



**Carnegie Foundation** for the Advancement of Teaching

# **CARNEGIE KNOWLEDGE NETWORK CONCLUDING RECOMMENDATIONS**

**DAN GOLDHABER**

**DOUGLAS N. HARRIS**

**SUSANNA LOEB**

**DANIEL F. MCCAFFREY**

**STEPHEN W. RAUDENBUSH**

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

**ATIL**

ADVANCING TEACHING - IMPROVING LEARNING

## CONCLUDING RECOMMENDATIONS

It is common knowledge that teacher quality is a key in-school factor affecting student achievement. While the quality of teaching clearly matters for how much students learn, this quality is challenging to measure. Evaluating teacher quality based on the level of their students' end-of-year test scores has been one method of assessing teacher quality, but this approach favors those teachers of students who begin the year already at a high academic level. Value-added methodology is an alternative use of annual test score data to assess teacher quality. As opposed to using the *level* of student achievement on a test at the end of the year to assess a teacher's effectiveness, value-added methodology seeks to isolate a teacher's contribution to student *growth* from other factors that contribute to student achievement, such as prior achievement or socio-economic status. Methodologists generally agree that value-added estimates can serve as a gauge of some dimensions of teacher effectiveness.

Yet the experts also agree that value-added measures have significant limitations. Most clearly, they do not capture a teacher's entire contribution to student learning because they are based solely on test performance, itself an imprecise signal of teacher effectiveness. There is disagreement among experts on the extent to which value added is biased. For instance, there is room for disagreement about whether or how much statistical models can or should account for factors such as poverty, or how teachers can be compared across different classroom contexts. The comparability of value added is greater among teachers working in similar contexts, which is true of all forms of performance evaluation. Moreover, value-added measures are likely to vary according to the tests used to compute them. Because tests cover different content, they reflect different areas of student knowledge. Value added measures what the tests measure, and because these tests capture only a slice of what students are learning, value added reflects only that slice. Teachers could be effective or ineffective on outcomes not captured by students' scores on these tests.

As important as it is to judge value-added measures on their validity and reliability, it is more relevant to know to what extent they can benefit schools and students if used in practice. Largely because value-added policies are so new, we know very little about how they work in practice. Thus there remains considerable debate about how value-added measures should be used to inform personnel policies, if they are to be used at all. As the real impacts of reforms using value added emerge and as researchers assess these effects, we will learn more about the consequences of different uses of value added. Until then, the CKN briefs have given us a rich picture of current research on the use of value-added measures for teacher evaluation. Reviewing the briefs and their implications for practitioners, we arrived at the following recommendations:

*1. When using value added, allow educational leaders to make judgments in interpreting the value-added results in light of other available measures of teacher quality and the principals' own assessments.*

Studies provide evidence that value-added measures meaningfully distinguish between teachers whose future students will consistently perform well and teachers whose students will not.<sup>1</sup> However, the studies also highlight the shortcomings of value added, among them instability in measures across time;<sup>2</sup> systematically lower values for teachers of some types of students;<sup>3</sup> and difficulty in comparing teachers across schools.<sup>4</sup> These issues are potential problems for any form of evaluation, including teacher observation.<sup>5</sup> The principal and other school leaders are often well situated to evaluate the many circumstances in which quantitative measures may

## CONCLUDING RECOMMENDATIONS

give an incomplete picture. For instance, school leaders can see a teacher’s indirect contributions to learning through her overall work in the school, or her contribution to student outcomes not measured by achievement tests. Educational leaders likewise can make allowances for unexpected events in a teacher’s environment—a sudden change in assignment, for instance — that can’t be captured by value-added models. Human judgment also may reduce teachers’ concerns about being assessed only by test scores or being subjected to statistical procedures they don’t trust or understand.

### *2. Use value added with other measures that are valid and have variation.*

Because education and teacher evaluations have multiple goals,<sup>6</sup> assessing progress toward these goals will require multiple measures. Moreover, evaluation programs that combine value added with other measures, at least in some instances, have led to instructional improvement.<sup>7</sup> While ultimately all measures should be evaluated on the same set of standards, each measure will have strengths and weaknesses. Using the measures together helps reduce errors and provides a sense of balance that may make them more acceptable to those being evaluated.<sup>8</sup> When multiple measures are included in an evaluation system, all of the measures must vary across teachers in order to matter at all for the evaluation. A measure that rates all teachers at the same level will carry no weight when it is combined with other measures to classify teacher performance. For example, suppose that a school system uses both observation and value-added scores to determine teacher effectiveness. Then suppose that all of the teachers in the school system are rated proficient according to observations by their principals. The only way a teacher can be deemed “in need of improvement” is if she has a low value-added score. The only way she can be classified as “exemplary” is if she has high value-added score. Even though the system uses two measures, it is only the value-added measure that differentiates on the summative measure.

Multiple measures don’t necessarily have to be combined into a single rating scale; for instance, the value-added estimate could be used mainly to identify candidates for more intensive observation.<sup>9</sup> Schools would then measure the performance of the identified teachers with more intensive measures that can provide a more accurate assessment of teacher performance but are too costly to collect for all teachers. Multiple measures may also take on many forms. Along with observations, scores on student learning objectives, and survey responses,<sup>10</sup> they might also include the value that teachers bring to student outcomes other than achievement test scores. These outcomes might include progressing through courses, graduating from high school, even attending or completing college.<sup>11</sup> If carefully developed and thoroughly evaluated, such long-term measures could help give schools and districts a richer picture of teacher effectiveness.

The decision to use value-added measures depends not on whether they are perfect—all measures are flawed. Rather, it depends on what other measures of effectiveness are available or feasible and the quality of those measures. If principals are skilled, have goals aligned with district or state goals, and have adequate time for evaluation, then they may be a better source of information on teacher effectiveness than value-added measures because they can observe more than test scores and see teachers throughout the course of the year. However, not all schools have principals who can do this effectively.

### *3. Choose a test that measures knowledge that is valued.*

## CONCLUDING RECOMMENDATIONS

Value added is sensitive to whatever test students take; different tests lead to different rankings of teacher effectiveness.<sup>12</sup> If we want value-added scores to yield information about teaching beyond what is currently measured by standardized tests, we must use a test that measures knowledge we value. Schools today focus largely on the content measured by standardized accountability tests. Tests used to determine value added should give teachers incentive to teach knowledge that is valued.

### *4. Consider differences in teaching contexts when using value added to compare teachers.*

Standard methods to statistically adjust for students' prior achievement can produce unbiased (or nearly unbiased) estimates of teacher value added. However, the research supporting this finding is based on studies in which teachers were randomly assigned to students within the same school, mainly within elementary schools, while most value-added systems compare teachers in different schools.

The distinction between within-school and between-school comparisons is an important one, because teachers within the same school share the same organizational conditions (leadership and resources); are subject to similar contextual factors (neighborhood safety, parental support, norms that favor academic achievement); and, particularly in elementary school, they tend to teach students with similar levels of prior achievement. In a heterogeneous district, schools can vary widely by all these factors. If well-managed schools foster better teaching and learning than do poorly-managed schools, then teachers in well-managed schools will outperform equally able teachers in poorly-managed schools. If some contextual conditions promote more student learning than do other conditions, teachers in schools with more favorable conditions will outperform equally able teachers in schools with less favorable conditions. Our confidence in value-added estimations of teachers will increase with the degree of similarity between the classrooms of the teachers being compared. Moreover, most tests are not designed to compare the learning gains of students who start at very different levels of achievement.

For these reasons, when education leaders are making consequential personnel decisions informed by value added, they would be wise to take into account the different contexts in which teachers work. They should also look at value-added scores across the district to detect general patterns in the distribution of teacher effectiveness. This practice serves as a check on the equitability of hiring and placement policies.

### *5. Take specific steps to ensure the overall credibility of the teacher evaluation system.*

Many of the CKN briefs identified threats to the validity of value added as a measure of teacher effectiveness for some teachers.<sup>13</sup> Inaccurate value-added estimates can lead to decisions harmful to both teachers and students.<sup>14</sup> For value-added measures to be useful, they must be subject to the following minimal conditions: the tests on which they are based should reliably capture valued student outcomes; data should accurately identify which students are in which teachers' classes; and teachers should work with enough students to produce value-added estimates that are relatively free of noise.

So school systems must look for patterns that suggest errors and use other data to confirm that teachers who consistently receive high or low value-added scores are truly high- or low-performers. For example, if all the special education teachers receive low value-added scores,

## CONCLUDING RECOMMENDATIONS

but if those scores are belied by classroom observations and student surveys, educational leaders should examine whether there might be problems with one or the other measure.<sup>15</sup> System leaders should clearly document the statistical methods used for calculating value added, taking steps to secure data accuracy. They should provide evidence of that accuracy and of the checks used to ensure the appropriateness of the evaluation models, the checks used to confirm the assumptions made by the models, and the tests of the robustness of the results. When communicating about individual teacher value-added scores, reports should indicate the nature of value-added scores as *estimates*, and be careful not to overstate the degree of their precision.<sup>16</sup>

Evaluation measures that capture meaningful differences in teaching quality are important components of a system that ensures quality learning opportunities for all students. There is great risk of teacher evaluation systems not performing as planned, however. The human resources literature is rife with data and theories on why these performance measurement systems fail to yield the rich information and informed decisions they were intended to provide.<sup>17</sup> Likewise, the literature on performance measurement for public employees offers numerous examples of unintended consequences, including employees gaming the system or trying to improve their performance ratings without improving their actual performance. Edward Deming, the quality improvement expert whose ideas have revolutionized the industry, cautioned organizations against using performance measurement for individuals, believing that doing so would lead to capricious behavior, as well as fear that stifles creativity.<sup>18</sup> This warning should be taken seriously. If school leaders are effectively evaluating and motivating teachers using measures other than value-added measures, then value-added measures may provide little benefit for students. If local evaluations are not effective, then value-added measures may be a beneficial tool, at least until schools implement better systems.

Clearly, it is challenging to use evaluation and performance monitoring to improve outcomes. States and school districts will need to take deliberate steps to increase the chances that their systems will work as intended. Because current evaluations by managers tend to be uniformly high, school systems will need to monitor evaluation data for variability in performance ratings and promote and reward accurate evaluations. Systems must make sure that teachers are not narrowing the curriculum or manipulating student assignments to improve their value-added scores. For instance, schools might assess students periodically on tests that are different from the state test but cover the same standards. Schools might also monitor student and teacher assignments to identify unusual changes over time.

School systems should identify and monitor the factors that should *not* change when value added is used to assess performance by watching for evidence that administrators or teachers are taking unwanted steps to influence their value-added scores. For instance, the proportion of students classified with certain learning disabilities, which is susceptible to manipulation, should not change.<sup>19</sup> School systems should also list the factors that *should* change with a new evaluation system, such as the number of classroom observations performed or the alignment between the type of professional development chosen and teachers' needs. They should then monitor these factors for evidence that changes are occurring, and plan for how to respond if the system fails to behave as expected.

Data from millions of students and thousands of teachers as well as careful thought and analysis have taught us much about the statistical properties of value added. Yet these measures are

## CONCLUDING RECOMMENDATIONS

only just beginning to be put into widespread use for teacher evaluation.<sup>20</sup> As these new systems roll out, states and districts should be learning and experimenting, determining what works and what doesn't. They should use this time to collect data and monitor their evaluation systems, using what they learn to make revisions. They should share their experiences with other school systems and learn from them in turn. The research community must also work with states and districts to identify the best practices for teacher evaluation, assessing whether the new systems really do yield better teaching and learning. Finally, the decisions school systems reach about value added—including whether and how to use the measure for evaluation—should not be seen as one-time events. School systems are now gathering a wealth of data from which we can learn how to make educational organizations more effective at conducting evaluations to improve teaching. Their evaluation systems must make use of that data, and evolve based on what that data shows.

## ENDNOTES

- 
- <sup>1</sup> Daniel F. McCaffrey, “Do Value-Added Methods Level the Playing Field for Teachers?” (Carnegie Knowledge Network Brief No. 2, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2012), <http://www.carnegieknowledgenetwork.org/briefs/value-added/level-playing-field/>; and Stephen W. Raudenbush, “What Do We Know About the Long-Term Impacts of Teacher Value Added?” (Carnegie Knowledge Network Brief No. 15, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, March 2014), <http://www.carnegieknowledgenetwork.org/briefs/long-term-impacts/>.
- <sup>2</sup> Susanna Loeb and Christopher A. Candelaria, “How Stable Are Value-Added Estimates across Years, Subjects, and Student Groups?” (Carnegie Knowledge Network Brief No. 3, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2012), <http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-stability/>.
- <sup>3</sup> Douglas N. Harris and Andrew Anderson, “Does Value Added Work Better in Elementary Than in Secondary Grades?” (Carnegie Knowledge Network Brief No. 7, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, May 2013), <http://www.carnegieknowledgenetwork.org/briefs/value-added/grades/>; and McCaffrey, “Value-Added Methods.”
- <sup>4</sup> Stephen W. Raudenbush, “What Do We Know About Using Value-Added to Compare Teachers Who Work in Different Schools?” (Carnegie Knowledge Network Brief No. 10, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, August 2013), <http://www.carnegieknowledgenetwork.org/briefs/comparing-teaching/>.
- <sup>5</sup> Douglas N. Harris, “How Do Value-Added Indicators Compare to Other Measures of Teacher Effectiveness?” (Carnegie Knowledge Network Brief No. 5, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2012), <http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-other-measures/>.
- <sup>6</sup> Douglas N. Harris, “How Might We Use Multiple Measures for Teacher Accountability?” (Carnegie Knowledge Network Brief No. 11, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2013), [http://www.carnegieknowledgenetwork.org/briefs/multiple\\_measures/](http://www.carnegieknowledgenetwork.org/briefs/multiple_measures/).
- <sup>7</sup> Susanna Loeb, “How Can Value-Added Measures Be Used for Teacher Improvement?” (Carnegie Knowledge Network Brief No. 13, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, December 2013), [http://www.carnegieknowledgenetwork.org/briefs/teacher\\_improvement/](http://www.carnegieknowledgenetwork.org/briefs/teacher_improvement/).
- <sup>8</sup> Harris, “Value-Added Indicators.”
- <sup>9</sup> Harris, “Multiple Measures.”
- <sup>10</sup> Harris, “Value-Added Indicators.”
- <sup>11</sup> Raudenbush, “Long-Term Impacts.”
- <sup>12</sup> Douglas F. McCaffrey, “Will Teacher Value-Added Scores Change when Accountability Tests Change?” (Carnegie Knowledge Network Brief No. 8, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, June 2013), <http://www.carnegieknowledgenetwork.org/briefs/value-added/accountability-tests/>.
- <sup>13</sup> Dan Goldhaber and Roddy Theobald, “Do Different Value-Added Models Tell Us the Same Things?” (Carnegie Knowledge Network Brief No. 4, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2012), <http://www.carnegieknowledgenetwork.org/briefs/value-added/different-growth-models/>.
- <sup>14</sup> Dan Goldhaber and Susanna Loeb, “What Do We Know About the Tradeoffs Associated with Teacher Misclassification in High Stakes Personnel Decisions?” (Carnegie Knowledge Network Brief No. 6, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, April 2013), <http://www.carnegieknowledgenetwork.org/briefs/value-added/teacher-misclassifications/>.

## CONCLUDING RECOMMENDATIONS

---

<sup>15</sup> Douglas N. Harris, “How Might We Use Multiple Measures for Teacher Accountability?” (Carnegie Knowledge Network Brief No. X, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, May 2013), [http://www.carnegieknowledgenetwork.org/briefs/multiple\\_measures/](http://www.carnegieknowledgenetwork.org/briefs/multiple_measures/).

<sup>16</sup> Stephen W. Raudenbush and Marshall Jean, “How Should Educators Interpret Value Added Scores?” (Carnegie Knowledge Network Brief No. 1, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, October 2012), <http://www.carnegieknowledgenetwork.org/briefs/value-added/interpreting-value-added/>.

<sup>17</sup> For a discussion of challenges to using performance measures for incentive programs, see Loeb, “Teacher Improvement.”

<sup>18</sup> Sharon L. Lohr, “The Value Deming's Ideas Can Add to Educational Evaluation,” *Statistics, Politics, and Policy* 3, issue 2 (2012).

<sup>19</sup> Daniel F. McCaffrey and Heather Buzick, “Is Value Added Accurate for Teachers of Students with Disabilities?” (Carnegie Knowledge Network Brief No. 14, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, January 2014), [http://www.carnegieknowledgenetwork.org/briefs/teacher\\_disabilities/](http://www.carnegieknowledgenetwork.org/briefs/teacher_disabilities/).

<sup>20</sup> Dan Goldhaber and Susana Loeb, “What Do We Know About the Tradeoffs Associated with Teacher Misclassification in High Stakes Personnel Decisions?” (Carnegie Knowledge Network Brief No. X, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, May 2013), <http://www.carnegieknowledgenetwork.org/briefs/value-added/teacher-misclassifications/>; and McCaffrey, “Do Value-Added Methods Level?”



## 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 [carnegieknowledgenetwork.org](http://carnegieknowledgenetwork.org).

## CITATION

Dan Goldhaber, Douglas N. Harris, Susanna Loeb, Daniel F. McCaffrey, and Stephen W. Raudenbush, "Carnegie Knowledge Network Concluding Recommendations," Carnegie Foundation for the Advancement of Teaching, Stanford, CA, January 2015.  
<http://www.carnegieknowledgenetwork.org/briefs/concluding-recommendations>



## **Carnegie Foundation** for the Advancement of Teaching

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](http://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/>.

**Carnegie Knowledge Network Concluding Recommendations**  
**January 2015**  
**[carnegieknowledgenetwork.org](http://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.*