



In new EPI report, leading educational testing experts caution against heavy reliance on the use of test scores in teacher evaluation

Student test scores are not reliable indicators of teacher effectiveness, even with the addition of value-added modeling (VAM), a new Economic Policy Institute report by leading testing experts finds. Though VAM methods have allowed for more sophisticated comparisons of teachers than were possible in the past, they are still inaccurate, so test scores should not dominate the information used by school officials in making high-stakes decisions about the evaluation, discipline and compensation of teachers.

NEWS FROM EPI

FOR IMMEDIATE
RELEASE,
Sunday,
August 29, 2010

CONTACT
Phoebe Silag or
Karen Conner
202-775-8810
news@epi.org

The Obama administration has encouraged states to adopt laws that use student test scores as a significant component in evaluating teachers, and a number of states have done so already. The Los Angeles Times recently used value-added methods to evaluate teachers in the Los Angeles Unified School District based on the test scores of their students, and Secretary of Education Arne Duncan supported the paper's decision to publicly release this information, asserting that parents have a right to know how effective their teachers are. The conclusions of this report suggest that the Los Angeles Times' analysis, which attempts to analyze teacher effectiveness, is unreliable and inaccurate.

The distinguished authors of EPI's report, *Problems with the Use of Student Test Scores to Evaluate Teachers*, include four former presidents of the American Educational Research Association; two former presidents of the National Council on Measurement in Education; the current and two former chairs of the Board of Testing and Assessment of the National Research Council of the National Academy of Sciences; the president-elect of the Association for Public Policy Analysis and Management; the former director of the Educational Testing Service's Policy Information Center and a former associate director of the National Assessment of Educational Progress; a former assistant U.S. Secretary of Education; a former and current member of the National Assessment Governing Board; and the current vice-president, a former president, and three other members of the National Academy of Education.

The co-authors make clear that the accuracy and reliability of analyses of student test scores, even in their most sophisticated form, is highly problematic for high-stakes decisions regarding teachers. Consequently, policymakers and all stakeholders in education should rethink this new emphasis on the centrality of test scores for holding teachers accountable.

Analyses of VAM results show that they are often unstable across time, classes and tests; thus, test scores, even with the addition of VAM, are not accurate indicators of teacher

-More-

effectiveness. Student test scores, even with VAM, cannot fully account for the wide range of factors that influence student learning, particularly the backgrounds of students, school supports and the effects of summer learning loss. As a result, teachers who teach students with the greatest educational needs appear to be less effective than they are. Furthermore, VAM does not take into account nonrandom sorting of teachers to students across schools and students to teachers within schools.

There are further negative consequences of using test scores to evaluate teacher performance. Teachers who are rewarded on the basis of their students' test scores have an incentive to "teach to the test," which narrows the curriculum not just between subject areas, but also within subject areas. Furthermore, creating a system in which teachers are, in effect, competing with each other can reduce the incentive to collaborate within schools—and studies have shown that better schools are marked by teaching staffs that work together. Finally, judging teachers based on test scores that do not genuinely assess students' progress can demoralize teachers, encouraging them to leave the teaching field.

Evaluating teachers accurately is a critical piece of the effort to improve America's schools, and VAM methods are appealing in that they seem to offer an objective and simplified way of comparing one teacher with another. However, as EPI's report makes clear, "There is simply no shortcut to the identification and removal of ineffective teachers." The authors conclude that that, "Although standardized test scores of students are one piece of information that school leaders may use to make judgments about teacher effectiveness, test scores should be only a small part of an overall comprehensive evaluation."

The report's co-authors are:

- **Eva L. Baker**, Professor of education at UCLA and Co-Director of the National Center for Evaluation Standards and Student Testing (CRESST)
- **Paul E. Barton**, former Director of the Policy Information Center of the Educational Testing Service
- **Linda Darling-Hammond**, Professor of education at Stanford University, former President of the American Educational Research Association
- **Edward Haertel**, Professor of education at Stanford University, former President of the National Council on Measurement in Education, Chair of the National Research Council's Board on Testing and Assessment, former Chair of the committee on methodology of the National Assessment Governing Board
- **Helen F. Ladd**, Professor of public policy and economics at Duke University, President-elect of the Association for Public Policy Analysis and Management
- **Robert L. Linn**, professor emeritus at the University of Colorado, former President of the National Council on Measurement in Education and of the

American Educational Research Association, former Chair of the National Research Council's Board on Testing and Assessment

- **Diane Ravitch**, Research Professor at New York University and historian of American education
- **Richard Rothstein**, Research Associate of the Economic Policy Institute
- **Richard J. Shavelson**, Professor of Education (Emeritus), former dean of the School of Education at Stanford University, and former president of the American Educational Research Association
- **Lorrie A. Shepard**, Dean and professor at the School of Education at the University of Colorado at Boulder, former President of the American Educational Research Association, immediate past President of the National Academy of Education

Each author is available for comment.

###

The Economic Policy Institute (EPI) is an independent, nonprofit, nonpartisan think tank that researches the impact of economic trends and policies on working people in the United States and around the world. EPI's mission is to inform people and empower them to seek solutions that will ensure broadly shared prosperity and opportunity.

Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains

July 2010

Peter Z. Schochet
Hanley S. Chiang
Mathematica Policy Research

Abstract

This paper addresses likely error rates for measuring teacher and school performance in the upper elementary grades using value-added models applied to student test score gain data. Using realistic performance measurement system schemes based on hypothesis testing, we develop error rate formulas based on OLS and Empirical Bayes estimators. Simulation results suggest that value-added estimates are likely to be noisy using the amount of data that are typically used in practice. Type I and II error rates for comparing a teacher's performance to the average are likely to be about 25 percent with three years of data and 35 percent with one year of data. Corresponding error rates for overall false positive and negative errors are 10 and 20 percent, respectively. Lower error rates can be achieved if schools are the performance unit. The results suggest that policymakers must carefully consider likely system error rates when using value-added estimates to make high-stakes decisions regarding educators.