

[Education Week's blogs](#) > [Assessing the Assessments](#)

Merits of International Assessments

By [Invited Contributor Listed Below](#) on April 14, 2014 12:08 AM | [No comments](#)

Today's guest contributor is **Henry Braun**, *Boisi Professor of Education and Public Policy Educational Research, Measurement, and Evaluation Department; Director, Center for the Study of Testing, Evaluation and Education Policy (CSTEPP), Boston College.*

In discussions of the future, the term *globalization* is ubiquitous, typically referring to the breakdown of national barriers to the movement of goods, services and people. Paralleling the emergence of a one-world economy, international large-scale survey assessments (ILSAs) have also risen to prominence, with acronyms such as TIMSS, PIRLS, PISA and SAS (PIAAC) now broadly recognized. ILSAs are garnering heightened media attention and, in many countries, exerting increasing influence on education policy. Not surprisingly, this trend has occasioned considerable criticism. Some is methodological, principally questioning the comparability of results given differential sample quality and the need to adapt the assessment instrument to dozens of different cultures and languages. Other criticisms are directed at the excessive attention paid to country rankings and the tendency to over-interpret the results in the search for productive policy strategies. One particular concern is that setting goals in terms of improving its ranking on one or more ILSAs could lead to a homogenization of education that does a disservice to a nation's distinctive culture and educational needs.

These critiques have merit. Indeed, ILSA sponsors and the organizations that actually conduct the assessments have worked to address methodological deficiencies, though much remains to be done. On the policy side, criticism of the pernicious impact of (naïve) country comparisons is certainly in order. At the same time, we should not lose sight of the many positive contributions that ILSAs make to education policy. The goal should be to strengthen their capacity to do good while working assiduously to minimize negative consequences, unintended or not. Let's look at some of these contributions.

Before the advent of ILSAs, each country's educational system was hermetically sealed -there was no way to make meaningful comparisons among them. (At the state level, this was the situation in the U.S. before NAEP's Trial State Assessment began.) One problem was that the claims made by those in charge of the system, often exaggerated and self-serving, could not be easily refuted. Today, the burden of proof lies on those making claims that run counter to the evidence provided by ILSAs, particularly if the divergence is substantial. Although country "league tables" play an out-size role in the minds of many policy makers, more useful comparisons are possible. For example, comparisons across jurisdictions of the variances in test scores, of the gradients of test scores on SES, or of gaps between immigrants and native born students can be informative and even a spur to action. Sophisticated statistical analyses are not needed to extract useful information from ILSAs, as is attested by the information-rich almanacs produced in conjunction with the release of the basic data.

A striking example is provided by the results of the latest Survey of Adult Skills (SAS), conducted under the auspices of the OECD's Programme for the International Assessment of Adult Competencies. SAS is a household survey that assesses adults ages 16 to 65 in literacy, numeracy and (this administration) in problem-solving in technology-rich environments. Thus, it is possible to compare skills both across age groups within a country and within age groups across countries. What do we find? With regard to the oldest cohort, adults aged 55-65, the U.S. is a leader in literacy among OECD countries. Further, in the U.S., as in other OECD countries, the youngest cohort, ages 16-24, has stronger literacy skills than the oldest cohort. However, in comparison to their age peers across the OECD, young adults in the U.S. are, at best, in the middle of the pack. The concern is not simply the drop in the rankings; rather, it is that the score gap between the U.S. and the leaders is substantively meaningful and serves as a leading indicator of a possible long-term decline in global competitiveness. In the absence of the SAS, it would be nearly impossible to draw such a strong - and policy relevant - conclusion.

As cross-sectional studies, ILSAs are limited in offering evidence to support directly the kinds of causal conclusions desired by policy makers and score differences of a few points at the national level are not particularly meaningful. But used wisely, the rich data generated by ILSAs, in conjunction with other relevant evidence, provide unique insights that can challenge unmerited complacency and establish worthy benchmarks for educators and policymakers to aim for.

Henry Braun

Center for the Study of Testing, Evaluation and Education Policy (CSTEPP)

Center for the Study of Testing, Evaluation and Educational Policy (CSTEEP)

Boston College

Categories: [Validity](#) [High-stakes testing](#) [School accountability](#) [International assessments](#) [Test use](#)

Tags: [ILSA](#) [International Large-Scale Assessments](#) [NAEP](#) [OECD](#) [PIAAC](#) [PIRLS](#) [PISA](#) [SAS](#) [TIMSS](#)