

Effects of value-added policies

Jesse Rothstein

Jesse Rothstein is Associate Professor of Public Policy and Economics at the University of California, Berkeley, and Research Associate at the National Bureau of Economic Research.

It is important to distinguish between two topics that have often been mixed together: (1) the properties of value-added models and (2) the effects of value-added-based policies. Most research to date has focused on the first, nearly always in low-stakes settings, and many researchers and others have drawn strong policy conclusions from that research. But at this point we know very little about the effects of policies that would use value-added scores to make decisions about teachers. That should be the focus going forward. What really matters is not the effect of individual teachers, which is what most research estimates, but the effect of a policy.¹

What do we know about the properties of value-added models?

A considerable amount of research has been devoted to developing models to estimate the contributions of individual teachers to student achievement. It is important to note that the things we have learned about the properties of value-added models nearly always come from low-stakes settings; that is, the value-added calculations for individual teachers have not generally been used to make decisions about teacher retention or bonuses. While much has been learned, there are still many unanswered questions. I'll review here what I see as a few of the most important outstanding issues.

Value-added measures have been shown to have substantial measurement error, although averaging a few years of data does help. The measures are also sensitive to student assignments. We know that assignment of students to teachers is not random, but it remains an open question whether assignment practices introduce large biases in individual teachers' evaluations. In a paper a few years ago, I showed that the available data were consistent with substantial biases or with essentially no bias.² Important papers by Kane and Staiger and Chetty, Friedman, and Rockoff have narrowed the plausible range somewhat.³ However, both the Kane-Staiger and the Chetty and colleagues estimates have had very wide confidence intervals, so we still do not know the importance of biases due to student assignments.

The Chetty and colleagues study revealed an important fact that has not been incorporated into most thinking about value-added models to date. Specifically, they found that teacher effectiveness changes over time: Some teachers

are ineffective at first but improve as they age, while others start better and then burn out. Under a policy that uses value-added measures to fire poor teachers and reward good ones, some teachers fired early for poor student achievement would have improved over time, while some teachers who receive early raises will continue to receive them even if the quality of their teaching declines. Both modeling and policy calculations will need to change to accommodate this fact, which could have important implications for the kinds of cost-benefit analyses that have been done to date (including in the Chetty and colleagues study).

Another unresolved issue is the choice of value-added specifications. Each author tends to focus on his or her preferred value-added model, and it isn't clear how much it matters. An important aspect of this issue is the distinction between within- and between-school comparisons. Researchers typically focus on within-school comparisons, including fixed effects to absorb any between-school differences. There is good reason for this, as while it is barely possible that students are randomly assigned to teachers within schools, it is clearly not the case that students or teachers are randomly assigned to schools. Proposed policy applications of value added, however, will need to make both within- and between-school comparisons. We do not have a consensus about how to do this, nor much evidence about how much it matters.

Finally, Chetty and colleagues show that teacher value added is predictive of students' future wages. However, the strength of this correlation is unknown. If we could measure teachers' impacts on student wages, would we find that their test score impacts (as measured by value added) were good proxies for them? We don't know. We also know very little about the interactions across grades; if a student has a high value-added teacher two years in a row, how should the values be combined to calculate the joint effect? Researchers typically treat the effects as additive, but there's no evidence for this and a good deal of reason to think it is incorrect.

What do we know about the effects of value-added-based policies?

Much less is known about the effects of value-added-based policies than about how to measure the contributions of teachers to student achievement. It is difficult to find studies that show that offering significant bonuses to high value-added teachers in the United States produces significant effects, and some of the highest-quality studies of the issue find no evidence of such effects.⁴ And there is essentially no evidence on the effects of policies that use value added for deselection (i.e., firing) or for professional feedback.

A study by Carrell and West provides a cautionary tale: adjunct Air Force academy professors, whose continued employment depends on their measured teaching performance,

outscored their regular faculty peers on value-added-type measures based on end-of-year tests, but their students performed poorly in follow-on classes.⁵ These results suggest the potential for teacher responses that improve the teacher value-added measure without improving future student outcomes.

What would we expect to happen if teacher policy is based on value added?

In the absence of extensive evidence on the effects of value-added policies, we can still make an educated guess using a long-standing principle in the education field known as Campbell's Law: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."⁶ Campbell also states that "achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways."

Thus, if teachers are told that their jobs depend on having a high value added, we should expect that value added will be high, but also worry that that might come at the cost of teachers not doing things that we would really like them to do, but that are not directly related to value-added scores. For example, since teachers are evaluated based on math and reading scores, they might spend less time teaching subjects that are not covered in achievement tests, such as history. Even within a tested subject, teachers might spend more time on topics that are covered on the test such as analogies, and less on topics that are not such as composition. There is anecdotal evidence that some teachers are unwilling to teach students whom they believe will not improve their value-added score. Teachers might also focus more on short-term learning (such as drills on multiple-choice questions) that is likely to be reflected in test scores, rather than on long-term learning that will serve students better after the tests are done. The Air Force Academy results mentioned above appear to indicate that these kinds of responses can be important.

David Figlio has done a lot of work looking at the unintended effects of school accountability, ranging from suspension of students who are expected to do poorly, to changing the food offered in the cafeteria on test day.⁷ There are a great deal of factors that may affect test scores without affecting learning, and this may not be how we want our school resources to be used. We do not currently have a sense of how large these distortions would be, and thus how much they would undermine a policy that was based on value-added measures, but it does appear possible that they could completely negate the effects of a teacher policy based on value added.

Personnel economists have spent years studying incentive compensation, and there are lessons from that field that clearly apply to education. When a task is multidimensional,

as teaching certainly is, and when a performance measure is subject to influence, as I believe value added is, it is important to ensure that the stakes are low for a particular measure; that multiple measures be used; that human discretion be part of the process; and finally that the process for helping employees improve be separate from the process through which personnel decisions are made. I believe that describes a viable teacher personnel policy, albeit one that looks quite different from what many districts are implementing. What would it take to implement this kind of policy? First there must be lots of administrators, all highly trained and carefully selected. It seems unreasonable for a single principal to be solely responsible for 40 teachers, accompanying staff, and all other aspects of a given school. While the consulting-world standard of one manager for every five workers is not likely to occur in the world of education, perhaps one administrator for every ten teachers is achievable? It is important that the administrator be capable—there is no reason to think that principal quality is any less important than is teacher quality. We should also be thinking at least as much about the best ways to develop and improve staff, rather than firing them. Finally, there should be an incentive pay component, but stakes need to be relatively low so as not to cause too much distortion of outcomes. ■

¹This point was made by D. B. Rubin, E. A. Stuart, and E. L. Zanutto, "A Potential Outcomes View of Value-Added Assessment in Education," *Journal of Educational and Behavioral Statistics*, 29, No. 1 (Spring 2004): 103–116.

²See J. Rothstein, "Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables," *Education Finance and Policy*, 4, No. 4 (Fall 2009): 537–571.

³See: T. J. Kane and D. O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," Working Paper No. 14607, National Bureau of Economic Research, 2008; and R. Chetty, J. N. Friedman, and J. E. Rockoff, "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," Working Paper No. 17699, National Bureau of Economic Research, 2011. In addition, the Gates Foundation's Measures of Effective Teaching (MET) Project recently released results of a large-scale experiment along the lines of that carried out earlier, on a smaller scale, by Kane and Staiger. Unfortunately, the experiment was plagued by high rates of noncompliance, which limited its ability to answer the question at hand. See: T. J. Kane, D. F. McCaffrey, T. Miller, and D. O. Staiger, *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*, MET Project Research Paper, Bill and Melinda Gates Foundation, Seattle, WA, January 2013; and J. Rothstein and W. J. Mathis, Review of *Have We Identified Effective Teachers?* and *A Composite Estimator of Effective Teaching: Culminating Findings from the Measures of Effective Teaching Project*, National Education Policy Center, Boulder, CO, January 31, 2013.

⁴See M. G. Springer, D. Ballou, L. Hamilton, V. Le, J. R. Lockwood, D. F. McCaffrey, M. Pepper, and B. M. Stecher, *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*, National Center on Performance Incentives at Vanderbilt University, Nashville, TN, 2010.

⁵S. E. Carrell and J. E. West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 118, No. 3 (2010): 409–432.

⁶D. T. Campbell, "Assessing the Impact of Planned Social Change," *Evaluation and Program Planning*, 2, No. 1 (1979): 67–90.

⁷See, for example, D. N. Figlio and S. Loeb, "School Accountability," in *Handbook of Economics of Education*, Volume 3, eds. E. Hanushek, S. Machin, and L. Woessmann (The Netherlands: North-Holland, 2001).