

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

Are Fourth Graders Who Don't Test Like Seventh Graders Really Failures?

By James Harvey on April 30, 2014 8:48 AM | [No comments](#)

This week **Madhabi Chatterji**, associate professor and founding director of the Assessment and Evaluation Research Initiative at Teachers College, Columbia University and **James Harvey**, executive director of the National Superintendents Roundtable wrap up this month-long conversation between measurement experts and educators on the front line. **Meiko Lin**, who managed the blog on behalf of Chatterji and Harvey, helped guide the discussion by formulating some questions about unresolved issues. Read their final thoughts and "takeaways" from *Assessing the Assessments*.

Meiko Lin: What emerged from this exercise? What gaps were left in the discussion?

James Harvey: This was an extremely useful undertaking. It revealed the eagerness of both measurement experts and front-line educators to engage around assessment issues. There seemed to be considerable concern on both sides about overuse and potential misuse of testing. Madhabi Chatterji's initial [commentary](#) and [blog entry](#) directly addressed these issues. I believe we got an excellent sense of what impact the assessment for accountability movement of the last decade has had at the district, school, and classroom level. From a practitioner's perspective, I learned that there's a strong sense of shared interest from academics like [Deanna Sands](#) and superintendents like [Kelley Kalinich](#) and [Steven Ladd](#) in strengthening formative assessment. There also seems to be some concern on the part of academics like [James Pellegrino](#) and [Edmund Gordon](#) about the importance of learning from the assessment mistakes of the past.

There also seems to be disagreement on some points that could be empirically settled. On one hand, we have the assertion from [William Schmidt](#) that education plays a crucial role in economic competitiveness and the (undoubtedly true) argument that U.S. performance cannot be attributed solely to poverty. On the other, we have the arguments, from [Iris Rotberg](#), that education, per se, plays at best a limited role in competitiveness and, from [Paul Ash](#), that 40 percent or more of outcomes by nation, on average, are related to socio-economic factors.

So, one gap is that we were never really able to iron out (or even highlight) these differences. It's not clear the blog had the capacity to change closely held beliefs on either side. Another is that I don't think we did the issue of Value-Added Measurement justice--either arguments for it or against it.

Finally, we didn't really address the implications of the fact that NCES continues to insist in each of its major reports that the NAEP benchmarks "should continue to be used on a trial basis and should be interpreted with caution." Yet each of the Common Core assessment consortia have with [almost no public discussion](#) defined the NAEP proficiency benchmark as the acceptable level of performance for every student in the United States. That decision by itself is enough to explain why the passing rates of New York students dropped like a stone when the Common Core assessments were implemented on a trial basis. Why should schools be deemed a failure because only a third of fourth-graders are comfortable with [items testing reading at the fifth- to seventh-grade levels?](#)

Meiko Lin: How do we guarantee that all stakeholders have a role in assessing teaching and learning? What should we expect from each of them?

James Harvey: This is a powerful and difficult question. It governs how measurement professionals and testing companies on one hand and policymakers, educational leaders, and principals and teachers, on the other, understand what each side is doing. One must not make the mistake of assuming that educators are not interested in assessing learning. They have always assessed it. The issues that need discussion from the educational side include: Should computer-based formative national assessments replace teacher-developed assessments? What has a reductive emphasis on reading and language arts done to the curriculum? What has it done to the school calendar? Does accountability at the school and district level really require that every child, in every school be tested every year, in every subject? While each assessment may be justified for different purposes in its own terms, the accumulation of assessments overwhelms schools, as [Richard Noonan](#) pointed out.

From the measurement side, I believe the testing companies (whether profit-making or non-profit) need to do a much better job of explaining how they construct and administer these assessments. While it may be difficult to explain the complexities of large-scale assessments to educators, the reality is that teachers, principals, and district administrators are entitled to a much clearer sense of how much judgment goes

into selecting items, samples, and the matrices which determine which students get which testing booklet. It is hard to find much evidence of practitioner involvement in the rapid development of the Common Core or its assessments. I'm overstating it when I say that I've often thought that the modified Angoff procedures used to establish benchmarks on some of these large-scale tests are little better than throwing darts at a blackboard, but I do know that many measurement experts themselves have serious questions about this process.

Meiko Lin: What are some of the ways you would recommend that the general public interpret international large-scale assessment (ILSA) results? What can we infer and what can't we infer from ILSA results?

James Harvey: I think I'd ask the general public and policymakers to remember the words of Daniel Kahneman, Princeton University Emeritus Professor of Psychology and winner of the 2002 Nobel Prize in Economics. In *Thinking, Fast and Slow*, Kahneman observed, "The errors of a theory are rarely to be found in what it asserts explicitly; they hide in what it ignores or tacitly assumes."

He explained it by a weakness he observed in himself: "I call it theory-induced blindness: once you have accepted a theory and used it as a tool in your thinking, it is extraordinarily difficult to notice its flaws." In fact the theory about how the world works can easily mislead us as to what is actually going on.

I truly thought Oren Pizmony-Levy's contribution to the blog broke new ground. He reminded us that ILSA's were developed not to rank countries against each other but to provide educators and policymakers within each country a better sense of how successful they were in pursuit of their own education goals. What the ILSA's (and to some extent NAEP) now tacitly assume is that they are assessing school outcomes. I believe they are ignoring what is actually going on and, in many ways, misleading policymakers as to what needs to be done. Granted, policymakers, in demanding report cards that rank countries, are complicit in this. I think measurement experts should be in the vanguard of those questioning these demands. What ILSA rankings are ignoring is the social context in which schools around the world function. They make the mistake of believing that because they assess learning in schools that's where it all takes place. What they are measuring is not school success alone (although that's part of it), but each society's commitment to the well-being of its next generation.

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

Tags: [accountability](#) [assessment](#) [high-stakes testing](#) [international large scale assessment](#) [test use](#) [validity](#)