Public Disclosure Authorized

Public Disclosure Authorized

The World Bank

EDUCATION NOTES

February 2012

67346



MEASURING LEARNING:

HOW EFFECTIVE STUDENT ASSESSMENT SYSTEMS CAN HELP ACHIEVE LEARNING FOR ALL

Why Assess Student Learning?

It is not enough for children to be enrolled in school and sitting in classrooms. For the benefits of education to accrue, children must be learning. But how do we measure whether children are learning and what do we do with that information? As governments and development partners strive to improve student learning outcomes, it is vital to develop strong systems for assessing student learning.

The importance of learning assessment is linked to growing evidence that learning drives prosperity. Research finds a one standard deviation increase in scores on international assessments of reading and mathematics is associated with a 2 percent point increase in annual growth rates of GDP per capita (Hanushek and Woessmann, 2007).

Assessment is the process **of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about what to do next in the educational process.**

An assessment system **is a group of policies, structures, practices, and tools for generating and using information on student learning. This system can support a variety of purposes and uses, such as informing instruction, determining progress, and providing data on accountability.**

Research on assessment reveals that the right kinds of assessment activities, and the right uses of data generated by those activities, can contribute to better outcomes—such as better learning and better informed policy decisions (Clarke, 2011). Evidence on assessment shows:

- A link between high-quality, formative classroom assessment activities and better learning outcomes, as measured by student performance on standardized tests.
- A link between countries that have exit examination policies and higher performance levels on international assessments.
- A link between the use of data from large-scale assessments to hold schools and educators accountable and improvements in student learning outcomes.

Challenges to Assessing Student Learning

Today, too few countries have in place the policies, structures, practices, and tools that would constitute an effective student assessment system. This is particularly the case for low-income countries, which stand to benefit most from systematic efforts to measure student learning outcomes—particularly as testing is among the least expensive innovations in education reform (Hoxby, 2002; Wolff, 2007).

While some low-income countries have experimented with standardized assessments of student learning, participation in such assessments has often been ad-hoc—neither integrated into an education strategy, nor sustained over time. One-time assessments may generate shock value and

create an opening for wider discussions of educational quality, but it is sustained assessment systems that allow countries to monitor learning trends and gain a better understanding of the relative contribution of various inputs and educational practices to changes in student learning outcomes.

What Do Assessment Systems Look Like?

Assessment systems tend to be made up of three main kinds of assessment activities, corresponding to three main information needs or purposes:

- **classroom assessments** -- which provide real-time information to support teaching and learning in individual classrooms;
- **examinations** -- which are used to make decisions about the progress of individual students through the education system (such as certification or selection for university entry); and
- **large-scale survey assessments** -- which monitor learning trends and provide both policy- and practitioner- relevant information on overall performance levels in an education system, and contributing factors.

Classroom assessments are those carried out by teachers and students in the course of daily activity. They include a variety of activities, tools and procedures for collecting and interpreting written, oral, and other forms of evidence on student learning. Research shows a strong link between effective classroom assessment activities and better student learning outcomes as measured by performance on standardized tests, with the largest gains being made by low achievers (Black and Wiliam, 1998).

Examinations provide information for crucial decisions about individual students—for example, whether they should be promoted to the next grade level, assigned to a particular type of school or academic program, graduate from high school, or gain admission to university (Greaney and Kellaghan 1995; Heubert and Hauser 1999). The high-stakes nature of most examinations means they exert a backwash effect on the education system in terms of what is taught and what is learned. This in turn impacts the skills and knowledge profile of graduates. Because exams can have negative consequences for individual students—particularly those from disadvantaged groups—their uses and outcomes must be carefully monitored.

Large-scale survey assessments are designed to provide information on system performance levels and related or contributing factors (Greaney and Kellaghan 2008), typically in relation to an agreed set of standards or learning goals. Assessment results inform both educational policy and practice. Examples include international assessments of student achievement levels such as TIMSS, PIRLS, and PISA; regional assessments such as PASEC in Francophone Africa, SACMEQ in Anglophone Africa, and LLECE in Latin America; national-level assessments such as SIMCE in Chile;[1] and sub-national assessments such as state-level tests in the USA and Canada.

Education systems can have very different profiles in these three assessment areas. For example, Finland's education system emphasizes classroom assessment as a key source of information on student learning and draws less on external examinations. On the other hand, China has traditionally placed considerable emphasis on examinations as a means to sort and select from its large student population, and less on classroom assessments (although this is changing.)

A recent study (Darling-Hammond and Wentworth 2010) reviewed the practices of high-performing education systems around the world (e.g., Australia, Finland, Hong Kong, Singapore, Sweden, UK), and found that the assessment systems in these countries:

- Provide feedback to students, teachers, and schools about what has been learned and "feed-forward" information that can shape future learning, as well as guide college- and career-related decision making.
- Closely align curriculum expectations, subject and performance criteria, and desired learning outcomes.
- Engage both teachers and students.
- Prioritize quality over quantity with respect to standardized testing.

[1]TIMSS – Trends in International Mathematics and Science Study; PIRLS – Progress in International Reading Literacy Study; PISA – Programme for International Student Assessment; PASEC – Programme d'Analyse des Systèmes Educatifs (Program on the Analysis of Education Systems); SACMEQ – Southern and Eastern Africa Consortium for Monitoring Educational Quality; LLECE – Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for Assessment of the Quality of Education). SIMCE – Sistema de Medición de la Calidad de la Educación (Assessment System for Measuring the Quality of Education).

Effective Assessment Systems

The effectiveness of an assessment system depends on the quality of the information that it generates, particularly for decision making. Major drivers of information quality are (Clarke, 2011):

- **the enabling context** -- the broader context in which assessment activity takes place and the extent to which that context is supportive of assessment;
- **system alignment** -- the extent to which assessment activities are aligned with the rest of the education system; and
- **assessment quality** -- the technical quality of the instruments, processes, and procedures used for assessment activity.

To be effective, comprehensive systems for learning assessment must feed into education practice and policies that improve student learning.

The enabling context refers to the broader legislative or policy framework for assessment activities; the institutional and organizational structures for carrying out and using the results from assessment activities; the availability of sufficient and stable sources of funding; and the presence of trained human resources. The enabling context is important to get right because it is a key driver of the long-term quality and effectiveness of an assessment system and no assessment system is sustainable in its absence.

System alignment includes the connection between assessment activities and system learning goals, standards, and the curriculum (Fuhrman and Elmore 1994). Alignment involves more than a simple match between what is tested and what is in the curriculum. For example, while the correspondence between a given country's curriculum and what is tested on international assessments such as PISA and TIMSS may be low, an assessment might still be aligned with (and useful for informing) the overall goals of its education system or related reforms underway. Indeed, the use of data from TIMSS, PIRLS, and PISA to identify drivers of performance, and monitor the impact of reforms on performance over time, has been key to the improvement of achievement levels in countries as diverse as Brazil, Jordan, and Poland.

The need for good **assessment quality** applies not only to large-scale survey assessments, but to any kind of assessment activity (AERA, APA, and NCME 1999). If an assessment is not sound in terms of its design, implementation, analysis, reporting, or use, it may contribute to poor decision making with respect to student learning and system quality. Two overarching technical issues for any assessment activity are reliability and validity, that is, whether an assessment activity produces precise data (reliability)—a particularly important consideration for high-stakes examinations and for monitoring trends over time—and whether test scores represent intended values and are used in intended ways (validity). Test score validity can, for example, be threatened by a difference between the language of instruction and the language of testing, which may make it difficult for a child to show what he or she knows and can do. Validity considerations include careful consideration of the consequences of the uses of test scores, including the social, economic, and other impacts on different groups in the population.

Assessment systems, and the activities that comprise them, may be characterized according to four levels of development:

- **latent** -- at the beginning stages,
- **emerging** -- on the way to meeting an acceptable minimum standard,
- **established** -- acceptable minimum standard, and
- **advanced** -- best practice.

Systems that make the shift from *emerging* to *established* (acceptable minimum standard) are distinguished by a concerted focus on reforms, inputs, and practices that strengthen the enabling context for assessment. The framework shown in Table 1 illustrates these development levels (with the exception of *latent*, which basically represents the absence of any assessment activity), based on specific recommended indicators for the enabling context, system alignment, and assessment quality. The indicators are most relevant to examinations and large-scale survey assessment activities, but with some modifications, can be applied to classroom assessment.

Table 1. Levels of Assessment System Development

	Emerging	Established	Advanced
Enabling Context	• No or limited policy framework • Few trained staff; high turnover • Unreliable funding • Unclear institutional structures/ arrangement	• Presence of policy framework • Training programs/trained staff with low turn over • Stable funding • Clear institutional structures/ arrangement	The same as for Established + strong focus on: • Assessment for learning • School-based and classroom assessment • Role of teachers • Innovation and research-based practices
System alignment	• Assessments not fully aligned with learning goals, standards, curriculum • Assessments out of sync with reforms in other areas	• Assessments aligned with learning goals, standards, curriculum • Assessments in sync with reforms in other areas	
Assessment quality	• Limited awareness or application of technical or professional standards	• Awareness and application of technical or professional standards	

Source: M. Clarke, 2011.

References

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 1999. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Black, P., and D. Wiliam. 1998. "*Assessment and Classroom Learning.*" Assessment in Education: Principles, Policy and Practice, 5(1), 7-73.

Clarke, M. 2011. "*Framework for Building an Effective Student Assessment System*." READ/SABER Working Paper. Washington, DC: World Bank.

Darling-Hammond, L., and L. Wentworth. 2010. "*Benchmarking Learning Systems: Student Performance Assessment in International Context*." Stanford Center for Opportunity Policy in Education, Stanford University, CA.

Fuhrman, S., and R. Elmore, eds. 1994. *The Governance of Curriculum*, Alexandria, VA: ASCD.

Greaney, V., and T. Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.

Hanushek, E., and L. Woessmann. 2007. *Education Quality and Economic Growth.* Washington, DC: World Bank.

Heubert, J., and R. Hauser. 1999. *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academy Press.

Hoxby, C. 2002. "*The Cost of Accountability*." NBER Working Paper Series, Vol. w8855. Available at SSRN: http://ssrn.com/abstract=305599.

Wolff, L. 2007. "*The Costs of Student Assessment in Latin America*." PREAL Working Paper 38. Partnership for Educational Revitalization in the Americas (PREAL), Washington, DC.

Education Notes is a series produced by the World Bank to share lessons learned from innovative approaches to improving education practice and policy around the globe. Background work for this piece was done in partnership, with support from the Russia Education Aid for Development (READ) program. For additional information or hard copies, please go to www.worldbank.org/education or contact the Education Advisory Service: eservice@worldbank.org.

Author: Marguerite Clarke **Photo Credit:** Liang Qiang/World Bank