Researchers evaluate metrics for measuring schools and teachers

June 14 2017, by Peter Dizikes

A new study by an MIT-based team of economists has developed a novel way of evaluating Value-Added Models (VAMs), which assess teachers and schools. Credit: MIT News

In recent years, 14 states in the U.S. have begun assessing teachers and

schools using Value-Added Models, or VAMs. The idea is simple enough: A VAM looks at year-to-year changes in standardized test scores among students, and rates those students' teachers and schools accordingly. When students are found to improve or regress, teachers and schools get the credit or the blame.

Perhaps not surprisingly, however, VAMs have generated extensive debate. Proponents say they bring accountability and useful metrics to education evaluation. Opponents say standardized tests are likely to be a misleading guide to educator quality. Although VAMs often adjust for some differences in student characteristics, educators have argued that these adjustments are inadequate. For example, a teacher with many students trying to overcome learning disabilities may be helping students improve more than a VAM will indicate.

A new study by an MIT-based team of economists has developed a novel way of evaluating and improving VAMs. By taking data from Boston schools with admissions lotteries, the scholars have used the random assignment of students to schools to see how similar groups of students fare in different classroom settings.

"Value-added models have high stakes," says Josh Angrist, the Ford Professor of Economics at MIT and co-author of a new paper detailing the study. "It's important that VAMs provide a reliable guide to school quality."

The researchers have found that existing VAMs tend to underestimate the amount of test score improvement that actually occurs at some schools. On the other hand, the scholars say, conventional VAMs do provide a ballpark figure for improvement that should not be discounted.

"Conventional Value-Added Models are biased, but we're able to show that the bias is modest," says co-author Peter Hull PhD '17, who will

soon join the University of Chicago's economics department as an assistant professor. He adds that, in Boston at least, VAMs "generate useful predictions of school quality."

The same approach that lets the MIT team evaluate VAMs also allows them to show how the metrics may be improved. In so doing, the paper states, the new method could help "improve policy targeting relative to conventional VAMs."

The paper, "Leveraging Lotteries for School Value-Added: Testing and Estimation," appears in the *Quarterly Journal of Economics*. The authors are Angrist; Hull; Parag Pathak, the Jane Berkowitz Carlton and Dennis William Carlton Professor of Microeconomics at MIT; and Christopher Walters PhD '13, an assistant professor at the University of California at Berkeley.

Boston public

The conclusion comes from an analysis of data from Boston's public school system, covering a period from the 2006-07 through the 2013-14 academic years. The data include a sample of roughly 28,000 students at 51 different schools, including some charter and pilot schools.

The test scores of students are taken from fifth- and sixth-grade results in the Massachusetts Comprehensive Assessment System (MCAS), in math and English language arts. The researchers use these data to replicate conventional VAMs and develop their own "hybrid" VAM model that combines the new school-quality estimates with the older approach.

The study exploits the fact that Boston's school system uses a centralized assignment system for students (which was designed in part by Pathak). This system uses a "lottery tie-breaking" feature to help determine which

students will attend schools in high demand. Thus, an element of chance helps determine where a large portion (around 77 percent) of sixth-graders will be enrolled in middle school. This, in turn, gives the researchers the random assignment they need to derive higher-resolution comparisons of the effects schools have on student achievement.

Because the students in this pool of applications differ (on average) only in where they were offered a place, researchers can make apples-to-apples comparisons to see how the students who are admitted via lottery perform, compared to those who were not admitted. The differences in performance then reflect school quality rather than differences in ability or family background.

By contrast, when comparing two schools without use of random assignment, it can be very difficult, if not impossible, to ensure that the students being evaluated are otherwise similar. In this scenario, what might look like a lack of student achievement, using a conventional VAM estimate, could result from a school having a larger number of disadvantaged students.

The study itself shows the difference created by the new VAM technique through a hypothetical scenario involving school closure and expansion: Suppose the lowest-rated Boston school were replaced by a school where students showed the average amount of improvement on test scores. In that case, the researchers find, those scores would increase by 0.24 of a [standard deviation](#) when judged by a conventional VAM method, and 0.32 of a standard deviation when using the new method. This reflects "the usefulness of conventional VAMs, despite their inability to perfectly control for student ability," as Hull observes.

Similarly, if replacing the lowest-ranked school in the survey with a a top-quintile school, student test scores would improve by 0.39 of a standard deviation using a conventional VAM, and 0.53 of a standard deviation

when using the MIT team's own VAM method.

The debate rolls on

The paper's authors note that the findings are situated within some broader political debates about education systems in general. Charter schools are often a subject of considerable public debate, since they receive public funding but may be privately operated and staffed by nonunion teachers, in contrast to traditional public schools. Pilot schools are a hybrid model, with more room for variations in scheduling and curriculum than most public schools, but with unionized teachers.

The 14 states using test-score based VAMs for policymaking are Alabama, Arizona, Florida, Indiana, Louisana, Maine, Mississippi, New Mexico, North Carolina, Ohio, Oklahoma, Texas, Utah, and Virginia.

In any case, Angrist notes, the topic of school performance is a vital one for researchers to examine and for educators to evaluate. Indeed it may be more pressing, he notes, in school districts where test scores have been perennially low, and where larger disparities in school quality may exist.

"For lower-income families, this is fateful," Angrist observes.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/*), a popular site that covers news about MIT research, innovation and teaching.*

Provided by Massachusetts Institute of Technology