

## Caution Urged in Using 'Value Added' Evaluations

Scholars say districts must be more careful

By Sarah D. Sparks

Top researchers studying new “value added” or “growth index” models for measuring a teacher’s contribution to student achievement completely agree on only one thing: These methods should be used in staff-evaluation systems with more care than they have been so far.

That area of agreement emerged in an Aug. 9 meeting that drew together a who’s who of a dozen of the nation’s top education researchers on value-added methods—in areas from education to economics—to build, if not consensus, at least familiarity within a disparate research community for value-added systems. The U.S. Department of Education’s research agency, which organized the forum, last week released the proceedings of the meeting, as well as individual briefs from each of the experts.

“There’s been a huge amount of research in this field in recent years, but it tends to be really siloed,” John Q. Easton, the director of the Institute of Education Sciences, told members of the National Board for Education Sciences, IES’s advisory group, during a briefing earlier this month. “People don’t seem to read each other’s work, and it’s published in totally different journals. It was so typical to read somebody’s study who was not citing all the others.”

### Pros and Cons

Value-added methods, which attempt to measure teachers’ performance based on their students’ test scores, have gained support in the last decade, as **studies** by Stanford University economist Eric A. Hanushek and others found inconclusive evidence to support a link between a teacher’s effectiveness and his or her degree credentials—the latter of which is the traditional basis for teacher pay. Massive federal support, in the form of the \$290 million Teacher Incentive Fund and the \$4 billion Race to the Top competition has led to rapid growth in the number of states and districts adopting these methods in their teacher evaluation systems.

[← Back to Story](#)

EDUCATION WEEK  
**Multi-User**  
LICENSING




**UNLOCK  
ESSENTIAL  
CONTENT**  
Open your  
team’s potential

**+ Click here for more info**

### Summit on 'Value Added' Research: Who's Who

The federal Institute of Education Sciences recently convened a meeting of a dozen top researchers on the use of value-added

Advocates argue that value-added methods can be more objective than principal observations alone, and if done well can provide information about areas in which a teacher needs to beef up instruction. Critics contend these scores can only be used for teachers of mathematics and English/language arts in tested grades, leaving out both a large proportion of district teachers and any contribution a teacher makes to untested subjects or skills, be they science or self-control.

**One influential study**  by Jesse Rothstein, a public policy and economics professor at the University of California, Berkeley, and a participant in the meeting, found a standard value-added model was biased because it did not take into account that parents and principals often push teachers to take certain students, rather than assigning them at random.

"[Value-added measures] will deteriorate—will become less reliable and less closely tied to true effectiveness—if they are used for high-stakes individual decisions," Mr. Rothstein wrote in a brief for the meeting. "How much will teachers change their content coverage, neglect nontested subjects and topics, lobby for the right students, teach test-taking strategies, and cheat outright? ... We simply don't know."

### Tools for Improvement

The Measures of Effective Teaching Project, funded by the Seattle-based Bill and Melinda Gates Foundation, is expected to release a report later this year in which class rosters were randomly assigned to clusters of teachers by school, grade, and subject area. (Education Week receives support from the Gates Foundation for coverage of the education industry and K-12 innovation.) This may help identify how the selection bias Mr. Rothstein mentioned takes place and can be prevented, according to Thomas Kane, an education and economics professor at the Harvard Graduate School of Education, and a meeting participant.

Mr. Kane and fellow Harvard assistant education professor Andrew Ho, contended that district leaders should focus less on using value-added systems to rank teachers, which Mr. Ho likened to hospital intake questionnaires that identify initial symptoms. "Medicine (and education) is not only about symptoms (and even less so about one-dimensional rankings of symptoms), but, far more critically, diagnosis and ultimately treatment," Mr. Ho

methods to measure teacher effectiveness:

- **DAMIAN W. BETEBENNER**, senior associate, National Center for the Improvement of Educational Assessment, Dover, N.H.
- **HENRY BRAUN**, director, Center for the Study of Testing, Evaluation, and Education Policy, and professor of education and public policy, Boston College
- **SEAN P. CORCORAN**, associate professor of educational economics, Steinhardt School of Culture, Education, and Human Development, New York University
- **LINDA DARLING-HAMMOND**, professor of education and faculty co-director, Stanford Center for Opportunity Policy in Education, Stanford University
- **JOHN N. FRIEDMAN**, assistant professor of public policy, John F. Kennedy School of Government, Harvard University, and faculty research fellow, National Bureau of Economic Research, Cambridge, Mass.
- **DANIEL GOLDHABER**, director, Center for Education Data and Research, Seattle, and interdisciplinary arts and sciences professor, University of Washington Bothell
- **ANDREW HO**, assistant professor, Harvard Graduate School of Education
- **THOMAS KANE**, professor of education and economics, Harvard Graduate School of Education, and faculty director, Center for Education Policy Research, Cambridge, Mass.
- **HELEN F. LADD**, professor of economics and public policy, Duke University
- **ROBERT C. PIANTA**, dean, Curry School of Education, University of Virginia, and director of the university's Center for Advanced Study of Teaching and Learning

said. "How can we use VAM results to improve teaching and the teacher corps?"

Education officials' tendency to average multiple measures or years of data into a single composite score worried many researchers.

From one year to the next, a teacher's ratings under some of the value-added systems now in use can vary by 4 percent to 25 percent, according to Linda Darling-Hammond, an education professor and faculty co-director of the Stanford Center for Opportunity Policy in Education at Stanford University in Palo Alto, Calif. She argued that researchers and policymakers must take into account the range of scores available on their state's tests when developing a value-added system. For example, a teacher of gifted students may not show up as very effective, because his or her students are already performing near the top of the test's ability to measure their progress.

Mr. Kane countered that teachers have such a strong effect on student achievement that if value-added measures help identify teachers in the bottom 5 percent of performance and bring them up to the district average, they can lead to an average increase in lifetime earnings for each student of \$52,000 as a result of being taught by that teacher for one year.

### Common-Core Concerns

Many of the experts see both promise and peril in the rollout of the Common Core State Standards and their effect on existing and emerging teacher evaluation systems.

In most districts, researchers voiced concern that evaluation systems do not take into account the time it will take for even the most effective teachers to adapt to new areas of focus in the standards—not to mention that the common core deliberately omits guidance on specific teaching strategies to meet the new requirements.

For example, Henry Braun, the director of the Center for the Study of Testing, Evaluation, and Education Policy at Boston College and a consultant with the Partnership for Assessment of Readiness for College and Careers, or PARCC, one of the two consortia developing tests for the common core, has been struggling with how to design an assessment which likely will end up being used for teacher evaluation. He worried that if the teacher accountability "tail" wags the student assessment "dog," tests won't be designed appropriately to measure students' learning rather than teacher behavior.

Experts called for state policy leaders to consider how their individual state tests will affect the validity of individual districts' evaluation systems.

• **JONAH E. ROCKOFF**, associate professor of business, Columbia Graduate School of Business, and faculty research fellow, National Bureau of Economic Research

• **JESSE ROTHSTEIN**, professor of public policy and economics, University of California, Berkeley, and research associate, National Bureau of Economic Research

SOURCES: Institute of Education Sciences, U.S. Department of Education

### RELATED BLOG



[Visit this blog.](#)