

May 15, 2011

To The New York State Board of Regents:

As researchers who have done extensive work in the area of testing and measurement, and the use of value-added methods of analysis, we write to express our concern about the decision pending before the Board of Regents to require the use of state test scores as 40% of the evaluation decision for teachers.

As the enclosed report from the Economic Policy Institute describes, the research literature includes many cautions about the problems of basing teacher evaluations on student test scores. These include problems of attributing student gains to specific teachers; concerns about overemphasis on “teaching to the test” at the expense of other kinds of learning; and disincentives for teachers to serve high-need students, for example, those who do not yet speak English and those who have special education needs.

Reviews of research on value-added methodologies for estimating teacher “effects” based on student test scores have concluded that these measures are too unstable and too vulnerable to many sources of error to be used as a major part of teacher evaluation. A report by the RAND Corporation concluded that:

The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools.<sup>1</sup>

The Board on Testing and Assessment of the National Research Council of the National Academy of Sciences stated,

*...VAM estimates of teacher effectiveness ... should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable.*

Henry Braun, then of the Educational Testing Service, concluded in his review of research:

VAM results should not serve as the sole or principal basis for making consequential decisions about teachers. There are many pitfalls to making causal attributions of teacher effectiveness on the basis of the kinds of data available from typical school districts. We still lack sufficient understanding of how seriously the different technical problems threaten the validity of such interpretations.<sup>2</sup>

According to these studies, the problems with using value-added testing models to determine teacher effectiveness include:

---

<sup>1</sup> Daniel F. McCaffrey, Daniel Koretz, J. R. Lockwood, Laura S. Hamilton (2005). Evaluating Value-Added Models for Teacher Accountability. Santa Monica: RAND Corporation.

<sup>2</sup> Henry Braun, Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models (Princeton, NJ: ETS, 2005), p. 17.

- Teachers' ratings are affected by differences in the students who are assigned to them. Students are not randomly assigned to teachers – and statistical models cannot fully adjust for the fact that some teachers will have a disproportionate number of students who may be exceptionally difficult to teach (students with poor attendance, who are homeless, who have severe problems at home, etc.) and whose scores on traditional tests have unacceptably low validity (e.g. those who have special education needs or who are English language learners). All of these factors can create both misestimates of teachers' effectiveness and disincentives for teachers to want to teach the neediest students, creating incentives for teachers to seek to teach those students those expected to make the most rapid gains and to avoid schools and classrooms serving struggling students.
- Value-added models of teacher effectiveness do not produce stable ratings of teachers. Teachers look very different in their measured effectiveness when different statistical methods are used.<sup>3</sup> In addition, researchers have found that teachers' effectiveness ratings differ from class to class, from year to year, and even from test to test, even when these are within the same content area.<sup>4</sup> Henry Braun notes that ratings are most unstable at the upper and lower ends of the scale, where many would like to use them to determine high or low levels of effectiveness.
- It is impossible to fully separate out the influences of students' other teachers, as well as school and home conditions, on their apparent learning. No single teacher accounts for all of a student's learning. Prior teachers have lasting effects, for good or ill, on students' later learning, and current teachers also interact to produce students' knowledge and skills. Some students receive tutoring, as well as help from well-educated parents. A teacher who works in a well-resourced school with specialist supports serving students from stable, supportive families may appear to be more effective than one whose students don't receive these supports.

These problems are exacerbated further by the fact that the kind of grade-level tests and end-of-course tests used in New York are not designed to measure student growth.

While value-added models based on student test scores are useful for looking at groups of teachers for research purposes – for example, to examine the results of professional development programs or to look at student progress at the school or district level, they are problematic as measures for making evaluation decisions for individual teachers.

We urge you to reject proposals that would place significant emphasis on this untested strategy that could have serious negative consequences for teacher and for the most vulnerable students in the State's schools.

---

<sup>3</sup> Rothstein, J. (2007). Do Value-Added Models Add Value? Tracking, Fixed Effects, and Causal Inference. National Bureau for Economic Research.

<sup>4</sup> Lockwood, J. R., McCaffrey, D. F., Hamilton, L.S., Stetcher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44 (1), 47 – 67.

Eva Baker, Distinguished Professor, UCLA Graduate School of Education  
Director, National Center for Research on Evaluation, Standards and Student Testing (CRESST)  
President, World Educational Research Association, 2010-2012  
Past President, American Educational Research Association

Linda Darling-Hammond, Charles E. Ducommun Professor of Education, Stanford University  
Past President, American Educational Research Association  
Executive Board Member, National Academy of Education

Edward Haertel, Vida Jacks Professor of Education, Stanford University  
Chair, Board on Testing and Assessment, National Research Council  
Vice-President, National Academy of Education  
Past President, National Council on Measurement in Education

Helen F. Ladd, Edgar Thompson Professor of Public Policy and Professor of Economics,  
Sanford School of Public Policy, Duke University  
President, Association of Public Policy and Management

Henry M. Levin, William Heard Kilpatrick Professor of Economics and Education, Teachers  
College, Columbia University  
Past President, Evaluation Research Society  
Past President, Comparative and International Education Society

Robert E. Linn, Professor Emeritus, University of Colorado at Boulder  
Past President, American Educational Research Association  
Past President, National Council on Measurement in Education

Aaron Pallas, Professor of Sociology and Education, Teachers College, Columbia University  
Fellow, American Educational Research Association

Richard Shavelson, Dean Emeritus and Margaret Jacks Professor Emeritus, Stanford University  
Past President, American Educational Research Association

Lorrie A. Shepard, Dean & Distinguished Professor, University of Colorado at Boulder  
Past President, American Educational Research Association  
Past President National Academy of Education  
Past President National Council on Measurement in Education

Lee S. Shulman, Charles E. Ducommun Professor Emeritus, Stanford University  
President Emeritus, Carnegie Foundation for the Advancement of Teaching  
Past President, American Educational Research Association