Value-added measures of teachers: Research and policy

Value-added models in education are used to attempt to measure the contributions to student achievement of individual teachers. Test scores for a particular teacher's students are compared to those of the same students in the previous year, as well as to those of students with other teachers in the same grade, in an effort to isolate the contribution of the given teacher. Advocates of these methods argue that these measures provide objective information that can be used to improve instruction, while critics counter that their validity as an indicator of teacher quality is still in question. School districts from Washington D.C. to Los Angeles have started to use value-added measures, and some teachers' ratings have been made publicly available, including recently in the Los Angeles Times *and the* New York Times.

On June 27, 2012, as part of IRP's annual Summer Research Workshop, three researchers participated in a roundtable discussion of teacher value-added measures. Raj Chetty presented outcomes from a long-term study of the effects of teachers on students from elementary school through early adulthood. Jesse Rothstein explored the potential dangers of using value-added measures to make teacher personnel decisions. Finally, Eric Hanushek looked at policy implications from a different perspective, exploring why and how value-added measures can be used most effectively. This set of articles summarizes the three presentations.

The long-term effects of teachers

There is considerable debate about the best way to measure and improve teacher quality. One method is to rate teachers based on their students' test score gains, known as the "value-added" approach.[1] School districts have begun to use these measures to make personnel decisions about teachers. For example, District of Columbia Public Schools lay teachers off or offer them bonuses using a teacher-performance assessment system that puts 50 percent weight on value-added scores. This article describes an extensive study designed to estimate the effects of teachers on student outcomes through early adulthood.[2]

Potential issues with value-added models

The debate about using teacher value-added models stems primarily from three issues. First is concern about the potential for bias in value-added estimates; do differences in test-score gains across teachers capture causal effects, or are they instead driven by student sorting? Second is lack of evidence on the long-term effects of teachers; do teachers who raise test scores also improve students' long-term outcomes, or are they simply better at teaching to the test? The third issue is measurement error; are estimates based on only a few years of data accurate enough to be used for policy decisions?

Assessing teachers over the long-term

The study done by Raj Chetty, John Friedman, and Jonah Rockoff addresses all of the above issues by using data on a million children, from childhood through early adulthood.[3] The researchers developed new quasi-experimental tests to assess bias in value-added estimates. They look at whether those who had high value-added teachers as children have better outcomes in adulthood. Finally, they assess the monetary gains to be made by selecting teachers with higher estimated value-added scores, given observed measurement error.

The estimated teacher effects include both direct and indirect outcomes. For example, having a good teacher in the fourth grade can improve a student's labor market outcomes in adulthood directly, but it can also have indirect effects, if students with good teachers in the fourth grade receive better teachers in subsequent grades.[4]

Are value-added estimates unbiased?

Since students are not assigned to teachers randomly, it is necessary to adjust for the composition of students assigned to a classroom; the standard approach is to control for prior year variables. Recent studies have reached conflicting conclusions about whether this approach is sufficient for obtaining consistent estimates of teacher effects.[5]

In this study, the researchers evaluate whether or not value-added estimates are biased by (1) testing for selection on observable characteristics, and (2) using quasi-experimental methods that make use of natural teacher turnover. In order to test for selection on observable characteristics, they look at whether parent characteristics are correlated with teacher value-added scores, and find no relationship. For example, the children of wealthier parents are no more likely to get higher value-added teachers. In order to test for selection on unobservable characteristics, they looked at changes in students' scores in the year before and after a switch in teachers due to teacher turnover. Again, they find no evidence of selection. Therefore, they conclude that their value-added measures provide unbiased estimates of teachers' causal effects on student test scores.

This article summarizes the presentation given by Raj Chetty.

Raj Chetty is Professor of Economics at Harvard University.

Effects on outcomes in adulthood

Next, the researchers assessed whether teachers who raise test scores also improve their students' outcomes in adulthood. They analyze the effects of teachers on three sets of outcomes; college attendance, earnings, and other indicators such as teenage birth rates.

Being assigned to a higher value-added teacher in a single grade significantly raises a student's likelihood of attending college. A one standard deviation increase in the value added of a teacher appears to increase the probability of that student attending college by age 20 by 1.25 percent. Students with higher value-added teachers are also more likely to attend a better college, as measured by projected average earnings at age 30.

Having a higher value-added teacher has a clear statistically significant effect on earnings. An increase in teacher value added of one standard deviation increases annual earnings at age 28 by $182. The lifetime financial value of having a teacher one standard deviation higher is approximately $4,600 per grade.[6]

Having a teacher one standard deviation higher in value added in a single year from grades 4 through 8 reduces the probability of a teen birth for female students by 1.25 percent.[7] Students with higher value-added teachers are also more likely to live in higher socioeconomic status neighborhoods as adults.

Measurement error and policy relevance

Any evaluation of teachers based on value-added measures must rely on only a few years of classroom data. This limited amount of data adds uncertainty to value-added estimates, thus potentially reducing their utility for performance evaluation. In order to evaluate how much the utility is reduced, it is necessary to look at a policy example. Thus, the researchers analyze the effects of retaining or firing teachers on the basis of their value-added scores.

On average, replacing a teacher in the bottom 5 percent with an average teacher for one year raises a child's cumulative lifetime income by $50,000. For a class of average size (28 students), the cumulative lifetime income gains from a high value-added teacher exceed $1.4 million. This is equivalent to $267,000 in present value at age 12, discounting future earnings gains at a 5 percent interest rate. Of course, data limitations do not allow certainty about which teachers are in the bottom 5 percent. In estimating the gains of deselecting teachers based on their *estimated* value added, there is still a substantial potential lifetime earnings gain. The present value of earnings gain from deselecting teachers below the fifth percentile increases with the number of classes observed per teacher. While the gain with even ten observed classes is still below the $267,000 value achievable with perfect knowledge of teacher rank, with even three or four observed classes, the lifetime gain is still around $200,000.

Policy implications

While the Chetty and colleagues study supports the idea that existing value-added measures are useful in identifying long-term effects of teachers, this conclusion alone is not sufficient to assess value added as a policy tool, for at least two reasons. First, it is necessary to weigh any potential gains against the cost of firing teachers. The researchers' calculations suggest that the financial benefits of such a policy far outweigh the costs. A second and more serious concern not addressed in this study is potential negative behavioral responses to testing when the stakes are so high, such as teaching to the test or even cheating.[8] It is possible that such responses, if sufficiently large, could completely counter any policy gains.

Parents should be interested in knowing the value added of their child's teacher, whether or not that information is useful as a policy tool. This analysis shows that high value-added teachers improve students' achievement and long-term outcomes. The most important lesson of this study is that finding policies to raise the quality of teaching—whether through the use of value-added measures, or through other tools such as salary structure changes or teacher training—is likely to have substantial economic and social benefits in the long run.■

[1]See, for example, E. A. Hanushek, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review Papers and Proceedings* 61, No. 2 (1971): 280–88; and R. J. Murnane, *The Impact of School Resources on the Learning of Inner City Children* (Cambridge, MA: Ballinger, 1975).

[2]The study summarized here is described in detail in R. Chetty, J. N. Friedman, and J. E. Rockoff, "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," NBER Working Paper No. 17699, National Bureau of Economic Research, 2011. http://www.nber.org/papers/w17699.

[3]The data link two large databases: student records from a large school district, including teachers, class assignment, and test scores from 1991–2009 for students in grades 3–8; and U.S. income tax records including both student outcomes (such as earnings, college, and teenage birth) and parent characteristics (such as income, savings, home ownership, mother's age at childbirth, and marital status).

[4]See, for example, T. J. Kane and D. O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER Working Paper No. 14607, National Bureau of Economic Research, 2008.

[5]Kane and Staiger, "Estimating Teacher Impacts on Student Achievement;" and J. Rothstein, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics* 125, No. 1 (2010): 175–214.

[6]A one standard deviation increase in teacher value added in a single grade results in increased earnings at age 28 of $182, which is 0.9 percent of mean earnings in the regression sample. The researchers assume that the percentage gain in earnings remains constant at 0.9 percent over the lifecycle, and that earnings are discounted at a 3 percent real rate (that is, a 5 percent discount rate with 2 percent wage growth) back to age 12, the mean age in the sample. Under these assumptions, the mean present value of lifetime earnings at age 12 in the U.S. population is approximately $522,000. Thus, the financial value of having a one standard deviation higher value added teacher is 0.9 percent of $522,000, or approximately $4,600 per grade.

[7]The "teenage birth" measure indicates whether a tax return was filed that included a dependent born while the mother was a teenager.

[8]G. Barlevy and D. Neal, "Pay for Percentile," *American Economic Review* 102, No. 5 (2012): 1805–1831.