

Prepubl draft: Wagner, D. A. (2012). What should be learned from learning assessments?
COMPARE: A Journal of Comparative and International Education, 40, 3, 510-512.

In special Compare Forum on: The debate on learning assessments in developing countries

What should be learned from learning assessments?

Daniel A. Wagner
University of Pennsylvania

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, this Forum; Wagner et al., 2011).

Similarly, knowing *when* an assessment should take place is non-trivial. As pointed out by Gove (see this Forum), 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 & Martin, this Forum).

The *why* question is less obvious, since there are often multiple reasons why learning assessments are put into place. For large-scale assessments, Lockheed (this Forum) 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 (this Forum), by contrast, argues that NGOs have a clear and present need to continuously track the implementation of its programs, 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*. NY: Firefly Books.

Wagner, D. A. (2010). Quality of education, comparability, and assessment choice in developing countries. *COMPARE: A Journal of Comparative and International Education*, 40, 6, 741-760.

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

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