Learning Assessments

DEVCO B4 Education discussion paper

December 2014

This Briefing Paper is aimed at EU staff working in country Delegations. Its purpose is to provide background information on learning assessments, which are central to improving teaching and learning in schools.

In the global discussion on post-2015 education goals, improving the learning outcomes of the world's children and youth is considered a high priority, alongside the need for equitable access to good quality education. An essential part of improving learning is to be able to measure it, otherwise how do you know if has improved? There is a critical need for better learning metrics and assessments to measure what learners know.

This Briefing Paper provides a brief overview of the key learning assessment issues. It highlights concerns about the development and use of learning assessments at the national level. The annexes provide an overview of the major regional and international learning assessments being conducted mostly at the primary and secondary school levels. While this Briefing Paper does not purport to be a comprehensive guide to learning assessments, it provides an introduction with sufficient breadth and depth to allow non-education specialists to participate in policy discussions on student learning assessments.

Introduction

Evaluating the quality of education begins with understanding one of the main education outcomes: learning. Over the past decade, access to basic education has increased, but often the education system struggles to provide a quality learning experience. Monitoring the quality of education requires identifying what students have learned and what they can do with their newly acquired skills. Testing student knowledge entails the use of student learning metrics and assessments, which are diverse in their scope, breadth and perspective.

Identifying what students have learned, especially as related to the main areas of literacy and numeracy, is essential to guiding policymakers on whether the education system is performing as expected and guiding eventual education reforms.

The results of learning assessments can influence policy and practices from what happens in the classroom to large scale changes and investments by education ministries. At the classroom level, learning assessments can inform teachers and help them adapt their pedagogical skills, beliefs or practices to enhance learning. Policymakers working can use learning assessments to monitor educational outcomes, assess the effect of national policies and identify weaknesses and take appropriate action. **Learning assessments are also essential for supporting international discussions on improving education quality**.[1]

[1] For example, discussions related to the post-2015 development agenda relate that learning must be a cornerstone of the next education priorities, with clearly defined measurement goals.

This Briefing Paper aims to answer those questions often asked by policymakers and education staff when considering the development of learning assessments. What is learning assessment? What options are there in measuring learning? Why embark in a national or international learning assessment? What will it accomplish? What scope and breadth should be considered? How will it be used? And what are some essential priorities to consider?

A brief overview of learning assessments

What is a learning assessment?

Learning assessments measure student performance as they relate to education standards which are usually established and specified in a national curriculum framework.[2] Student learning assessments come in a variety of shapes and sizes. Learning assessments measure student knowledge and performance at a given grade or age, and can be conducted nationwide as well as in distinct sub-national areas.

Three types of learning assessments are used in the education system:

- **Diagnostic assessments** identify the students' skills and abilities against basic norms and standards. These are used to identify learning disabilities or other barriers to learning.
- **Formative assessments** provide teachers with feedback on how their students are performing on a regular basis during the school year. Usually, these include in-class examinations, observations, homework or in-class activities.
- **Summative assessments** measure learning at the end of a period of learning (e.g., a school year, at the end of a course) and assess against the learning expectations from the curriculum. International and national assessments are usually summative assessments.

Countries can choose to develop their own national assessment and/or participate in a larger-scale international learning assessment.[3] But the fundamental characteristics of learning assessments – whether national or international (described in more detail in the following sections) – are often very similar and include the following points:

- Assessments evaluate students using a **paper-based (for older children) or oral test (for younger children)** on a particular subject. Computer-based tests are evolving in some developing countries (see **Error! Reference source not found.**).
- A **selected target group** of students (by age or by grade) are the focus of the assessment. Assessments can be made in transition years, such as the end of primary or lower secondary

[2] Classroom assessments and school diagnostic assessments also assess student learning but are not included in the scope of this briefing paper. They differ significantly and are more focused on individual classroom or school improvement goals.

[3] Other assessment types – such as diagnostic and formative assessments – exist in the education sector but are not the subject of this paper.

education to evaluate student learning in that education level, or at other key points in the education cycle.

- Assessment can be administered to the **entire student population (census-based assessment)** at a given point in the education cycle (e.g. end of primary) or to a **sample population (sample-based assessment)** representative of the entire targeted population.
- Assessing the entire population, often through major public examinations, has the major advantage of providing every child, parent and school with performance information.
- Sample assessments do not provide information on individual pupils and schools; their main application is for policy making.
- Sample testing may be more useful in large countries where testing the entire population is not feasible, financially or logistically. The complexity of the learning assessment can also be a factor in deciding whether to sample the population.
- Assessments **typically evaluate specific cognitive outcomes** – most often reading and numeracy – and are developed based on the curriculum. For younger children of pre-primary school age, instruments assess the child's development (e.g., motor skills, language/cognitive, socio-emotional).
- **Background and contextual information** may be collected through additional questionnaires to students, parents, teachers or school principals. These instruments seek out specific information on the child's learning environment as well as the socio-economic background information which can influence learning outcomes. Considering learning achievement against other socio-economic variables is important in determining policy that directs resources towards certain groups or areas.
- Individual scores are aggregated to achieve **grade-level information.** Student, parent, teacher and school characteristics can be included in analyses to determine their impact on learning. Often, results identify learning disparities within countries (e.g., by wealth, sex, location). Some census-based assessments are used to attach sanctions to school performance.

Standardised instruments can be used for the evaluation of learning acquisition and have been developed by regional alliances (e.g., SACMEQ, LLECE). Traditional forms of assessment (e.g., multiple-choice, fill-in-the-blank, true/false questions) privilege memorization and other rote learning skills that demand a relatively low-level of cognitive effort. Some assessment instruments are more evolved and require applying and processing skills learned in the classroom (e.g., by analysing a text), which cannot be measured with traditional instruments. It is important that the complexity and sophistication of an assessment system does not exceed the capacity to manage it.

Examples of standardised tests include:

- PIRLS 2016 questionnaire are available for download at http://timssandpirls.bc.edu/pirls2016/downloads/P16_FW_Appendix_B.pdf.
- Sample PISA questions are available at http://nces.ed.gov/surveys/pisa/Items.asp?sub=yes&SectionID=1&CatID=1.
- Examples from various national and international assessments http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1222888444288/National_assessment_Vol2.pdf.

High-quality learning assessments are measured by their technical value as well as by how the results are interpreted, disseminated and ultimately used to implement policy changes. For more detail on this important aspect of learning assessments, see Volume 5 of World Bank National Assessments of Educational Achievement Series, link available in Further reading section).

International learning assessments

International learning assessments can be global, regional or simply a collection of countries. The scope of subjects and grades covered can vary among and within assessments. For example, some countries choose to assess both 4th and 8th grades using the TIMSS/PIRLS, but not all. The number of countries included in one assessment ranges from as few as four (Young Lives) to 65 for PISA (OECD).

Regional assessments are usually contained within a region (although PASEC has expanded to include francophone countries in Asia and neighbourhood countries).

Multiple-country assessments are conducted in a small cluster of countries which may not have a geographical or cultural proximity.

Assessment type	International	Regional	Multiple country
Names	EGMA EGRA PIRLS PISA SABER[4] TIMSS	PASEC PERCE/SERCE/TERCE PILNA SACMEQ	ASER ELDS LMTF PRIDI READ Uwezo

The features and objectives of these international assessments vary considerably. While they are introduced in more detail individually in the Annexes, several common points are noteworthy:

- An increasing number of countries have been participated in international learning assessments.
- Most assessments are school-based, although home-based assessments such as the ASER and Uwezo can reach out to include out-of-school populations.
- Most international assessments collect additional background information on students, teachers, principals, schools, communities or parents. Combined with student characteristics, contextual information can elucidate concerns about the impact of the environment, teacher qualifications, parental education and household poverty on learning outcomes. Such information is valuable for guiding policy decisions.

[4] The SABER programme is not technically a national assessment because it does not evaluate students. It is included because it evaluates the national capacity for learning assessment according to a common set of standards.

- EGRA and EGMA are developed mostly as school-based evaluations which aim to provide a sub-national or district level comparison, rather than a national picture of learning, which is the object of larger sample-based international surveys (e.g., PISA, TIMSS, SACMEQ).
- International learning assessments are conducted regularly on a three to five-year scale, although one-off assessments are possible (e.g., PILNA).
- Not all international assessments aim to have international comparability. Limited data are compared in PASEC and SACMEQ.
- Some assessments are easier to implement than others because they require fewer technical processes (e.g., training, national capacity building).

A multiple country longitudinal assessment: Young Lives

Young Lives is a cross-country survey of children funded by the United Kingdom's Department for International Development (DfID), Irish Aid and the Netherlands Ministry of Foreign Affairs. It is a 15-year international longitudinal study of children in four developing countries (Ethiopia, India, Peru and Viet Nam). The study examines childhood poverty, with specific attention to the background and contextual factors that define life trajectories. In each country, 3 000 children, their primary caregivers and community members are surveyed every 3 to 4 years. Additional research components in each country include: qualitative case studies of a sub-sample of 50 children, their parents, teachers and community representatives; a sample questionnaire on the schools attended; and specific context-based studies (e.g., transitions in early childhood). The youngest children were 6 months old and the oldest 8 years old at the time of the first survey in 2002. The last round of questionnaires is expected to be conducted in 2016.

For more information, see:

- Brock, K., 2011. Young Lives: A Multi-disciplinary Longitudinal Study of Childhood Poverty, Young Lives Method Guide.
- Young Lives, 2014. Young Lives: What we do (webpage). http://www.younglives.org.uk/what-we-do.

National learning assessments

National learning assessments are conducted at the country level to evaluate student performance in specific grades or at specific ages. Teachers and school authorities are sometimes asked to fill separate questionnaires to provide background analytical information which might influence learning outcomes.

National learning assessments are created and administered at the national level, under the direction of ministries of education or national evaluation and research institutes. Donors such as USAID and the World Bank have been financing national assessments for many years (in addition to international ones).

An increasing number of countries are performing national assessments. Between 2000 and 2013, 142 countries conducted at least one national assessment.[5]

Many countries have operated national assessments of student performance for many years, which have been well-accepted by national stakeholders. Where this happens (e.g. Ghana, Uganda, Zambia) trends in performance can be tracked over time. Having local control over the assessment strategy, design and analysis strengthens national ownership of monitoring students' learning. In some countries, however, the results of national assessments are not validated or disseminated appropriately and the policy value of a national assessment is diminished.

The intrinsic values of national assessments

- **Low-stake national student assessments** are usually conducted on a large scale, so as to be representative of the student population. They are low-stake because there are no incentives or consequences attached to the performance of those sampled. These assessments are designed to be informative about learning outcomes, classroom practices and other quality-related variables. They do not impact students directly, except for the time taken to administer the test. Low-stakes evaluations are usually implemented as formative assessments, but can also be summative assessments of learning outcomes.
- **High-stake national assessments** are often mandatory in order to complete a given level of education or gain admission to the next level (i.e., summative assessment). These exams are common at the end of primary or secondary education. In some countries, floors or ceilings of "passing" rates are established to determine, respectively, the minimum or maximum share of students who can pass the assessment. For this reason they are high stake: future choices are often conditional upon performance. These guidelines are often set to meet political expectations or reflect education system constraints (e.g., capacity limits at the next level). Because high-stake assessments by definition influence the student's trajectory, they can have a significant effect on teaching and learning. Teachers might feel the pressure to educate more narrowly ("teach to the test") and students might lose motivation in non-assessed subjects.

Selecting learning assessments

Why embark in a national or international learning assessment? What will it accomplish? What scope and breadth should be considered? How will it be used? And what are some essential priorities to consider? These questions are essential to consider before embarking on a student learning assessment. This section addresses these concerns, while providing a comparison between national and international assessments.

[5] World Education Blog, "Improving, not over-hauling learning assessments post-2015," by Nihan Köseleci Blanchy, 19 November 2014, https://efareport.wordpress.com/2014/11/19/improving-not-over-hauling-learning-assessments-post-2015/.

Top priorities when implementing learning assessments

- **Operational costs should not be an excuse to avoid the development of learning assessments.** Assessments are not expensive policy options, relatively speaking. It is often argued that extensive assessments are very expensive to implement. But assessments are shown to be less expensive than other innovations in education reform, which are enacted without the valid data to justify the investment. Assessments can cost less than increasing teacher salaries or reducing class size (World Bank, SABER Framework Paper). National resistance to assessments – for example due to dedicated teaching time, teacher frustration and school labelling – can engender costs of both a political and administrative nature.

- **Sufficient resources are needed, however, to support both technical capacity and dissemination of results.** Often financing is not dedicated to the post-survey phase which allows for data analysis, dataset sharing, report drafting and dissemination of findings. With fewer resources, the data analysis process can be quite lengthy and results appear less useful when they appear two years or so after the data collection. The ASER reports come out less than one year after data collection. Donors supporting assessments could support improved data standardisation (e.g., providing documentation, data quality control) and sharing of results.

- **Assessments need to be comparable over time.** Countries should be able to compare learning assessments over time, so as to analyse the impact of policy changes and provide higher accountability on education reform.

- **Assessments need to include marginalised populations**. This is a problem with assessments conducted in school that do not include a household survey component. Out-of-school children are excluded from such assessments, and efforts such as ASER which are home-based can expand coverage to all children. Obtaining information on key household factors linked to education outcomes such as poverty, ethnicity, location (remote, urban, rural) and gender are also relevant to affecting policy change.

- **Conducting assessments should be linked to education policy change and reform.** The results of an assessment should be independently validated and disseminated in such a way that all stakeholders have access to the data and can analyse them.

Whether to implement a national learning assessment

> *"Regular and reliable national measurement of learning outcomes […] will play a critical role in monitoring progress towards the proposed education goals. There are several reasons why national data are important. First, more frequent and locally-relevant data can be collected through national systems; and second, in addition to providing the primary source of information for countries to track progress towards goals, national and regional tracking serve as the basis for global tracking over time. Finally, for some constructs, the standards required for global tracking may not be met or feasible across all areas of the proposed targets, but may be more feasibly tracked at the national level."*
>
> Post-2015 Education Indicators Technical Advisory Group of the EFA Steering Committee (2014), Towards indicators for a post-2015 education framework, version 2, November 2014, p. 6.

National learning assessment are a valuable tool for monitoring learning outcomes, especially as they relate to national policy standards and curriculum expectations. Countries can choose to conduct their own assessments which specifically analyse the effects of national education policies. They can also use their own timeframe (when and how often), rather than depend on international assessment schedules. Moreover, the lack of timeliness and regularity for some international assessments can preclude the opportunity for sustaining momentum across time in the way an annual national assessment can.

The results of national assessments have often been overlooked in the international sphere as they do not provide the basis for cross-country comparisons. But with an increased global focus on **learning**, national learning assessments are essential to provide context-specific information on national learning outcomes.[6] As such, governments can be held more accountable for education quality, even if their students' learning outcomes are not being compared directly with those of other countries.

Both National and International assessments are considered indispensable for informing national education policy. With the exception of ASER and EGRA which can be administered on an annual basis, most international assessments have several years between phases. This lack of frequency can be a limitation for a country trying to track trends and improve education quality and may warrant the development of a national learning assessment.

[6] Post-2015 global education goals might require national authorities to fine-tune measurable targets, which would give national assessments greater visibility.

Advantages	Challenges
National assessment characteristics	
• Can be cheaper to conduct • More timely, and can be more frequent • Able to evaluate a wider range of subjects, grade levels and school types, based on national needs. • More likely to not have cultural bias in questionnaires, if developed nationally • Can be more fitted to policy concerns on national learning outcomes (e.g., by focusing on equity) • Builds national capacity in statistical and information systems	• Same elevated level of technical capacity and complex methodologies to produce valid results. •
Policy implications	
• Can choose calendar based on national education reform and changes, so as to measure impact	• May not include sufficient background information • Can encourage teachers to teach to the test • Results might be invalid if poor methodology or implementation occurs • Does not implicate cross-country comparison and yield policy activity as a result

Source: Benavot, Aaron, http://www.unicef.org/ceecis/BenavotCAFE(Final).pdf.

The case of Brazil highlights that a national assessment can be a sufficiently robust source of information to implicate significant policy changes in education. Although Brazil is part of PISA, it is the use of the national learning assessment programme which was the motivating factor in improving learning quality.

Prova Brasil – focusing on inequality in teaching and learning

Prova Brasil is conducted every two years to measure mathematics and Portuguese ability among 4th and 8th grade students. The results are combined with enrolment and progression data to create an index for basic education development for every school. Instead of being assessed against an arbitrary national score (e.g., average), a school's performance is measured against its past performance. The overall goal is to reach the average PISA score by 2021.

Two notable positive outcomes include:

- The regular assessment creates a disincentive for schools to automatically promote students, as well as to hold them back (repetition) or encourage dropping out, but rather helps them focus on improving learning.
- The index has been linked to teacher bonus programmes, creating financial incentives to improve teaching and learning.

Prova Brasil results also highlighted the inequality and inequity in the Brazilian primary education. School resources (e.g., teacher level of education, level of ICT, infrastructure, safe school, well-equipped libraries) are distributed unevenly among populations groups and impact student learning. Brazil introduced the results of the national learning assessment into a larger analysis of learning and teaching performance.

For more information, see:

- Bruns et al. (2012), "Achieving World-Class Education in Brazil: The Next Agenda," Washington, DC, World Bank, https://openknowledge.worldbank.org/bitstream/handle/10986/2383/656590REPLACEM0hieving0World0Class0.pdf?sequence=1.
- IADB (2012), "Assessing Educational Equality and Equity with Large-Scale Assessment Data: Brazil as a case study," Technical Note, http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=36744258.

In South Africa, the Annual National Assessments (ANAs) are a recent policy development to improve teaching and decision-making at the school level. Since 2010, the ANAs are linked to national education policy as a primary output in the Minister of Basic Education's Delivery Agreement for Outcome 1 ("to improve the quality of basic education") and provides feedback to parents, teachers and school administrators on student learning.

South Africa's Annual National Assessment

The Annual National Assessment, introduced in 2011, is a census-based assessment, which tests all children enrolled in South Africa in Grades 1 to 6 and Grade 9. The tests focus on literacy and numeracy and are administered in 11 official languages. The development of the ANAs is part of the national education plan to reach the objective that at least 60% of all students achieve acceptable levels of literacy and numeracy by 2014 and at least 90% by 2024.

The ANA is designed to be aligned with the curriculum so that they do not add significantly to the teachers' workload. The assessments also are developed with a diagnostic notion so that teachers can interpret and use the results to improve their teaching and identify those children falling behind. About 80% of teachers followed the guidelines on how to interpret results and used them in class.[7] School managers also use the ANAs to improve decision-making. Rather than compare results among schools, parents use the ANA to identify challenges and hold schools accountable for their children's learning. Responding to poor ANA results, provinces developed literacy and numeracy strategies which included teacher training, intensive teacher mentoring and increased learning resources.

The ANAs have been criticised for their results and the lack of external verification reduces their validity.[8] South Africa also participates in international assessments (PIRLS, TIMSS and SACMEQ), although these results often show South Africa coming in last or nearly.

Whether to participate in an international learning assessment

Being part of a rigorous international assessment programme has shown to be a useful motivating factor for policymakers around the world. By comparing countries of similar economic levels or cultural affinity, the results of cross-country learning assessments can have greater value to stakeholders. This in turn can lead to more effective policy dialogue, and generate more support to enact changes in education policy.

[7] UNESCO (2014), Education for all Global Monitoring Report 2013: Teaching and Learning Achieving Quality for All. Paris: UNESCO.

[8] Nicholas Spaull (2013), South Africa's Education Crisis: The quality of education in South Africa 1994-2011, October, http://www.section27.org.za/wp-content/uploads/2013/10/Spaull-2013-CDE-report-South-Africas-Education-Crisis.pdf. Linda Chisholm and Russell Wildeman (2013), "The Politics of Testing in South Africa," Journal of Curriculum Studies, 45:1, p. 89-100, http://mg.co.za/article/2011-04-08-sa-education-the-poorest-choice.

Benefits of participating in an international assessment	Risks of participating in an international assessment
• Adherence to high technical standards of assessment design, instrumentation, sampling, administration, analysis and reporting • Technical capacity building exercise, as most assessments require national implementation teams • Increase transparency regarding education system outcomes and human capital development in a cross-national context • More likely to be subject to analytic work enabled by internationally available datasets • Highly-validated results, which can be a political benefit when strong performance • Usually can adapt to national context, language or culture and, in some cases, curriculum	• Participation costs are usually elevated • Occur less frequently and calendar not chosen by country. • Highly-validated results, which can be a political risk when weak performance • Cross-country comparisons are complex to defend politically • Disaffection with the international exercise if its assessment framework is of limited relevance and responsiveness to the country joining • Failure to adapt the survey instruments to national context, leading to discrediting of the results by stakeholders • Dissemination calendar is not controlled by national government

Source: Adapted from DfID (2012), "National and International Assessments of student achievement," Student Guidance Note: A DfID practice paper.

Some learning assessment instruments are more accessible and less costly to implement than others. For example, the large international educational assessments (e.g., PIRLS, PISA SACMEQ, TIMSS) require much time and effort to set up, train and implement in a country. Other simpler instruments can be more affordable and easier to implement (e.g., ASER), yet remain large enough to produce valid and robust results on specific population samples. Implementation costs, however, remain mostly a function of the density of the country (large, remote compared to small and compact) and salary levels (data collectors and analysts).

Comparing assessment types

Some countries might question whether to participate in both international and national assessments. **There are advantages and challenges to both types of assessments and they can work in a complementary fashion as well**.

Countries should not forgo participation in one type of assessment at the expense of the other, as they can both serve to enhance the quality of education systems, even at the classroom level. For example, the EGRA and EGMA have been used extensively to improve teaching methods for early literacy and numeracy.

The case of Ghana shows that the combination of both assessment types (national and TIMSS) led to a re-evaluation of the quality of education, because of very poor performance in the TIMSS. Without the national assessment, the government would not have been able to identify problem areas and affect policy change. The positioning of Ghana among the lowest performing countries in the TIMSS served to motivate policy change more effectively than the national assessment.

Ghana – using national and international assessments to enact quality changes

Students in Ghana participated in national and international assessments during the 2000s.

- The National Education Assessment (NEA) assesses students every two years since 2005 in mathematics and English in grades 3 and 6 and allows for comparison across districts and regions. It has been funded notably by USAID.
- Ghana participated in the TIMSS for 8th grade mathematics and science assessments (2003, 2007, 2011), funded by a grant from the World Bank.

Between 2003 and 2011, Ghana's TIMSS results improved in math and science, but in 2011 it was still the lowest performing country among all participating countries (about 15%-25% of students' scores were too low to be estimated reliably). Ghana is not scheduled to participate in the next TIMSS, given that the exam is not sufficiently sensitive (calibrated) to lower performing countries.

The national assessment is able to better capture and identify problem areas than the international assessment given Ghana's relative low performance in the latter. Yet, the NEA content is more limited than the TIMSS data as it does not take into account certain school or student characteristics, and it does not test problem-solving or written expression (as does the TIMSS). The NEA's disaggregated data (e.g., gender, type of school) are evaluated by district to develop policy recommendations.

Although stakeholders are aware of the results of the NEA, the TIMSS results garnered more attention mostly likely because of the international comparison aspect. Policymaking has been responsive to TIMSS results: tracking the impact of reforms on student achievement levels, informing curriculum improvement, and informing other assessment activities in Ghana.

For more information, see:

- SABER country report: http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAS/SABER_SA_Ghana_CR_Final_2013.pdf.

A common criticism of the major international assessments is that they are insufficiently sensitive to developing country contexts: one reason for Ghana's withdrawal from TIMMS. In response to this, the OECD is piloting a PISA for Development assessment in five to seven countries (starting in 2013) "*using enhanced survey instruments that are more relevant to the contexts found in developing countries but which produce scores that are on the same scales as the main PISA assessment".* http://www.oecd.org/pisa/aboutpisa/pisafordevelopment.htm

"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? These factors […] have real costs – political and budgetary."

Wagner et al. (2012), The debate on learning assessments in developing countries, p. 511.

More relevant than the choice of assessment type is the political backing to develop, conduct and analyse results from a learning assessment. There are often many valid reasons for conducting a learning assessment (i.e., curriculum review), as well as less reasonable ones (i.e., political infighting). But if the assessment is to have any value in improving student learning and teaching, then its results must be assessed with implications on teacher training and curriculum reform policies. As such, governments must be favourable to implicating change across various sectors and budget lines to support improvement in learning outcomes.

Further reading on learning assessments

- Altinok et al. (2013), "A New International Database on Education Quality: 1965-2010," http://halshs.archives-ouvertes.fr/docs/00/91/00/62/PDF/DT_2013-5.pdf. A standardised measure made of various learning assessments to track changes over time in developing countries. Also known as the Altinok dataset.

- Australian Council for Education Research, Centre for Global Education Monitoring:
  - Reviews of Learning Assessments, http://www.acer.edu.au/gem/reviews-of-learning-assessments. Valuable collection of information on learning assessments, with individual summary pamphlet per assessment.
  - Learning Assessments at a Glance, Summary table, http://www.acer.edu.au/files/Gem-la-review-Table.pdf. Includes detailed information about sampling methods, administration, reporting and dissemination approaches to major assessment studies.

- Clarke, Marguerite (2012), "What Matters Most for Student Assessment Systems: A Framework Paper," SABER Working Paper Series, Number 1, April, World Bank, http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/SAS/Framework_SABER-Student_Assessment.pdf. Provides theory and evidence on the various important qualities of successful student assessment programmes.

- DfID (2012), "National and International Assessments of student achievement," Student Guidance Note: A DfID practice paper, http://www.heart-resources.org/wp-content/uploads/2012/04/04-Learning-Outcomes-How-To-Note.pdf?062746. A more extensive background document on learning assessments than this paper, which includes detail on administrative, design and technical decisions. Also provides comparative estimates of costs across selected national, regional and global assessments.

- Global Partnership for Education (2014), "Towards Nuts and Bolts for a data revolution," blog entry by Luis Crouch, 21 August 2014. http://www.globalpartnership.org/blog/towards-nuts-and-bolts-data-revolution. Advocacy for the use of learning assessments to improve international education goals.

- SACMEQ training modules for quantitative research in educational planning (2005). Modules 5 (Item writing for tests and examinations) and 6 (Overview of test construction) are particularly relevant for designing national assessments. http://www.sacmeq.org/training-modules.

- Schleicher, Andreas (OECD PISA) (2014), "Ten things policymakers should know about learning goals and assessments," at the 4th READ Global Conference, St. Petersburg, 2014, http://www.worldbank.org/content/dam/Worldbank/Event/education/Day_2_Keynote_2_Schleicher_EN.pdf. Examples of how policymakers should use assessment results.

- UIS (2006), *Teachers and Educational Quality: Monitoring Global Needs for 2015*, http://www.uis.unesco.org/Library/Documents/teachers06-en.pdf.

- UNESCO Education for All Global Monitoring Report (2014), "The challenges and rewards of measuring global learning after 2015," World Education Blog, 3 June 2014, http://efareport.wordpress.com/2014/06/03/the-challenges-and-rewards-of-measuring-global-learning-after-2015/.

- World Bank website, National Assessment of Educational Achievement Series, http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTEDUCATION/0,,contentMDK:22301663~menuPK:5495844~pagePK:148956~piPK:216618~theSitePK:282386,00.html. Different reports to assist countries in the development of national assessments (NA), including introduction to NA, how to develop tests and questionnaires, how to implement a NA, how to use the NA results, and analysing data from NA.

Annexes

Acronyms

AFD	Agence française de développement
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (German Federal Ministry for Economic Cooperation and Development)
CONFEMEN	Conférence des Ministres de l'Education des Etats et Gouvernements de la Francophonie (Conference of Education Ministers of French-speaking Countries and Governments)
DfID	Department for International Development (UK)
ECD	Early Child Development
EMIS	Education Management Information Systems
GIZ	Gesellschaft für Internationale Zusammenarbeit (German Federal Enterprise for International Cooperation)
IADB	Inter-American Development Bank
ICT	Information and Communication Technologies
IEA	International Association for the Evaluation of Educational Achievement (Netherlands)
JICA	Japan International Cooperation Agency
NGO	Non-governmental organisation
NORAD	Norwegian Agency for Development
OECD	Organisation for Economic Co-operation and Development
ORLEAC	UNESCO Regional Bureau of Education in Latin America and the Caribbean
PISA	Program for International Student Assessment
RTI	Research Triangle Institute International (United States)
SPBEA	Secretariat of the Pacific Board for Educational Assessment
UIS	UNESCO Institute for Statistics
UNICEF	United Nations Children's Fund
USAID	U.S. Agency for International Development

Assessment abbreviations

ANA	Annual National Assessment (South Africa)
ASER	Annual Status of Education Report
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
ELDS	Early Learning and Development Standards
LLECE	Latin American Laboratory for Assessment of the Quality of Education
LMTF	Learning Metrics Task Force
PASEC	Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (Programme for Analysing Education Systems of CONFEMEN's member countries)
PERCE	Primer Estudio Regional Comparativo y Explicativo (First Regional Comparative and Explanatory Study)
PILNA	Pacific Islands Literacy and Numeracy Assessment

PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
PRIDI	Programa Regional de Indicadores de Desarrollo Infantil (Regional Programme on Child Development Indicators)
READ	Russia Education Aid for Development Trust Fund
SABER	Systems Approach for Better Education Results
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SERCE	Segundo Estudio Regional Comparativo y Explicativo (Second Regional Comparative and Explanatory Study)
TERCE	Tercer Estudio Regional Comparativo y Explicativo (Third Regional Comparative and Explanatory Study)
TIMSS	Trends in International Mathematics and Science Study