

SUSANNA LOEB STANFORD UNIVERSITY

CARNEGIE KNOWLEDGE NETWORK What We Know Series: Value-Added Methods and Applications

HIGHLIGHTS

- Value-added measures are not a good source of information for helping teachers improve because they provide little information on effective and actionable practices.
- School, district, and state leaders may be able to improve teaching by using value-added to shape decisions about programs, human resources, and incentives.
- Value-added measures of improvement are more precise measures for groups of teachers than they are for individual teachers, thus they may provide useful information on improvement associated with practices, programs or schools.
- Many incentive programs for staff performance that are based on student performance have not shown benefits. Research points to the difficulty of designing these programs well and maintaining them politically.
- Value-added measures for selecting and improving programs, for informing human resource decisions, and for incentives are likely to be more useful when they are combined with other measures.
- We still have only a limited understanding of how best to use value-added measures in combination with other measures as tools for improvement.

INTRODUCTION

The question for this brief is whether education leaders can use value-added measures as tools for improving schooling and, if so, how to do this. Districts, states, and schools can, at least in theory, generate gains in educational outcomes for students using value-added measures in three ways: creating information on effective programs, making better decisions about human resources, and establishing incentives for higher performance from teachers. This brief reviews the evidence on each of these mechanisms and describes the drawbacks and benefits of using value-added measures in these and other contexts.

WHAT IS THE CURRENT STATE OF KNOWLEDGE ON THIS ISSUE?

On their own, value-added measures do not provide teachers with information on effective practices. But this does not mean that value-added measures cannot be useful for educators and leaders to improve instruction through other means, such as identifying practices that lead to higher academic achievement or targeting professional development toward teachers who need it most.ⁱ

As discussed in other Carnegie Knowledge Network briefs, value-added measures have shortcomings as measures of teaching performance relative to other available options, although they also have strengths.ⁱⁱ To begin with, the use of student test scores to measure teaching practice is both an advantage and a disadvantage. The clear benefit is that, however imperfect, test scores are direct measures of student learning, which is an outcome that we care about. Students who learn more in school tend to complete more schooling, have greater earnings potential, and lead healthier lives. Measures such as those based on classroom observations and principals' assessments lack that direct

link to valued outcomes; evidence connects these other measures to student outcomes only weakly. Basing assessments of teachers on the learning of students also recognizes the complexity of the teaching process; many different teaching styles can benefit students. Conversely, value-added measures are dependent on specific student tests, each of which is an incomplete measure of all the outcomes we want to see. The choice of test can affect a teacher's value-added estimate, so it is important to choose tests that truly measure desired outcomes and to recognize that tests are imperfect and incomplete.^{III}

Because value-added measures adjust for the characteristics of students in a given classroom, they are less biased measures of teacher performance than are unadjusted test score measures, and they may be less biased even than some observational measures. As described in an earlier brief, some research provides evidence that value-added measures—at least those that compare teachers within the same school and adjust well for students' prior achievement—do not favor teachers who teach certain types of students.^{iv} The adequacy of the adjustments across schools in value-added measures, which would allow the comparison of teachers in different schools, is less clear because schools can contribute to student learning in ways apart from the contribution of individual teachers and because students sort into schools in ways that value-added measures may not adjust for well.^v

While a fair amount of evidence suggests that value-added measures adequately adjust for differences in the background characteristics of students in each teacher's classroom—much better than do most other measures—value-added measures are imprecise. An individual teacher's score is not an accurate measure of his performance. Multiple years of value-added scores provide better information on teacher effectiveness than does just one year, but even multiple-year measures are not precise.

Given the imprecision of value-added measures and their inability to provide information about specific teaching practices, we might logically conclude that they cannot be useful tools for school improvement. But this conclusion would be premature. For state, district, and school leaders, value-added measures may aid in school improvement in at least three ways: improving programs, making decisions about human resources, and developing incentives for better performance.

Program Improvement

When deciding how to invest resources or which programs to continue or expand, education leaders need information on whether programs are achieving their goals. Ideally, they would have access to random control trials of a particular program in its own particular context, but this type of information is impractical in most situations. If the programs aim to improve teacher effectiveness, value-added measures can provide useful evidence. For example, a district might want to know whether its professional development offerings are improving teacher performance in a particular subject or grade. If the program serves large enough numbers, value-added measures can distinguish between the improvement of teachers who participated in the program and those who did not. Value-added measures are likely to be at least as informative as, and probably more informative than, measures such as teachers' own assessments of their experience or student test scores that are not adjusted for prior student characteristics. Databases with student test scores and background characteristics provide similar information, but they require greater expertise to analyze than do already-adjusted value-added scores. That is, assessing programs with value-added measures is easier than it is with test scores alone because the value-added measures account for differences in the students that teachers teach.

One might worry that value-added measures are too imprecise to measure teacher improvement. However, because multiple teachers participate in programs, program evaluators can combine measures across teachers, reducing the problem of imprecision. Imprecision is even more of an issue with value-added measures of teacher *improvement* than it is for measures of *static effectiveness*. If we think about improvement as measuring the difference between a teacher's effectiveness at the beginning of a period and her effectiveness at the end, the change over time will be subject to errors in both the starting and the ending value. Thus, value-added measures of improvement, particularly those based only on data from the beginning and end of a single year—will be very imprecise. We need many years of data both before and after an intervention to know whether an individual teacher has improved.

Imprecision is less of a problem when we look at groups of teachers. As with years of data, the more teachers there are, the more precise the measure of the average effect is. With data on enough teachers, researchers can estimate average improvements of teachers over time, and they can identify the effects of professional development or other programs and practices on teacher improvement. As examples, studies that use student test performance to measure teachers' effectiveness—adjusted for prior achievement and background characteristics-demonstrate that, on average, teachers add more to their students' learning during their second year of teaching than they do in their first year, and more in their third year than in their second. Similarly, a large body of literature finds that professional development programs in general have mixed effects, but that some interventions have large effects. The more successful professional development programs last several days, focus on subject-specific instruction, and are aligned with instructional goals and curriculum.^{vi} The day-to-day context of a teacher's work can also influence improvement. Teachers gain more skills in schools that are more effective overall at improving student learning and that expose teachers to more effective peers.^{vii} They also appear to learn more in schools that have more supportive professional environments. Value-added measures allow us to estimate average improvements over time and link those to the experiences of teachers.

Value-added measures of static effectiveness, in contrast to value-added measures of improvement, can also lead to teaching improvement through a school or district's adoption of better programs or practices. For example, when districts recruit and hire teachers, knowing which pre-service teacher preparation programs have provided particularly effective teachers in the past can help in selecting teachers that are most likely to thrive. Value-added measures can provide information on which programs certify teachers who are successful at contributing to student learning. Similarly, value-added measures can help identify measurable candidate criteria—such as performance on assessments—that are associated with better teaching performance. If teachers with a given characteristic have higher value-added, it might be beneficial for districts to consider that characteristic in the hiring process.^{viii}

Human Resource Decisions

The second means through which value-added measures may be used for improvement is by providing information for making better human resource decisions about individual teachers. While value-added measures may be imprecise for individual teachers, on average teachers with lower value-added will be less effective in the long run, and those with higher value-added will be more effective. Education leaders can learn from these measures. As an example, instead of offering professional development to all teachers equally, decision-makers might target it at teachers who need it most. Or they might use the measures as a basis for teacher promotion or dismissal. While these decisions may benefit from

combining value-added data with other sources of data, value-added can provide useful information in a multiple-measures approach.

Research is only now emerging on the usefulness of value-added measures for making policies about human resources. One urban district gave principals value-added information on novice teachers, but continued to let principals use their own judgment on teacher renewal. Principals changed their behavior as a result, providing evidence that value-added gave them either new information or leverage for acting on the information they already had.^{ix} In another set of districts, the Talent Transfer Initiative identified high-performing teachers and offered them incentives for moving to and staying in low-performing schools for at least two years. The teachers who moved were found to have positive effects on their students in their new schools, especially in elementary schools.^x There are many possible uses of value-added, some of which may lead to school improvement and a more equitable distribution of resources.

There is less evidence on the uses of value-added measures at the school level. An elementary school principal might consider creating separate departments for teaching math and reading if data showed differences in value-added for math and reading among the fifth grade teachers.^{xi} Similarly, a principal could use the data to help make teaching assignments. If a student has had teachers with low value-added for two years in a row in math, she might make sure to give the student a teacher with a high value-added the next year.^{xii} The imprecision of the value-added measures alone suggests that combining them with other measures would be useful for these kinds of decisions. There is now little research on these school-level practices and their consequences.

Incentives

A third potential way to use value-added measures to improve schools is to provide teachers with incentives for better performance. Conceptually, linking salaries to student outcomes seems like a logical way to improve teachers' efforts or to attract those teachers most likely to produce student test score gains. Performance-based pay creates incentives for teachers to focus on student learning, or for high-performing teachers to enter and remain in the profession.

Performance-based pay carries possible drawbacks, though. Value-added measures do not capture all aspects of student learning that matter, and by giving teachers incentives to focus on outcomes that are measured, they may shortchange students on important outcomes that are not. Further, it is difficult to create performance-based pay systems that encourage teachers to treat their students equitably. For example, formulas often reward teachers for concentrating on students more likely to make gains on the test. Performance-pay policies may also discourage teachers from cooperating with each other.

While there is some evidence that performance pay can work, most evidence from the U.S. suggests that it does not increase student performance. The largest study of performance incentives based on value-added measures comes from a Nashville, Tennessee study that randomly assigned middle-school math teachers (who volunteered for the study) to be eligible for performance-based pay or not. Those who were eligible could earn annual bonuses of up to \$15,000. The study showed very little effect of these incentives on student performance.^{xiii} Similarly, a school-based program in New York City that offered staff members \$3,000 each for good test performance overall had little effect on student scores, as did a similar program in Round Rock, Texas.^{xiv} These results do not mean that value-added measures can't be used productively as incentives; evidence from Kenya and India, at least, suggest that they

can.^{xv} But the failure of the programs in the U.S. provides evidence that they are difficult to implement.^{xvi}

The null results from most studies contrast with some recent evidence suggesting that programs that combine value-added with other measures of performance, or which use only other measures, can lead to improvement. The observation-based evaluations in Cincinnati, for example, have led to improvements in teacher effectiveness,^{xvii} as has the IMPACT evaluation system in Washington, D.C.^{xviii} Both of these programs provide feedback to teachers on their instructional practices. It is not clear how much of the improvement comes from this feedback in combination with the incentives, and how much comes from the incentives on their own.

Summary

While value-added measures may be used as instruments of improvement, it is worth asking whether other measures might be better tools. As noted, value-added measures do not tell us which teacher practices are effective, whereas other measures of performance, such as observations, can. But these other measures can be more costly to collect, and we do not yet know as much about their validity and bias.^{xix} Value-added may allow us to target the collection of these measures, which in turn can help school leaders determine which teachers could most benefit from assistance.

WHAT MORE NEEDS TO BE KNOWN ON THIS ISSUE?

While there is substantial information on the technical properties of value-added measures, far less is known about how these measures are put into actual use. It may, for example, be helpful to assign more students to the most effective teachers, to give less effective teachers a chance to improve, to equalize access to high-quality instruction, and to pay highly effective teachers more. But we have no evidence of the effectiveness of these approaches. Some districts are using value-added measures to make sure that their lowest performing students have access to effective teachers.^{xx} But, again, we know little about the outcomes of this practice. There is also little evidence that value-added can be justified as a reason for more intensive teacher evaluation or can give struggling teachers more useful feedback.^{xxi} Even formal observations (generally three to five per year) provide limited information to teachers. With triage based on value-added measures alone, teachers might receive substantially more consistent and intensive feedback than what they now get. These approaches are promising, and research could illuminate their benefits and costs.

We could also benefit from more information on the use of value-added as a performance incentive. For example, while we have ample evidence of unintended consequences of test-based accountability—as well as evidence of some potential benefits—we know less about the consequences of using value-added measures to encourage educators to improve.^{xxii} Given that performance incentives have little effect on teacher performance in the short run, it may not be worth pursuing them as a means for improvement at all. However, incentives may affect who enters and stays in teaching, as well as teacher satisfaction and performance, and we know little about these potential effects. Moreover, the research suggests that value-added scores may be more effective as incentive tools when they are combined with observational and other measures that can give teachers information on practices.^{xxiii} It would be useful to have more evidence on combining measures for this and other purposes.

WHAT CANNOT BE RESOLVED BY EMPIRICAL EVIDENCE ON THIS ISSUE?

While research can inform the use of value-added measures, most decisions about how to use these measures require personal judgment, as well as a greater understanding of school and district factors than research can provide. The political feasibility and repercussions of the choices will vary by jurisdiction. Some districts and schools are more accepting than others of value-added as a measure of performance. And value-added provides more relevant information in some districts than it does in others. When school leaders already have a good understanding of teachers' skills, value-added measures may not add much, but they may be helpful for leaders who are new to the job or less skilled at evaluating teachers through other means. Moreover, in some places, the tests used to calculate value-added are better measures of desired outcomes than they are in others. Tests are an interim measure of student learning, one indicator of how prepared students are to succeed in later life. Communities will no doubt come to different conclusions about how relevant these tests are as measures of the preparation they value.

The capacity of systems to calculate value-added and to collect other information will also affect the usefulness and cost-effectiveness of the measures. Some critical pieces must be in place if value-added is to provide useful information about programs. Data on crucial student outcomes, as measured by well-designed tests, must be collected and linked from teachers to students over time. Without this capacity, collecting and calculating value-added measures will be costly and time-consuming. With it, the measures can be calculated relatively easily.

Value-added measures will better assess the effectiveness of programs serving teachers, and that of the teachers themselves, than will other measures—such as test score comparisons—which do not adjust for the characteristics of students. For estimating the effects of programs or practices for the purpose of school improvement, value-added measures are not superior to random control trials. That is because these trials do a better job adjusting for differences in teaching contexts. However, experimental approaches are often impractical. By contrast, value-added approaches allow for real analyses of all teachers for whom test scores are available.

CONCLUSION

Value-added measures clearly do not provide useful information for teachers about practices they need to improve. They simply gauge student test score gains relative to what we would expect. Value-added scores also have drawbacks for assessing individual teacher effectiveness because they are imprecise. While on average they do not appear to be biased against teachers teaching different types of students within the same school, they are subject to measurement error. Nonetheless, they can still be useful tools for improving practice, especially when they are used in combination with other measures.

In particular, value-added measures can support the evaluation of programs and practices, can contribute to human resource decisions, and can be used as incentives for improved performance. Value-added measures allow education leaders to assess the effectiveness of professional development. They can help identify schools with particularly good learning environments for teachers. They can help school leaders with teacher recruitment and selection by providing information on preparation programs that have delivered effective teachers in the past. They can help leaders make decisions about teacher assignment and promotion. And they may even be used to create incentives for better teacher performance.

Value-added is sometimes better than other measures of effectiveness, just as other measures are sometimes superior to value-added. Value-added data can help school leaders determine the benefits of

a particular program of professional development. New observational protocols are likely to be more useful than value-added measures because they provide teachers with information on specific teaching practices. But these observational tools have other potential drawbacks, and they have not been assessed nearly as carefully as value-added measures have for validity and reliability. In addition to being costly to collect, observational measures that favor particular teaching practices may encourage a type of teaching that is not always best. Observational measures also may not adjust well for different teaching contexts and even when educators use these protocols, additional, less formal feedback is necessary to support teachers' development.

Value-added measures are imperfect, but they are one among many imperfect measures of teacher performance that can inform decisions by teachers, schools, districts, and, states.

ENDNOTES

ⁱ Harris, D. N. (2011). *Value-Added Measures in Education What Every Educator Needs to Know*. Cambridge, MA: Harvard Education Press.

^{II} Harris, D. N. (May 2013). How Do Value-Added Indicators Compare to Other Measures of Teacher Effectiveness? Carnegie Knowledge Network. <u>http://www.carnegieknowledgenetwork.org/wp-</u>

content/uploads/2012/10/CKN 2012-10 Harris.pdf

^{III} McCaffrey, D. (October 2012). "Do Value-Added Methods Level the Playing Field for Teachers?" Carnegie Knowledge Network. <u>http://www.carnegieknowledgenetwork.org/wp-content/uploads/2013/06/CKN_2012-</u>10 McCaffrey.pdf

^{iv} Ibid.

^v Raudenbush, S. (August 2013). "<u>What Do We Know About Using Value-Added to Compare Teachers Who Work in</u> <u>Different Schools?</u>" Carnegie Knowledge Network. <u>http://www.carnegieknowledgenetwork.org/wp-</u> content/uploads/2013/08/CKN Raudenbush-Comparing-Teachers FINAL 08-19-13.pdf

^{vi} Goldhaber, D. (November 2013). "What Do Value-Added Measures of Teacher Preparation Programs Tell Us?" Carnegie Knowledge Network. <u>http://www.carnegieknowledgenetwork.org/wp-content/uploads/2013/11/CKN-Goldhaber-TeacherPrep_Final_11.7.pdf</u>

^{vii} Rockoff, J.E., Staiger, D., Kane, T., and Taylor, E. (2012). "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools." *American Economic Review*.

http://www.nber.org/papers/w16240.pdf?new_window=1

^{viii} Goldhaber, D., 2013, ibid

^{ix} Rockoff, et al., 2012, ibid.

^x Glazerman, S., Protik, A., The, B., Bruch, J., and Max, J. (2013). "Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment." Princeton: Mathematica Policy Research. http://ies.ed.gov/ncee/pubs/20144003/pdf/20144003.pdf

^{xi} Fox, L. (2013). "Using Multiple Dimensions of Teacher Value-added to Improve Student-Teacher Assignments." Working Paper.

^{xii} Kalogrides, D., & Loeb, S. (Forthcoming). "Different teachers, different peers: The magnitude of student sorting within schools." *Educational Researcher*.

^{xiii} Springer, M., & Winters, M.A. (2009). "New York City's School-Wide Bonus Pay Program: Early Evidence from a Randomized Trial." TN: National Center on Performance Incentives at Vanderbilt University.

https://my.vanderbilt.edu/performanceincentives/files/2012/10/200902_SpringerWinters_BonusPayProgram15.p df

A similar study in New York City also found no effects of financial incentives on student performance, see: Fryer, R. (Forthcoming). "Teacher Incentives and Student Achievement: Evidence form New York City Public Schools." *Journal of Labor Economics*.

^{xiv} Springer, 2009, ibid.

Springer, M., J. Pane, V. Le, D. McCaffrey, S. Burns, L. Hamilton, and B. Stecher. (2012). "Team Pay for Performance: Experimental Evidence From the Round Rock Pilot Project on Team Incentives." *Educational Evaluation and Policy Analysis*, *34*(4), 367–390.

Yuan et. al. (2012). "Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices: Results from Three Randomized Studies." *Educational Evaluation and Policy Analysis* 35 (1), 3-22.

^{xv} Muralidharan, K. & Sundararaman, V. (2011). "Teacher Performance Pay: Experimental Evidence from India." Journal of Political Economy, University of Chicago Press, 119(1), 39-77.

http://www.nber.org/papers/w15323.pdf?new_window=1

Glewwe, P., Ilias, N, and Kremer, M. (2010). "Teacher Incentives." *American Economic Journal: Applied Economics*, 2(3), 205-27. <u>http://www.povertyactionlab.org/publication/teacher-incentives</u>

^{xvi} A recent study finds that incentives are more effective if they are incentives to avoid harm—such as dismissal or loss of salary—instead of incentives for promotion or higher pay. See:

Fryer, R., S. Levitt, J. List, and S. Sadoff. (2012). "Enhancing the Efficacy of Teacher Incentives through Loss Aversion." NBER Working Paper, No 18237. <u>http://www.nber.org/papers/w18237</u>

^{xvii} Taylor, E.S. and Tyler, J.H. (2012). "The Effect of Evaluation on Teacher Performance." *American Economic Review*, *102*(7), 3628-3651.

^{xviii} Wyckoff, J. and Dee, T.S. (2013). "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." Working Paper.

^{xix} Harris, D. N., 2013, ibid.

^{xx} Tennessee policy.

^{xxi} Harris, D. N. (November 2013). "How Might We Use Multiple Measures for Teacher Accountability?" Carnegie Knowledge Brief. <u>http://www.carnegieknowledgenetwork.org/wp-</u>

content/uploads/2013/11/CKN_2013_10_Harris.pdf

^{xxii} Figlio, D. and Loeb, S. (2011). "School Accountability." In E. A. Hanushek, S. Machin, and Woessmann, L. (Eds.), Handbook of the Economics of Education, Vol. 3, San Diego, CA: North Holland, 383-423.

^{xxiii} Wyckoff, J., 2013, ibid.

AUTHOR



Susanna Loeb is the Barnett Family Professor of Education at Stanford University, Faculty Director of the Center for Education Policy Analysis, and a Co-Director of Policy Analysis for California Education (PACE). She specializes in the economics of education and the relationship between schools and federal, state and local policies. Her research addresses teacher policy, looking specifically at how teachers' preferences affect the distribution of teaching quality across schools, how pre-service

coursework requirements affect the quality of teacher candidates, and how reforms affect teachers' career decisions. She also studies school leadership and school finance, for example looking at how the structure of state finance systems affects the level and distribution of resources across schools. Susanna is a senior fellow at the Stanford Institute for Economic Policy Research, a faculty research fellow at the National Bureau of Economic Research, a member of the Policy Council of the Association for Policy Analysis and Management, and Co-Editor of Educational Evaluation and Policy Analysis.

ABOUT THE CARNEGIE KNOWLEDGE NETWORK

The Carnegie Foundation for the Advancement of Teaching has launched the Carnegie Knowledge Network, a resource that will provide impartial, authoritative, relevant, digestible, and current syntheses of the technical literature on value-added for K-12 teacher evaluation system designers. The Carnegie Knowledge Network integrates both technical knowledge and operational knowledge of teacher evaluation systems. The Foundation has brought together a distinguished group of researchers to form the *Carnegie Panel on Assessing Teaching to Improve Learning* to identify what is and is not known on the critical technical issues involved in measuring teaching effectiveness. Daniel Goldhaber, Douglas Harris, Susanna Loeb, Daniel McCaffrey, and Stephen Raudenbush have been selected to join the Carnegie Panel based on their demonstrated technical expertise in this area, their thoughtful stance toward the use of value-added methodologies, and their impartiality toward particular modeling strategies. The Carnegie Panel engaged a User Panel composed of K-12 field leaders directly involved in developing and implementing teacher evaluation systems, to assure relevance to their needs and accessibility for their use. This is the first set of knowledge briefs in a series of Carnegie Knowledge Network releases. Learn more at <u>carnegieknowledgenetwork.org</u>.

CITATION

Loeb, Susanna. Carnegie Knowledge Network, "How Can Value-Added Measures Be Used for Teacher Improvement?" December 2013.

http://www.carnegieknowledgenetwork.org/briefs/value-added/teacher_improvement/



Carnegie Foundation for the Advancement of Teaching 51 Vista Lane Stanford, California 94305 650-566-5100

Carnegie Foundation for the Advancement of Teaching seeks to vitalize more productive research and development in education. We bring scholars, practitioners, innovators, designers, and developers together to solve practical problems of schooling that diminish our nation's ability to educate all students well. We are committed to developing networks of ideas, expertise, and action aimed at improving teaching and learning and strengthening the institutions in which this occurs. Our core belief is that much more can be accomplished together than even the best of us can accomplish alone.

www.carnegiefoundation.org

We invite you to explore our website, where you will find resources relevant to our programs and publications as well as current information about our Board of Directors, funders, and staff.



This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 Unported license. To view the details of this license, go to: http://creativecommons.org/licenses/by-nc/3.0/.

Knowledge Brief 13 December 2013 carnegieknowledgenetwork.org

Funded through a cooperative agreement with the Institute of Education Sciences. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.