NCME national council on measurement in education

Volume 18, Number 1 NEWSLETTER March 2010

Pages prior to and following main article were removed.

LEARNING TO ASSESS STUDENT LEARNING IN AMÉRICA LATINA

Michael C. Rodriguez, University of Minnesota

The United States Agency for International Development (USAID) granted a 4-year award in 2005 to the *Programa Estándares e Investigación Educativa Guatemala* (Education Standards and Research Program), as part of its Central America and Mexico Regional Strategy to improve educational outcomes. The USAID-Guatemala personnel work closely with the Ministry of Education, local universities, and related government and community organizations, providing financial and technical assistance to improve education outcomes and inform national education dialogues. The project produced national K-12 curriculum standards and standards-based assessments in language arts and mathematics. Throughout the 5 years of the project, USAID-Guatemala and the Ministry of Education encountered significant challenges, some of which could be addressed by current measurement theory and practice, some of which could not.

The project involved consultants to provide needed expertise. Since 2007, I have worked with the USAID-Guatemala team by providing psychometric technical assistance and training. A symposium discussing the challenges and lessons learned in this project and related efforts by others in Honduras and Chile will be presented at the 2010 NCME meeting.

Assessment has a long history. Although selection and placement assessments can be traced back to 2200 BC in China, national assessments of student learning are much more recent. In the USA, the technology of testing (scoring, scaling, equating) has been advanced through national assessment efforts, in part fueled by the NAEP assessments during the 1960s. The 1983 publication of *A Nation at Risk* was a call to arms in assessment. Successive reauthorizations of the Elementary and Secondary Education Act have required greater attention to learning and assessment of student performance, particularly the 2001 reauthorization (NCLB).

Internationally, there is a wide range of experiences in the assessment of learning. In much of Europe, Asia, Australia, and a few other regions, there is a long history of assessment. However, many nations around the world, particularly throughout the Caribbean, Central and South America, and Africa, have limited experience with large-scale assessment. In many nations, education reform efforts attend to more basic needs, including the structural safety of schools, effective resource allocation, professional development of teachers, establishment of curriculum standards, and equitable access to schools.

In a recent analysis of USAID assistance to basic education in the developing world, Chapman and Quijada (2009) reviewed 33 projects that were awarded over $2.3 billion. Some have argued that international development assistance to education efforts is misguided, that addressing more urgent needs like health and food security produces a more certain return on investment. Chapman and Quijada found 15 projects involved test development. The area of test development is where professional expertise is limited, particularly in Central America. Although there are high quality postsecondary institutions throughout the region, there are no doctoral programs in psychometrics, providing no opportunity for local expertise development in this area. International development assistance provides one opportunity to secure such expertise, through technical assistance, technical consultation, and professional development training.

A limited history of national assessment has presented an immense challenge – the language of large-scale assessment is absent in education agencies, schools, communities, and families in Central America. A great amount of institutional capacity building has been required. Fortunately, many countries in the region are experiencing new goal setting and decision making capacity within their respective Ministries of Education. In Guatemala, this has required the creation of additional bureaucracy and regulation making procedures. The technical challenges have been particularly acute, especially given the limited access to professional training opportunities. Overall, local capacity building has been evident and assessment technological advancements are seen as programs move toward modern measurement theory.

The USAID-Guatemala project has developed the current testing system through an IRT measurement model and has provided training for staff within the Ministry on modern measurement theory and advanced statistics. They continue to faces challenges, including the multi-language assessments in early primary levels, extreme limited opportunities to learn, and limited teacher preparation and professional development. Guatemala has a population of nearly 14 million where more than 50% speak one of 24 Mayan languages as their first language. The nation has worked very hard to provide bilingual (Spanish & Mayan) education for the first three years of school but few native Mayan speakers obtain teaching certificates, resulting in stark inequities in education quality. Most of the Mayan population live in rural areas where the schools have very few educational resources compared to those in the urban areas, creating additional inequities. Also, the educational system has serious efficiency problems. About 34% of the students fail their first year; 42% finish primary school; less than 10% finish secondary school. Moreover, teacher preparation occurs at the secondary level, not postsecondary.

The recently developed national curriculum standards were designed to serve several purposes, including establishing clear curriculum and performance goals for each grade and equalizing the quality of education. Before the year 2006, Guatemalan

achievement tests where norm referenced and sample based, designed through university-based evaluation projects. Since 2009, the national assessments are administered annually near the end of the school year in grades 1, 3, 6, 9 and 12, addressing Mathematics and Language Arts standards. Operational forms have been developed through a common-item linking design to facilitate equating across years. The tests, linking, equating, and standard setting, have all been supported through the Rasch measurement model. A technical manual was developed during the assessment development process and was used as a guide to evaluate the degree to which each step was consistent with the *Standards for Educational and Psychological Testing*.

USAID staff have addressed several issues that became more complicated because of extreme limited opportunity to learn. For example, assessment design has been difficult with respect to assessing context effects on common items that change position across forms and years. In the context of standards-based tests where there are significant opportunity-to-learn concerns and where a large number of students do not complete the test or tend to skip items, changing the location of items beyond one page may create complications in equating and maintaining item performance consistency and score stability over time.

The choice of IRT measurement model was also complicated. The Ministry of Education considered pattern scoring and the use of a 2PL or 3PL model. For individual examinees, pattern scoring may have a positive effect, while for others, it could have a negative effect. For example, in a few areas, Mayan numbers are covered well and students learn and practice the use of this number system (a first grade standard). In many regions and schools, students do not spend much time learning or practicing the Mayan number system. Item parameters will reflect this by resulting in high levels of difficulty with low levels of discrimination. These items will not discriminate between high and low ability students because performance is not a function of ability but is a function of opportunity to learn. Now, ability scores (person-theta values) will not be a function of the items covering Mayan numbers because the low discrimination will be used to weight these items less than the other items. Students who get all of the Mayan-number questions correct will not get as much credit toward their ability score as students who may get other questions correct and all of the Mayan-number items incorrect. Equating assumes that the same content is being covered from year to year, and although it is not part of the assumptions for estimation, differential opportunity to learn has implications for the validity of equating.

Because standard setting, as a conceptual framework and a process, was new to Guatemala, the first implementation became an opportunity to investigate the methodology. The Bookmark method was selected as the most appropriate method, well suited to setting standards on multiple-choice tests and one that could implemented with sufficient fidelity. It was important to provide validity-related evidence regarding the appropriateness and feasibility of employing a standard setting method in Guatemala. As a learning opportunity, a small study was designed to investigate variability in 3rd grade results from replications of the process, including three independent panels for Language Arts and three for Mathematics. There were significant differences in some cases, including cut score differences as large as 0.44 logits on the Rasch scale. To some degree, participant experiences and perceptions of the process helped explain variation within and between panel results (Rodriguez, Rego, & Rubio, 2009).
The USAID-Guatemala project is an important example of how the educational measurement community has a great deal to contribute to and learn from education reform and assessment efforts in developing countries, particularly Central America. For a general overview of the USAID-Guatemala projects, visit: http://www.usaid.gov/gt/health_education.htm.

References

Chapman, D.W., & Quijada, J.J. (2009). An analysis of USAID assistance to basic education in the developing world. *International Journal of Educational Development, 29*, 268-280.

Rodriguez, M.C., Rego, O., & Rubio, F. (2009, April). *Examining variation in independent replications of the Bookmark standard setting method on two tests*. Paper presented at the annual meeting of the National Council on Educational Measurement, San Diego, CA.

Thanos Patelis
The College Board
45 Columbus Avenue
New York, NY 10023

Phone: 212.649.8435
Fax: 212.649.8427
e-mail: tpatelis@collegeboard.org