoi. et al. 2011). Assessment results communicate the impact of these joint teaching and community interventions to all – from parents to ministers – and analysis available within months arrives in time to inform the next intervention phase. As highlighted by Wagner (2011), this near-immediate feedback on learning enables SC and partners to quickly enter dialogue to adjust interventions when results are not favourable to local stakeholders, while advocating for improvements where results prove the impact on learning. The second impact of assessment evidence on SC therefore has been to improve both implementation and national advocacy.

The reading measures that SC uses serve multiple stakeholders, an implementing NGO and its closest partners – local district education offices, local NGOs, communities, teachers and parents – and have revealed unanticipated trends in reading skills development. The Literacy Boost Assessment has a practitioner's focus that both enables direct linkages between measures and programme responses while enhancing face validity with ministers, teachers and parents. It establishes the range of children's skills on which to build by exploring what skills children possess – beginning from emergent literacy and the percentage of the alphabet mastered – and using that detail to inform interventions. For example, it allows children to read beyond the 60-second time limit of typical EGRA studies (RTI 2010). This alternate approach to measuring accuracy without a fixed time limit has led to important findings about individual differences in the Philippines, Ethiopia, Yemen, Malawi and Nepal, where there are slow, meticulous readers who demonstrate full comprehension; while in Guatemala, Vietnam, Pakistan and Mozambique there are fluent rapid readers with little or no comprehension (Dowd 2011). So the third impact of using assessments at SC has been to provide a significant counterpoint to donor calls for an exclusive focus on reading fluency and in a time-limited assessment tool. Evidence has driven SC to advocate for reading with comprehension and long-term solutions that ensure children leave primary education with the ability to read to learn.

Finally, the Literacy Boost Assessment gathers data on the local literacy environment among the marginalized populations SC serves. Most other assessments do not collect such detailed data on readers, reading and reading materials in the children's lives. For example, PIRLS looks at home reading habits prior to primary education and the total number of books/children's books. By contrast, the Literacy Boost Assessment considers readers, reading and materials in the life of the child at the time of the assessment with items, including the percentage of family members who read to the child last week and the variety of types of books in their homes. SC believes that such contextual detail is crucial for understanding the impact of age-appropriate books of quality in low-income countries where so few

children's books are available and children's exposure to print prior to schooling is limited.

Investigating trends among readers, reading and materials offers guidance for honing intervention strategies and reveals inequities in the system to address with partners to help all children reach their potential. Baseline analysis shows which of these factor(s) influence reading skills and may be driving inequities. In Pakistan and Ethiopia, for example, baseline data showed a positive relationship between reading habits and skills: the higher the percentage of family members who read to the child last week, the higher the predicted reading score. After intervening to enhance the literacy environment, the relationship described above was still present among comparison students: the more readers at home, the higher the score. Amongst Literacy Boost students the trend was no longer detectable after the intervention: all children showed the same levels of mastery. This result suggests that Literacy Boost produced a benefit for those struggling most to learn – in this case those with a low literacy environment at home (Mithani et al. 2011; Cao. et al. 2011). These children began the school year with significantly lower scores and Literacy Boost helped them to close existing learning gaps. Thus, the final impact of assessments and the evidence-base they produce for programming has been to heighten the level of specificity SC can bring to targeting basic education interventions. SC aims to reach the most marginalized children, and assessments have enhanced how effectively it can do so and have also provided more precise evidence of impact.

Overall, assessment has transformed SC basic education programming. Four years and an estimated $4 million have been invested in understanding why children in poor countries reach grade 5 without basic skills and devising and testing a means to address this crisis. This system of evidence-based programming has led to improved global technical guidance as well as enhanced national partnerships, advocacy and practice. It has also challenged the global debate on early grade reading by insisting on a focus on *reading with comprehension* and on those hardest to reach. Finally, because Literacy Boost employs and generates a continuous cycle of evidence about learning and what influences it, the results have led staff and partners to have a more nuanced understanding of learning. Each new cycle of intervention brings new questions and conversations with colleagues in early childhood development and others in school health and nutrition, about new inquiries into school readiness and linkages between health and learning outcomes. It has also created a new group of innovators thinking through maths learning as well as learning in emergency settings.

In these ways, the choice of a programme-focused assessment and its use in an evidence-based system has altered the way that SC works in basic education. In 2010, its impact directly benefited over 60,000 children in 95 schools in five countries. In 2011, it is benefiting children in more than 640 schools in nine countries, with new countries and schools slated

for 2012. But the central issue is not how large this can grow, but whether and how the lessons of this evidence can be leveraged into action by government and other institutions working to ensure educational quality in each country.

References

Cao, Y., A.J. Dowd, O. Mohammed, S. Hassen, T. Hordofa, F. Diyana, and C. Ochoa. 2011. *Literacy Boost Dendi, Ethiopia: Three-month report*. Washington, DC: Save the Children.

Dowd, A.J. 2011. Fluency, accuracy and comprehension across the globe: Unpacking the utility of 60 words per minute. Presentation to the Comparative and International Education Society, Montreal, May 2, 2011.

Goldenberg, C. 2001. Making schools work for low-income families in the 21st century. In *Handbook of early literacy research*, ed. S.B. Neuman and D.K. Dickinson, 211–31. New York: The Guilford Press.

Hood, M., E. Conlon, and G. Andrews. 2008. Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology* 100, no. 2: 252–71.

Mithani, S., I. Alam, J.A. Babar, A.J. Dowd, and C. Ochoa. 2011. *Literacy Boost Pakistan: Year 1 report*. Washington, DC: Save the Children.

Mungoi, D., N. Mandlante, I. Nhatuve, D. Mahangue, J. Fonseca, and A.J. Dowd. 2011. *Endline report of early literacy among pre-school and primary school children in Elma-supported schools Gaza Province, Mozambique*. Maputo: Save the Children.

New, R.S. 2001. Early literacy and developmentally appropriate practice. Rethinking the paradigm. In *Handbook of early literacy research*, ed. S.B. Neuman and D.K. Dickinson, 245–62. New York: The Guilford Press.

Pang, E.S., A. Muaka, E.B. Bernhardt, and M.L. Kamil. 2003. *Teaching reading*. Paris: UNESCO International Bureau of Education and International Academy of Education.

Pinto, C. 2010. *Impact of Literacy Boost in Kailali, Nepal 2009–2010: Year 1 report*. Washington, DC: Save the Children.

Research Triangle Institute. 2010. *Early Grade Reading Assessment Toolkit*. Washington, DC: The World Bank, Office of Human Development.

Snow, C.E., S.M. Burns, and P. Griffin. 1998. *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Wagner, D. 2010. Literacy. In *Handbook of cultural development science*, ed. M. Bornstein, 161–74. New York: Psychology Press.

Wagner, D.A. 2011. *Smaller, quicker, cheaper: Improving learning assessments in developing countries*. Paris: UNESCO-IIEP and FTI/World Bank.