

EDUCATION WEEK

Published Online: October 23, 2012

Published in Print: October 24, 2012, as 'Value Added' Use at Secondary Level Questioned

'Value Added' Measures at Secondary Level Questioned

By **Stephen Sawchuk**

Academic tracking in secondary education appears to confound an increasingly common method for gauging differences in teacher quality, according to two recently released studies.

Failing to account for how students are sorted into more- or less-rigorous classes—as well as the effect different tracks have on student learning—can lead to biased "value added" estimates of middle and high school teachers' ability to boost their students' standardized-test scores, the papers conclude.

"I think it suggests that we're making even more errors than we need to—and probably pretty large errors—when we're applying value-added to the middle school level," said Douglas N. Harris, an associate professor of economics at Tulane University in New Orleans, whose [study](#) examines the application of a value-added approach to middle school math scores.

High-school-level findings from a separate [second study](#), by C. Kirabo Jackson, an associate professor of human development and social policy at Northwestern University in Evanston, Ill., complement Mr. Harris' paper.

"At the elementary level, [value-added] is a pretty reliable measure, in terms of predicting how teachers will perform the following year," Mr. Jackson said. "At the high school level, it is quite a bit less reliable, so the scope for using this to improve student outcomes is much more limited."

Expanded Use

Value-added is a form of statistical modeling that measures students' learning gains over the course of a year, while controlling for factors that could skew the estimates. Teachers' impact on those test scores is estimated as the difference between students' projected growth and their actual performance.

Once primarily used to analyze large sets of education data, the technique has expanded into policy.

States and districts are increasingly coupling value-added data with other sources of information, such as

[← Back to Story](#)

EDUCATION WEEK
Multi-User
LICENSING



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

+ Click here for more info

surveys of students and observations of teachers' instructional skill, to inform teacher evaluations—and in some cases, to identify teachers for performance bonuses or promotions.

The technique has been dogged with controversy, though, especially after newspapers in Los Angeles and New York City released reports on individual teachers based on the calculations, which critics deemed flawed and misleading.

More than 30 states now require evaluations of teachers to include student academic performance, according to the National Council on Teacher Quality, a research and advocacy group in Washington.

Value-added is only one of the measures being used, and its use is **limited to certain subjects and grades**. But its influence appears to be growing.

Tracking Bias

Mr. Harris' paper will be presented at the annual conference of the Association for Public Policy and Management, to be held in Baltimore next month. Mr. Jackson's study was published in working form by the National Bureau for Economic Research in January and updated several times since.

The researchers used similar methodologies to probe the question of how academic tracking might affect the teacher-performance estimates.

Mr. Harris and a University of Wisconsin graduate student, Andrew A. Anderson, examined several years of Florida achievement data. Their question: Is the population of students that takes each set of classes skewed in such a way that it biases the teacher calculations?

Each course in the state's database was coded into one of three types: remedial, midlevel, and advanced. Then, the scholars analyzed whether the value-added estimates for teachers changed when they introduced a control for tracking—in essence, by subtracting out the average achievement difference among students in the relevant tracks.

They found that the estimates differed markedly as a result. After the authors accounted for tracking, between 30 percent and 70 percent of teachers in the sample were placed in a different performance quartile. (The degree of miscategorization depended on which quartile each teacher began in, and which track they predominantly taught.)

The scholars' analysis also showed that teachers who taught more remedial classes tended to have lower value-added scores, on average, than those teachers who taught mainly higher-level classes.

That phenomenon was not due to the best teachers' disproportionately teaching the more-rigorous classes, as is often asserted. Instead, the paper shows, even those teachers who taught courses at more than one level of rigor did better

Studies at a Glance

"Bias of Public Sector Worker Performance Monitoring: Theory and Empirical Evidence from Middle School Teachers"

Douglas N. Harris, Tulane University; Andrew A. Anderson, University of Wisconsin-Madison

Examines six years of Florida performance data between the 2003-04 and 2008-09 school years. The data include some 1.3 million students and 14,500 teachers across more than 330,000 middle school math courses.

Findings: Academic tracking introduces sorting bias into value-added estimates of middle school math teachers' performance.

"Teacher Quality at the High School Level: The Importance of Accounting for Tracks"

when their performance teaching the upper-level classes was compared against that from the lower-level classes.

The idea of tracking is a common-sense scenario for any parent who has faced the politics of middle school, Mr. Harris said. "It's not that surprising when you think about how tracking works," he said. "Part of it is based on whether your parents are the ones who are more savvy about this and are going to call the counselor and lobby for you to be in these higher courses."

But if such bias is not accounted for in policy, a teacher could, in effect, boost his or her value-added score simply by teaching all higher-level courses, the paper notes.

Value-added is less commonly applied at the high school level, and Mr. Jackson's paper questions whether it makes sense to expand its use there.

He based his study on a set of end-of-course test-score data in English 1 and Algebra 1 from North Carolina. Mr. Jackson wanted to determine whether the content of specific tracks—for instance, taking several Advanced Placement courses, participating in a program that teaches study skills, or experiencing a sequence of higher-level science courses—affects the formulas.

"It seems perfectly reasonable and plausible that when you learn Newtonian physics, you're learning skills that make it easier to learn math," he said. "So if you're comparing two algebra teachers, you would want to see the students don't differ in their exposure to mathematical concepts outside of algebra classes."

To find out, he separated students into tracks defined by whether they took a popular sequence of 10 courses, took remedial or advanced English and Algebra 1, and attended the same school.

Poor Predictors

Mr. Jackson found that, after controlling for such tracks, the value-added method was not able to determine top from bottom teachers in Algebra 1 with much precision. In English 1, the estimates picked up almost no meaningful differences between high- and low-performing teachers.

In addition, Mr. Jackson concluded that the estimates were generally poor at predicting how successive classes of students taught by those same teachers would do, calling into question whether policies based on the estimates would have much effect overall in helping to retain the best teachers.

"We know there are other ways in which we could be spending our energy to improve student outcomes," he said. "My takeaway is that this is not it."

Both papers stand somewhat in contrast with other studies that use value-added methodology, most of which have been conducted on elementary-level data.

The Bill & Melinda Gates Foundation's \$45 million Measures of Effective Teaching study, for example, found

C. Kirabo Jackson, Northwestern University

Examines five years of North Carolina performance data spanning 2005 to 2010 and covering nearly 400,000 students, 4,200 English 1 teachers, and 3,500 Algebra 1 teachers.

Findings: Value-added estimates of Algebra 1 and English 1 teachers' performance are prone to "noise" caused by the varying academic experiences of students in different tracks.

SOURCE: *Education Week*

RELATED BLOG



[Visit this blog.](#)

that value-added estimates, over time, tended to be **more predictive** of future teacher performance than measures such as observations performed by principals or students' impressions of their teachers. The value-added estimates were somewhat volatile, though, from year to year. (The foundation also provides grant support for *Education Week's* coverage of business and innovation.)

Concerns that selection bias may skew value-added calculations **have been raised at the elementary level**, too, but other researchers have concluded that using additional years of achievement data helps to mitigate that problem.

Addressing the issue at the middle and high school levels may be a complex matter, Mr. Harris said, because it isn't yet clear whether adding controls, as the scholars have done in their research, accounts for all the ways tracking could introduce bias.

"It's not clear that simply adding that indicator [of tracking] solves the problem," Mr. Harris said. "One reason is that how well we measure tracks, and even what's happening in those different tracks, is going to vary across schools."

For instance, the factors that determine how students are assigned to tracks may differ from school to school, making it more difficult to properly control for them, he said.