

6

Developing Value-Added Measures for Teachers and Schools

Eric A. Hanushek

Caroline M. Hoxby

Measuring schools' and teachers' contributions to student achievement is a central part of school accountability systems. The key idea behind accountability systems is providing information and incentives so that the people who staff schools make better decisions and thereby improve student learning. If we can measure teachers' and schools' contributions accurately, we have the information we need to diagnose learning problems, management problems, teachers' skill deficits, and so on.

Because measuring teachers' and schools' contributions to achievement is central to an accountability system, using accurate measures is key. In particular, it is important that teachers and schools be held accountable for *their* contributions to achievement, not outcomes outside of their control. Only part of a student's current achievement is, however, determined by his current teacher and school. His level of achievement also reflects what he learned from past teachers and schools, from his family, and from his peers.

Take, for example, students who come to school less prepared

because they have family-related disadvantages that undermine their readiness to learn. If a teacher is rewarded simply on the basis of her students' achievement when they are tested in her classroom, teachers may be reluctant to work with disadvantaged students—the very students most in need of good instruction.

These issues have been widely recognized in the past. For example, some accountability systems provide information about the family backgrounds of students along with information about test performance. Such systems explicitly or implicitly encourage us to evaluate schools relative to others that are “comparable”—that is, other schools that serve similar populations. Other accountability systems evaluate a school by comparing its students' average score this year to its students' average score last year. Some year-to-year comparisons are separated by grade so that third graders this year are compared to third graders last year and so on.

The methods (of addressing the problem that a student's current achievement reflects more than just his current teacher and school) that we have just described are common, but they are also crude. That is, they are considerably less informative than methods that we can apply to the data that schools and states have today.

Today, we can use longitudinal data on students to compute the specific contribution that an individual teacher makes to student learning. This individual contribution is often referred to as a teacher's “value-added.” Similarly, we can determine what each school's value-added is—that is, the contribution made by a specific school to student learning. Value-added information for teachers and schools can be used for a variety of purposes. Administrators can use it to discover the strengths and weaknesses of their team. Parents can use it as the basis of discussions with the school about their expectations. Legislators and other high-level policymakers can use it to evaluate policies they enact.

Value-Added Measures for Teachers and Schools 101

Perhaps most importantly, value-added information can be used as the basis of schemes that reward successful teachers and schools.

To compute a teacher's value-added, we do not give her credit for her students' incoming level of achievement, but we do give her credit for *gains* in achievement that students regularly make under her guidance. In this way, we separate a teacher's effect from what her students bring to the classroom. Using parallel procedures, we can compute a school's value-added. Computations of value-added typically account for grade and year effects, which arise if a test is particularly hard in a certain grade or a certain school year.

All calculations of teachers' and schools' value-added are based on very similar and very conventional statistical analyses. The statistics involved are not at all difficult and alternative procedures that are equally valid tend to produce similar results.¹A

1. For recent computations see Daniel Aaronson, Lisa Barrow, and William Sanders, "Teachers and Student Achievement in Chicago Public High Schools" (working paper 2002-28, Federal Reserve Bank of Chicago, June 2003); Jonah Rockoff, "The Impact of Individual Teachers on Student Achievement: Evidence from the Panel Data," *American Economic Review* 94, no. 2 (2004): 247-52; Jonah Rockoff, "Essays on the Finance and Production of Public Education" (Ph.D. dissertation, Harvard University, 2004); and Eric A. Hanushek, et al., "The Market for Teacher Quality" (working paper 11154, National Bureau of Economic Research, Feb 2005).

For comparisons among procedures see Dale Ballou, William Sanders, and Paul Wright, "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational Behavioral Statistics* 29, no. 1 (Spring 2004): 37-65; and Hanley Chiang and Caroline M. Hoxby, "From Teacher Effects to Teacher Rewards: The Empirics of Computing and Rewarding Teachers' Contributions to Student Achievement" (working paper, NBER, 2005). Chiang and Hoxby describe a variety of practical considerations a state should understand before beginning its computations.

See also William Sanders and Sandra P. Horn, "The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment," in *Teacher Evaluation: Guide to Effective Practice*, ed. Anthony J. Shinkfield and Daniel L. Shufflebeam (Boston: Kluwer Academic Publishers, 1995), 337-76; and Sanders and Horn, "Research Findings from the Tennessee

state need not worry that it is putting itself in the hands of a individual statistician if it decides to compute value-added. State statisticians can be trained to make the value-added computations as matter of routine.

Research on teachers' and schools' value-added provides a number of lessons, two of which are especially important.

First, assessments of value-added must take students' initial knowledge into account. Teachers should not be penalized for having students who start behind; nor should they be rewarded for having students who start ahead. Essentially, attention should focus on annual gains in student achievement. By comparing students' scores when they enter a classroom to their scores at the end of the school year, allowance is made for differences in students' starting knowledge and abilities. Thus, it is easiest to compute value-added when students are tested every year, in every grade. No Child Left Behind currently sets minimum grade and subject requirements for state testing. A state that intends to make value-added computations may want to test more grades and more subjects than it is required to test. Also, a state should ensure that its test-maker provides scale scores (scores designed to facilitate computing the improvement in a student's achievement) and not merely proficiency levels.

Second, a student's score on any one test is an imperfect measure of his achievement. A student's measured achievement will fluctuate somewhat depending on the specific test and test-taking circumstances (such as whether the student has a cold on the day of the test). This fluctuation is known as measurement error. Measurement error on individual students' tests averages out

Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education* 12 (1998): 247–56. Although Sanders' analyses of value-added are often described as proprietary, the statistical methods he uses are conventional and not mysterious.

Value-Added Measures for Teachers and Schools**103**

when we observe numerous students associated with a teacher: some students might be having bad days but others will be having good days. Measurement error can also affect a whole cohort of students (all the students in a particular grade in a particular year). For instance, the whole cohort of third-graders in a school may have colds on a particular testing day. However, if we look over a few years, it is unlikely that all the cohorts have colds. In short, it is useful to have multiple years of data for a teacher before we attach consequences to the value-added we compute for her. A reasonable rule appears to be that a state needs a minimum of three annual observations for a teacher before her value-added measure triggers rewards or other incentives. Fewer observations can be used for preliminary information that could be shared with teachers for their self-assessment. More observations will be needed for a teacher who has just a few students in her classroom each year.

These lessons suggest that we cannot readily use value-added to assess starting teachers or teachers whose subjects are not tested. For this reason, evaluations and assessments that move beyond test scores can be a useful complement to value-added calculations in a state's accountability system. It is worth emphasizing that value-added can be based on outcomes other than test scores, so long as they are annual and measured in absolute terms. For instance, we could compute a teacher's contribution to regular attendance or to (a lack of) disciplinary infractions.

Although the discussion to this point has concentrated on the value-added of teachers, computing the value-added of schools is analogous. As with individual teachers, it is important to allow for differences in students' initial achievement. The current school assessments under No Child Left Behind do this partially by using one of the "crude" methods described above. For instance, some states compare the achievement of this year's cohort of students to last year's.

The issues related to measurement error also apply to computation of schools' value-added.² In particular, we may need several years of data from a school that enrolls only a small number of students before we can compute its value-added with confidence.

The bottom line is that, by computing teachers' and schools' value-added, we can measure the contribution of individual teachers and schools in manner that is far more informative than are common measures of performance. While there are a few issues involved in computing value-added (mentioned above), the big picture is that they are based on conventional statistical methods so that states' statisticians can compute them as a matter of routine. Value-added measures can be used for self-assessment, diagnosing learning problems within a school, informing parents about school performance, and as the basis of reward schemes for teachers and schools.

Recommendations

1. In order to support better accountability and reward systems, Arkansas should track the performance of individual students and should calculate the value-added by teachers and schools to pupil achievement in each of the tested subjects.
2. The calculation of value-added should use established statistical procedures and should be subject to verification by independent contractors.

2. Thomas J. Kane and Douglas O. Staiger, "Volatility in School Test Scores: Implications for Test-Based Accountability Systems," in *Brookings Papers on Education Policy 2002*, ed. Diane Ravitch (Washington, D.C.: Brookings, 2002), 235–69.