

“This Is Not About Accountability”

A three-year study released by the Bill & Melinda Gates Foundation might shed light on the proper role of testing in public education reform; or the study’s most important findings might be ignored—overwhelmed by the national educational-testing juggernaut.

The Gates Foundation has long used its substantial clout and resources to promote value-added measures, or a practice that relies on students’ standardized test scores to evaluate teacher performance. Across the nation many policymakers and reform advocates have been quick to seize upon value-added as the key tool to improve public schools. Advocates of value-added claim that the measures can produce easily understood and scientific teacher rankings to identify most-effective and least-effective teachers.

Educators, backed by strong research evidence, have mostly claimed that the measures are woefully inexact, and their fears of inappropriate use have been borne out. For example, value-added has been used in some districts or states as the dominant or even single teacher-evaluation measure; it has been promoted as the basis for merit rewards; it has been used to “motivate” teachers by publicly posting their individual “scores” in newspapers.

The new report, *Measures of Effective Teacher*ⁱ (MET), confirmed some of the concerns that value-added critics anticipated years ago when Gates first became a powerful education-reform player. The philanthropic foundation continues to tout value-added measures, but now it takes a more nuanced stand, urging that a *mixture* of multiple measures, including student surveys and classroom observations, yields more predictive results (Washington Postⁱⁱ, Wall Street Journalⁱⁱⁱ, Reuters^{iv}, Huffington Post^v, School Finance 101^{vi}).

“If you select the right measures, you can provide teachers with an honest assessment of where they stand in their practice that, hopefully, will serve as the launching point for their development,” said Thomas Kane, a professor of education and economics at Harvard University, who headed the study (Education Week^{vii}). Since 2009, the nearly \$50 million study observed 3,000 teachers in six districts nationwide—Charlotte-Mecklenberg, N.C., Dallas, Denver, Hillsboro County, Fla., Memphis, Tenn., and New York City.

The key, then, becomes figuring out just how much weight to put on each of these measures and what to do with the drawn conclusions. This is a matter of some urgency as more and more states—under pressure from the federal government—are reforming their teacher evaluation systems to include standardized test scores (Hechinger Report^{viii}). The MET report concluded that test scores should count for between 33 percent and 50 percent of a teacher’s overall evaluation. It’s unclear how the researchers arrived at this sweeping estimate for how to “weigh” teacher evaluations. Perhaps, in the real world of schools and classrooms, it should be taken

to mean that student performance on state standardized tests can help guide but should not determine teacher evaluation.

And what of the remaining 50 percent to 67 percent? Will these other measures receive the attention, funding, and study comparable to that given to testing? Surveys and classroom observations are especially useful, along with student performance that is not measured by tests—including students' practical application of skills, student improvement, demonstrations, portfolios, and so forth.

Classroom observation poses challenges to ensuring evaluator competence and fairness. The observations recommended in the MET report have to be more discerning than once-a-year drop-ins by a principal. Such single observations were a poor indicator in the study, whereas averaging multiple visits by multiple observers improved reliability. Randi Weingarten, president of the American Federation of Teachers, said, "Just dropping by a teacher's classroom and writing up an evaluation must be replaced with a more serious process that actually helps improve teacher practice and student learning," Weingarten said (SchoolBook^{ix}).

Using teacher observations to improve student learning requires extensive training for observers, support services to help struggling teachers, and changed attitudes that favor teacher learning over teacher competition. Without such commitment and investment as a starting point, the default position will surely be to settle on the "silver bullet" of using student test data to determine which teachers to reward and which to punish.

There may be hints of hope in the words of Kane, the study's lead researcher, whose earlier work is often cited by advocates of using student test scores to make high stakes decisions on teachers' employment. "If we want students to learn more, teachers must become students of their own teaching" Kane said (Los Angeles Times^x). "This is not about accountability. It's about providing feedback every professional needs to strive toward excellence."

ⁱ http://metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf

ⁱⁱ http://www.washingtonpost.com/national/gates-study-weve-figured-out-what-makes-a-good-teacher/2013/01/08/05ca7d60-59b0-11e2-9fa9-5fbd9530eb9_story.html

ⁱⁱⁱ

<http://online.wsj.com/article/SB10001424127887323706704578230121388712506.html>

^{iv} <http://www.reuters.com/article/2013/01/08/us-education-teachers-idUSBRE90713020130108>

v http://www.huffingtonpost.com/2013/01/08/gates-foundation-met-report-teacher- n 2433348.html?utm_hp_ref=education

vi <http://schoolfinance101.wordpress.com/2013/01/09/gates-still-doesnt-get-it-trapped-in-a-world-of-circular-reasoning-flawed-frameworks/>

vii

http://www.edweek.org/ew/articles/2013/01/08/17teach_ep.h32.html?utm_source=fb&utm_medium=rss&utm_campaign=mrss

viii <http://hechingered.org/content/report-more-states-using-student-data-in-reform 5997/>

ix <http://www.schoolbook.org/2013/01/08/gates-study-finds-test-scores-can-measure-teacher-effectiveness/>

x <http://latimesblogs.latimes.com/lanow/2013/01/student-test-scores-one-way-to-gauge-most-effective-teaching-new-study-says.html>