

EDUCATION WEEK

Published Online: May 5, 2008
Published in Print: May 7, 2008

Updated: May 6, 2008

Eye  Research

[Complete Coverage ▶](#)

Scrutiny Heightens for 'Value Added' Research Methods

By **Debra Viadero**


Madison, Wis.

[← Back to Story](#)

As value-added research designs gain in popularity and undergo increasing scrutiny, experts are beginning to wave cautionary flags about how best to make use of them in education.

"We're making progress," said Douglas N. Harris, an assistant professor of educational policy studies at the Wisconsin Center for Educational Research, based at the University of Wisconsin-Madison. "But we need to be cautious in using value-added modeling in a high-stakes way. Using it at the school level is probably a better approach than we have now, but I think it's more ambiguous whether we should go to teacher-level value-added."

Mr. Harris helped organize an April 22-24 conference here in which researchers from a variety of academic disciplines presented studies that both bolstered the method's credibility and introduced caveats about its usefulness in schools.

A more pointed critique  of one popular value-added method also appeared in March in the journal *Educational Researcher*. In that article, Audrey Amrein-Beardsley, an assistant professor at Arizona State University in Phoenix, contends that the model hasn't been shared widely enough or independently tested enough to merit its widespread use in any kind of educational accountability system. (**'Value Added' Pioneer Says Stinging Critique of Method Is Off-Base,** May 7, 2008.)

"My personal opinion is that this model is promising way more than it can deliver," she said in an interview with *Education Week*. "The problem is that when these things are being sold to superintendents, they don't know any better."

Value-added techniques for measuring student achievement appeal to administrators and policymakers at all levels of education because they quantify the gains that students make from one school year to the next rather than relying on reporting percentages of students



ACCEPTABLE

Access to the v
<http://ad.doubleclick.net/a1;sz=300x250;tile=3;0>
listed in the category of: **W**
has been denied because it may viol
District's Acceptable Use Policy (

who get passing or proficient scores at the end of the school year.

Theoretically, that means that teachers or schools get credit only for the value they add to students' learning trajectories, and that they don't get penalized for any learning gaps that students bring with them on the first day of school.

Used in Tennessee and in the Dallas school district for a decade or more, value-added models for tracking students' academic progress first drew national attention in the 1990s when studies documented the dramatic impact that a good—or bad—teacher can have students' academic progress.

The methodology got another boost from the federal No Child Left Behind Act in 2002 because the law provided an impetus for states to develop sophisticated data-collection systems that lend themselves to value-added analyses.

Tied to Pay

Researchers have been using the methods all along to tap into the student-achievement data piling up in states. By tracking students by classroom, and over time, they aim to ferret out the characteristics that mark high-quality teachers so that schools can better focus on how to recruit, train, and retain them.

Robert Meyer, the director of the Value-Added Research Center here at the university, also estimates that 50 districts experimenting with pay-for-performance programs also are using the method to help them pinpoint which teachers qualify for bonuses or raises.

Meanwhile, the U.S. Department of Education recently gave the go-ahead for every state to apply for approval to test "growth models"—which are akin to value-added research designs—as a way to measure whether schools are making adequate yearly progress, or AYP, under the NCLB law.

Amid all the activity, exactly what constitutes a value-added system is not always clear. For example, said Adam Gamoran, who directs the Wisconsin research center, the growth models so far approved by the federal Education Department are not pure value-added systems, principally because the growth models focus more on projected rather than actual growth by gauging whether students are on track to attain proficiency by the 2013-14 school year, as the NCLB law requires.

"This is no mere academic exercise," Mr. Gamoran said of the effort to air issues surrounding value-added models at the University of Wisconsin conference. "This is a crucial policy issue in the landscape of education."

The two-day event was jointly sponsored by the Carnegie Corporation of New York, the Joyce Foundation, and the Spencer Foundation. (All three foundations underwrite coverage in *Education Week*.)

Top Rankings Change

One reason that some researchers are skeptical about using the methodology to reward or

punish teachers is that the results can vary sharply over time.

For example, researchers who used the methodology to rank teachers in seven large Florida school districts found that the proportion of teachers who stayed in the top quintile in terms of advancing student test scores from one year to the next ranged from about a quarter in one district to about a half in another.

Tim R. Sass, the Florida State University professor who presented those findings at the Wisconsin conference, characterized correlations in the percentages of teachers who stayed in the same quintile from one year to the next as "moderate."

But, he added, "if anybody's going to be using these things for high-stakes policy decisions, we want to add a large grain of caution here." The percentages of top-ranking teachers might be more stable from year to year, he added, if researchers could adjust the data to take into account differences in teachers' class sizes—a technique his research team has yet to try.

Part of the problem for statisticians is that value-added calculations operate on several assumptions that have yet to be proved. For example, results might be biased if it turns out that a school's students are not randomly assigned to teachers—if, for instance, principals routinely give high-achieving students to the teachers who are considered the school's best.

Several studies presented at the conference examined that assumption with districtwide testing data from Los Angeles and statewide results from North Carolina and Texas exams. But the researchers arrived at different conclusions about the extent to which principals intentionally assign particular students to particular classrooms and the extent to which it matters for the results, with one determining that it's not a problem and two others suggesting that it could be.

If future studies confirm the problem, said Jane Cooley, an assistant professor of economics at the University of Wisconsin-Madison, "the implication is that we could be rewarding or punishing teachers incorrectly."

Missing test-score data, an issue with any standardized assessment system, can also pose particular problems with value-added models because they rely on data from several school years, thus compounding the overall amount of data that will be missing, according to experts.

Some conference-goers also noted that value-added results are only meaningful if the scales for scoring the tests on which they are based are evenly spaced, so that a 1-point difference all along the test scale translates to a 1-point increment of learning.

But that would mean that every question is of equal difficulty, said Dale Ballou, an associate professor of public policy and education at Vanderbilt University, in Nashville, Tenn., and some statisticians and psychometricians are skeptical of that idea.

"I, too, have scaling concerns, but I think there are analytical statistics that can be used to combat the most serious problems," said William L. Sanders, who pioneered the use of value-added measurements in education more than 20 years ago. "There are ways to balance the risk versus the reward," added the statistician, who now manages the value-added research

and assessment center for the SAS Institute, a private firm based in Cary, N.C.

Buy-In a Challenge

Researchers, in fact, are already developing more sensitive statistical techniques to account for some of those potential problems and biases. Mr. Meyer, for instance, presented findings showing that models incorporating three or more years of testing data for every student could produce more reliable estimates.

Other research suggested that some worries about value-added models might not be as serious as previously thought. Researchers have long worried, for instance, that results from value-added analyses might be misleading if teachers are not similarly effective with students of different ability levels.

In their study, though, J.R. Lockwood and Daniel F. McCaffrey, from the Pittsburgh office of the RAND Corp. found that such differences account for only 2 percent to 4 percent of the classroom-to-classroom variation in teachers' effects on student achievement.

"In other words, if you give teachers a different group of kids, they would've still gotten answers that differed, but not by a lot," Mr. Lockwood explained.

But the more sophisticated the technique, the less understandable it could become for practitioners. The question is whether the added accuracy will make it harder for teachers and administrators to buy into value-added accountability systems, several experts said.

In tandem with the Wisconsin center, the federally funded Center for Analysis of Longitudinal Data in Education Research, or CALDER, is hosting a May 23 conference in Washington on value-added modeling, specifically for policymakers.

"Value-added models are providing us with new information about teachers, information that we've never had before," said Jane Hannaway, the director of CALDER, which is housed at the Washington-based Urban Institute. "But the information we get from value-added modeling is not perfect information. And we're still learning a lot about what the measures mean and don't mean and, given the limitations of the measures, how we should be using them."

Coverage of education research is supported in part by a grant from the Spencer Foundation.

Vol. 27, Issue 36, Pages 1,12-13