Education Week

Commentary

We Must Teach for 'Range' and 'Depth'

By James Nehring

There is a problem baked in to federal and state accountability policies when it comes to assessment of learning. The problem is so fundamental that any effort to develop assessment systems by way of the same accountability recipe will produce the same flat-as-a-pancake result.

The problem is this: Human judgment is poison to accountability, but it is the basic ingredient for assessment of learning.

Public accountability systems exist largely to ensure a meritocracy uncontaminated by nepotism, political favoritism, or anything that would threaten *merit* as the sole criterion for consequential decisions. This is the logic behind civil service exams, going back to ancient China and extending to modern public bureaucracies around the world. The system seeks to ensure that a person gets a job based on his or her qualifications, not who they know or how much money they have or which political party they favor. The idea is to so marginalize—or, better, eliminate—human judgment in the evaluation process so that no one can charge bias. Objective *measures*. A precise and scientific enterprise.

The testing regime at the heart of education accountability embraces this premise. Human judgment is a contaminant, so it's best to bleach it out and replace it with clean, pure, multiple-choice, and, where unavoidable, "prose constructed response items" bolted down firmly with headache-inducing rubrics. Let the data drive decisions.

But judgment is essential to learning assessment. Sure, we can *measure* discrete skills, but the ability to tackle a complex project where identifying the right questions is part of the process, where work unfolds over days and weeks and revision is essential, does not lend itself to mere measurement.

However, such work can be *assessed*, a more nuanced but less precise enterprise involving evidence, intelligence, conversation, and judgment. As long as policy, in an era of accountability, relies chiefly on *measurement*, our schools have little incentive or pressure to teach for range or depth. Instead, we teach decontextualized, discrete skills, unsuited to most tasks offered up by the real worlds of work, citizenship, or personal life. If we switch to an *assessment* paradigm, however, everything changes.

In recent years, the public has recognized this problem, and the system has responded in the only way the system can: with a more elaborate test, more complicated *measurement*. The PARCC and Smarter Balance tests tied to the Common Core State Standards claim to incorporate "performance assessment," co-opting the language of their detractors but delivering only a pathetic conceit: Fill-in-the-blank

replaces multiple-choice, and open-response items pile on ever-more-elaborate rubrics. Associated time and costs weigh the whole system down. Testing has thus entered its own rococo phase. It has become an elaborate parody of itself.

What if, instead of marginalizing human judgment in the assessment of learning, we honor it. What if we admit that a test, no matter how valid, reliable, and aligned, is simply not up to the task because all it can do is *measure*, and what we need requires something more. What if we build a system around human judgment that minimizes its vagaries while bolstering its strengths? It's how our legal system works.

Consider: In complex matters involving human motive, incomplete information, context, and ethics, our best recourse is the collective judgment of informed adults—an impaneled jury. The jurors are provided with all the evidence and the best arguments on all sides, but the decision, ultimately, lies with them. Data is helpful, but it does not "drive." Measures inform, but people, in the end, make an *assessment*. In looking for a model upon which to base assessment of learning, we'll do far better studying our democratic traditions than our civil service.

Some educational assessment systems, in both low- and high-income jurisdictions, work from this more appropriate starting point. Teams of teachers assess portfolios of student work using sensible rubrics keyed to public standards. Students present in juried exhibitions. Peer review ensures trustworthiness through schoolwide audits.

Such a system denies the policy world what it craves, simple numbers on a line, but there are still plenty of things about schools that we can usefully subject to measurement: resource allocation, student attendance, parent satisfaction, school climate, graduation rates, and more. When it comes to learning, however, in all its complexity, it is wise to remember the maxim *not everything that counts can be counted*.

Yes, assessment is flawed. Human judgment is imperfect. Juries make bad decisions. But I'll sooner entrust a group of informed, vetted, and thoughtful school teachers with my child's educational future than Pearson or PISA or the Educational Testing Service. If we believe that people, acting together with good information in good faith can't do the work, then we are lost as a democracy. Thomas Jefferson got it exactly right over 200 years ago when he wrote: "I know no safe depository of the ultimate powers of society but the people themselves, and if we think them not enlightened enough to exercise their control with a wholesome discretion, the remedy is not to take it from them, but to inform their discretion."

It's time to stop measuring what can't be measured, acknowledge the stunning complexity of learning, and build a system based on human judgment and authentic *assessment*.

JAMES NEHRING is an associate professor of education at the University of Massachusetts Lowell. He served for 25 years as a high school teacher and school leader. His most recent book, *Why Teach? Notes and Questions From a Life in Education* (Rowman & Littlefield), is due out in September. His website, which features his blog, is www.jamesnehring.org.