

Measuring Education Outcomes: Moving from Enrollment to Learning

Wednesday, June 2, 2010, 1:00 pm – 5:00 pm

[Center for Universal Education](#) at Brookings, 1775 Massachusetts Ave NW, Washington, DC

On Wednesday, June 2, the Center for Universal Education at Brookings hosted a discussion on the need to refocus the international education dialogue from school enrollment to learning achieved in developing countries. Participants, who included education experts from academia, international organizations and government, assessed the current state of systematic efforts at the global level to measure learning outcomes.

Center for Universal Education Co-Director and Senior Fellow Jacques van der Gaag opened the event by charting the landscape of learning, including education outside the primary school classroom, during early childhood development and the importance of acquiring both cognitive and non-cognitive skills for ensuring learning outcomes.

Peter Savelyev of the University of Chicago presented work he completed with James J. Heckman that uses data from the Perry Preschool program. Conducted more than four decades ago, the Perry Preschool program provided high-quality early care and education on low-income three- and four-year-old children and has followed them throughout their lives. This provides insight into the factors that contribute to children's successes both in school and outside of it. Mr. Savelyev reported that non-cognitive skills are malleable and affect life outcomes and that participation in the early childhood program shifted participants' behaviors positively. A range of traits produce human achievement, non-cognitive capabilities are enhanced by the intervention of early childhood education, and policy discussions should consider how non-cognitive skills are acquired.

Ludger Wößmann of the University of Munich addressed the relationships between economic outcomes and learning, based on his work with Eric Hanushek of Stanford University. Dr. Wößmann asserted that what really counts for long-run development is not school enrollment, but the learning achieved while enrolled. In evaluating years of education and economic growth, the positive relationship between the two factors is eliminated when one controls for the quality of that education.

This means that an additional year of schooling that doesn't actually result in learning has no effect on economic growth. Examining long-run growth rates of countries with available student achievement scores shows that countries with higher test scores have higher growth rates. The powerful effects of cognitive skills on individual earnings, on the distribution of income across society, and on economic growth support a causal interpretation of the results. Simply getting children into a classroom does not equate with learning.

When indicators beyond school enrollment and completion are considered, the current situation in developing countries is much worse than generally pictured. To develop a better understanding of learning levels, an increasing number of countries are considering and implementing large-scale assessments of reading, writing, mathematics, and science. However, there are shortcomings with scaling up the internationally-comparable learning assessments used in these studies. For example, the OECD Programme for International Student Assessment (PISA), which is geared toward surveying 15-year-old students in industrialized countries, is simply too advanced for the average student in a developing country. Given that the vast majority of the students will not do well on these tests means that the individual country results will be "noisy" and have little value for countries trying to compare themselves to their peer nations. Moreover, it sheds little light on what the country should focus on to improve learning.

Finally, to translate these findings into actionable policy, Dr. Wößmann considered the range of demand-side programs used in education – including conditional cash transfers, school fee reductions, and food and nutrition supplements. Each of these typically have a positive and significant impact on attendance and persistence but, with the exception of a Kenyan merit scholarship, there is little or no apparent impact on achievement and therefore cannot, in their current format, stand as the means to increase educational outcomes in developing countries.

Discussions following the opening presentations included questioning the usefulness of drawing conclusions from the small-scaled Perry Preschool program from nearly half a century ago; the validity of using international assessments in developing countries; and discussing the possibility of setting a universal standard for learning outcomes.

Nicholas Burnett of the Results for Development Institute began the discussion by providing a broad overview of the learning assessment landscape, including international and regional assessments, national exams and local diagnostic tests. In summarizing the primary options at these different levels, Dr. Burnett identified several of the major issues. For the international assessments (PISA, TIMSS, and PIRLS), there are very few options for assessing learning in the early and middle primary grades in the international and regional assessments. In addition, the cost of administering these tests is prohibitive for the poorest countries and the data generated by the tests are relatively inaccessible to those in a position to make real changes at the country-level. While nearly all countries engage in some type of national assessment in order to ascertain whether educational goals are achieved, the utility of the data depends highly upon the quality and relevance of the assessment and the technical expertise of those involved.

At the local level, there is an important new trend of citizen groups evaluating and publicly disseminating student learning, such as ASER (India and Pakistan) and Uwezo (Kenya), which link directly with increasing accountability mechanisms at the country-level. Although researchers and macro-level policymakers search for international comparisons, the most important aspect for improving learning with a country is to be able to compare internally and over time. In terms of clear gaps, there is scarcity of work being done around assessment of vocational skills and international level assessments in the early grades.

Marguerite Clarke of the World Bank presented on their current work on educational assessments. While the World Bank has included a learning assessment component in just under half of its education projects since 1998, this support has typically supported the implementation of a national assessment of basic reading and math. However, these initiatives have had limited sustainability, retention of trained staff, or meaningful use of results. The Russia Education Aid for Development (READ) Trust Fund is a partnership between the Russian government and the World Bank to help low-income countries improve learning outcomes through improved assessment systems. Since countries use a variety of assessments in different ways and frequency, READ is focused on helping countries develop systematic assessment cycles that include measurement, analysis and utilization of data to influence relevant policies and practices. READ's "roadmap" for this systematic approach considers the different assessment types and their function (classroom assessments, high-stakes examinations and large scale assessments) and the three main drivers of quality (the enabling environment, system alignment and technical quality) to help countries self-diagnose their assessment systems, and develop indicators, performance levels, strategies, and appropriate stages of development to determine what is needed to build a more effective system.

Amber Gove of RTI International presented the work of a multi-stakeholder community of practice engaged in the Early Grade Reading Assessment (EGRA) or its equivalent that has been conducted in 43 developing countries and 58 languages thus far. EGRA is typically adapted for use in a particular country and language. It is a brief, individually-administered oral assessment primarily used to document student reading skills to inform policymakers of the education system's needs for improving teaching and learning. The assessment can also be adapted for use as diagnostic, remediation and monitoring tools for individual student progress. Dr. Gove shared results that demonstrated basic reading skills in many of these countries are severely lacking. For example, many students at the end of grade 2 cannot read a single word of connected text. Even children taught in their "mother tongue" are performing poorly, with eight out of every ten Malian second graders not able to read and more than half of the children in Uganda receiving zero scores.

Summarizing the results publicly available thus far, it is clear that many teachers lack the support and basic content knowledge to teach reading, students have few interactions with print materials outside of the classroom, and many countries have inconsistent and unused policies relating to language of instruction. These results, if utilized, can have powerful effects. For example, in The Gambia, the EGRA results influenced national policy to switch to teaching in mother tongue after it became clear that teachers tasked with conducting their classes in English did not know the language themselves. The impact of EGRA and EGRA-like assessments can be leveraged by effective communication and dissemination of results. This can establish reasonable and relevant goals and benchmarks within each country and move beyond basic literacy to measure higher level skills (an Early Grades Mathematics Assessment is in a pilot phase). The purpose of the EGRA work is demonstrate that early reading is an essential entry point for ensuring quality education at all levels. At the global level, the EFA-Fast Track Initiative is adopted new indicators on the quality of education that have countries calculating the percentage of children, who after two years of formal schooling, demonstrate sufficient reading fluency and comprehension to be able to “read to learn.”

Jordan Naidoo reported that UNICEF is currently shifting its focus from access to quality and ensuring that children complete the education cycle, meet the appropriate learning targets, and acquire relevant skills. It is important to note that given its education work in over 70 countries, UNICEF occupies a unique position to influence system-wide education policies that should be considered and leveraged in movement toward a greater global focus on learning. UNICEF’s child-friendly schools focus on a rights-based, child-centered school environment that utilizes learning as one indicator of success and supports a comprehensive definition of the learning process beyond pure cognitive outcomes to include social-emotional development. However, UNICEF recognizes that it is not in the position to take a global lead on developing a systematic approach to learning assessments. UNICEF seeks partnerships with organizations that have expertise in learning outcomes measurement and support a collaborative approach, including greater attention to the foundational phase from early child development through the early primary grades.

Building upon the presentations, the participants then discussed a number of additional issues, including:

- **Challenges with EGRA:** The EGRA instrument is not designed to be used for international comparisons since different languages have different fluency characteristics (e.g. each language’s typical word length impacts how many words per minute should be read to achieve reading fluency). At this point, the sole purpose of the assessment is to inform the country itself. For example, in Liberia, EGRA assessments illustrated where the students needed to improve and informed the development of comprehensive lesson plans for teachers that included biweekly diagnostic assessments to continually inform teacher practice. Yet even the relatively simple 15 minute assessment has proven to be too technical in some countries; there is a clear need to develop more simple assessments that people in the country can conduct and analyze without external assistance.
- **Evaluating Young Students:** In response to the observation that the bulk of the international assessments do not assess children in the early primary grades, it was noted that children in grades 1 and 2 are notoriously unreliable testers, which, taken to a large-scale, might produce inaccurate pictures of the national education system. The one-on-one method utilized by EGRA is able to mitigate some of this unreliability by interacting directly with students that is sacrificed for larger-scale sample surveys of the student population but cannot be scaled up to be nationally-representative for the same reason.
- **Shortcomings of Large-Scale Assessments:** A number of shortcomings in large-scale international assessments were identified for developing country participation. First, some countries will receive data that reveals the floor effect of participating in assessments with industrialized nations, showing unilateral poor performance without any variance that would provide the detailed data needed to make actionable adjustments to the system. Second, too often the data generated doesn’t make its way to the district and school levels where changes in teaching and learning practices can be made. Finally, too often countries do not know what to do with the information provided and participate in the tests to fulfill an international donor expectation. Again, it is clear

that for these countries with the lowest levels of capacity to administered a systematic approach to evaluation, classroom-based assessments is where the best and most useful information will be generated.

- **Reading as Foundational:** A number of participants asserted that acquiring basic literacy is the threshold to all the other academic subjects. When a child is not doing well in reading, then he or she is not doing well in math and science either. Parents expect that if their children spend several years in school they should be able to read, write and calculate.
- **Intellectual Leapfrogging:** Many countries in Latin America have been working on implementing effective assessments over the past three decades. Experts working in the area maintained that due to the technical quality of testing and the effective utilization of testing results, the learning process for countries is long. Others offered optimism that intellectual leapfrogging from the experiences in Brazil, Chile and others could help countries make quicker strides toward implementing effective assessment systems, citing the role that civil society assessments have played in pushing national ministries of education to pay attention to the results. Ultimately, testing is about accountability; as soon as governments, local authorities, teachers, students, parents are held accountable, they change the way they behave. When OECD introduced PISA, a number of industrialized countries were suddenly held accountable for their education systems in broader international perspective. In the case of developing countries, there is a sense of a need to break the cycle of low expectations, whether they come from students, parents, schools, communities, ministries or donors.
- **Bringing Teachers in as Partners:** As the key link between government policy, national curriculum and student learning, teachers can be simultaneously praised and demonized. Ensuring that teachers are adequately prepared with learning assessment modules in their teacher education courses is one essential component to building effective assessment systems. One participant asserted that when the debate around quality education focuses on improving performance in the classroom, and is not a result of external demands, then we will see better results. Bringing teachers in as collaborators in the improvement process rather than targeted as a broken link in the system could have a tremendous impact on the way teaching and learning are conducted in the classroom. The teachers' union has the ability to play an instrumental role in bringing these teachers to the table in a meaningful discussion toward an actionable plan.
- **Trajectory of Learning Assessment Systems:** Thus, while citizen-run community-based or one-on-one classroom assessments may represent the best approach in the short term, ultimately countries need to move toward a systematic way of creating an assessment structure so that it becomes a regular part of doing business and a country can eventually begin to understand how it fits into the global education landscape.
- **Lacking a Sense of Urgency:** If it is clear that simply going through the motions in education will not lead to the results we seek, then it is surprising that there is no sense of urgency in rectifying these low levels of learning at a global level. The major international players do not have consistent frameworks or even general approaches to ensuring that learning is a central component of education reform. The private sector needs to be activated in support of increased learning in the classroom and could play an instrumental role in turning secondary school tutoring that has become a necessity in a many countries to pass exit examinations into early grade support for improved teaching and learning in the classroom.