



**Carnegie Foundation** for the Advancement of Teaching

# DO VALUE-ADDED METHODS LEVEL THE PLAYING FIELD FOR TEACHERS?

**DANIEL F. MCCAFFREY**  
EDUCATIONAL TESTING SERVICE

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

**ATIL**

ADVANCING • TEACHING • IMPROVING • LEARNING

### HIGHLIGHTS

- Value-added measures partially level the playing field by controlling for many student characteristics. But if they don't fully adjust for all the factors that influence achievement *and* that consistently differ among classrooms, they may be distorted, or "confounded."
- Simple value-added models that control for just a few tests scores (or only one score) and no other variables produce measures that underestimate teachers with low-achieving students and overestimate teachers with high-achieving students.
- The evidence, while inconclusive, generally suggests that confounding is weak. But it would not be prudent to conclude that confounding is *not* a problem for all teachers. In particular, the evidence on comparing teachers across schools is limited.
- Studies assess general patterns of confounding. They do not examine confounding for individual teachers, and they can't rule out the possibility that some teachers consistently teach students who are distinct enough to cause confounding.
- Value-added models often control for variables such as average prior achievement for a classroom or school, but this practice could introduce errors into value-added estimates.
- Confounding might lead school systems to draw erroneous conclusions about their teachers – conclusions that carry heavy costs to both teachers and society.

### INTRODUCTION

Value-added models have caught the interest of policymakers because, unlike using student tests scores for other means of accountability, they purport to "level the playing field." That is, they supposedly reflect only a teacher's effectiveness, not whether she teaches high- or low-income students, for instance, or students in accelerated or standard classes. Yet many people are concerned that teacher effects from value-added measures *will* be sensitive to the characteristics of her students. More specifically, they believe that teachers of low-income, minority, or special education students will have lower value-added scores than equally effective teachers who are teaching students outside these populations. Other people worry that the opposite might be true -- that some value-added models might cause teachers of low-income, minority, or special education students to have *higher* value-added scores than equally effective teachers who work with higher-achieving, less risky populations.

In this brief, we discuss what is and is not known about how well value-added measures level the playing field for teachers by controlling for student characteristics. We first discuss the results of empirical explorations. We then address outstanding questions and the challenges to answering them with empirical data. Finally, we discuss the implications of these findings for teacher evaluations and the actions that may be based on them.

### WHAT IS KNOWN ABOUT HOW VALUE-ADDED MODELING LEVELS THE PLAYING FIELD?

Value-added modeling uses statistical methods to isolate the contributions of teachers from other factors that influence student achievement. It does this by using data for individual students, such as scores on standardized tests, special education and English-learner status, eligibility for free and reduced-price meals (a proxy for poverty), and race and ethnicity. It sometimes controls for the average classroom or school-wide scores on previous years' tests or for Census data, class size, or a principal's years of experience.

Despite these controls, people still worry that value-added estimates may be distorted, or, in statisticians' terms, "confounded." Value-added measures are said to be confounded if they are subject to change because of students' socio-economic backgrounds or other student-level characteristics, and also if teachers who are equally effective have persistently different value-added scores because of the types of students they teach. For example, confounding occurs if teachers of low-income or minority students have lower –or higher – scores than equally effective teachers who teach groups that tend to be higher-achieving. In short, confounding means that we cannot determine the educators' contributions distinct from those of the students they teach.

Confounding might occur because the statistical model doesn't measure or properly control for *all* the factors that contribute to student achievement. For example, it might not fully account for students with unique disabilities.

For confounding to occur, these kinds of factors must *consistently* be associated with the students of a particular teacher, so that they result in a value-added score that *consistently* underestimates or overestimates her effectiveness. For example, suppose that *every year* a fifth grade teacher is assigned highly gifted students whose learning is not captured by the yearly achievement test, and that her value-added measure does not account for the gifted status of these students. We consider this teacher's value-added score to be confounded. It is persistently too low.

On the other hand, suppose that by complete chance, a fifth grade teacher has five highly gifted students in her class *in one year*, but that she does not have similar students *in other years* and might not even have similar students in other class assignments she could have had that same year. This teacher's value-added score is not confounded, even though the model does not completely account for her students' gifted status, because it is only *by chance* that this teacher had such good students, and it was not assumed that her luck would continue. Even if this chance assignment does not result in confounding, it does create an error in the teacher's value-added for the year.<sup>1</sup>

The concern with confounding is that student characteristics will conflate measures of teacher effectiveness in predictable ways: teachers in high-poverty schools might consistently receive scores that are too low, teachers of English language-learners might consistently receive scores that are too high, and so on. To avoid these risks, value-added models must fully control for background variables that could be persistently associated with any teacher.

### RESULTS FROM THE LITERATURE

It is difficult to test for confounding of value-added estimates using data collected on students and their teachers in standard settings. We do not know how effective teachers really are, so we can't compare our value-added estimates to the "truth." Moreover, the only student data we have for teachers are from the types of students who we think might cause confounding, and the very variables we might manipulate to test for confounding are the same ones the value-added model uses as controls in the first place.

Still, researchers have conducted some clever tests to assess the likelihood of confounding. The research supports one conclusion: value-added scores for teachers of low-achieving students are underestimated, and value-added scores of teachers of high-achieving students are overestimated by models that control for only a few scores (or for only one score) on previous achievement tests without adjusting for measurement error.<sup>2,3</sup> An example of such a model is a student growth percentile<sup>4</sup> model used to calculate median growth percentiles for a teacher of fourth grade mathematics that controls only for students' prior third grade mathematics scores.<sup>5</sup>

Beyond this point, the evidence is contradictory. There are studies suggesting confounding does not exist<sup>6</sup> and others suggesting it might.<sup>7</sup> All of the studies have limitations, and because they are so hard to conduct, none is definitive. The evidence may favor the conclusion that confounding generally is weak. But because the evidence supporting this claim comes from just a few limited studies, and because other studies contradict some of that evidence, it would be unwise to conclude that confounding is *not* a problem. Moreover, the studies take in broad samples of teachers; they can't rule out the possibility that some teachers consistently teach students who are distinct enough in some way to cause confounding.

### STUDIES THAT SUGGEST CONFOUNDING DOES NOT EXIST

Two rather compelling studies find no evidence on confounding, although neither has been replicated, and neither addresses all sources of confounding. Clearly, value-added models do not account for every factor that might contribute to student learning; the question is whether they account for *enough* variables so that any factors not controlled by the model are not persistently associated with teachers.<sup>8</sup>

The first study that found no evidence of confounding looked at large samples of students and teachers from a single urban district over several years.<sup>9</sup> It merged student test scores with data from the tax returns of the students' families to see if income and other data from the returns<sup>10</sup> were related with value-added estimates. The study found that the data (not typically available to districts) were not associated with value-added estimates.<sup>11</sup>

The same study used several years of value-added estimates to devise another clever test of potential confounding. The researchers ask us to suppose that a top-performing teacher had been teaching grade 5 at the same school for several years. She is one of three grade 5 teachers at the school. Then the researchers ask us to suppose that the teacher leaves. Given that the school is unlikely to replace this exceptional teacher with an equally good performer, the grade 5 cohorts will average lower achievement in the years after she leaves because one-third of the students will now have a less

## DO VALUE-ADDED METHODS LEVEL THE PLAYING FIELD FOR TEACHERS?

effective teacher than did the students in the previous cohorts. In other words, if we follow sequential cohorts of students from a grade level in a school, we should hypothesize to see a drop in achievement when a teacher with a very high value-added estimate leaves the school (or a particular grade), and an increase in achievement when a teacher with low value added leaves. The magnitude of the change will depend on the value added of the teacher who transfers.

The study tested this hypothesis and found that these patterns held.<sup>12</sup> And the degree of change was very much in line with the predictions based on the teachers' value added. If confounding were strong, the value-added estimate for a really good teacher would be due to the students the teacher was assigned, not to the teacher herself. Consequently, the loss of a teacher with high value added would not affect average achievement, since removing that teacher would have no change on the cohorts, and since cohort-to-cohort changes in achievement would be smaller than that predicted by value added. The fact that the changes are consistent with value-added predictions suggests that confounding is limited.<sup>13</sup>

The second study<sup>14</sup> that found no evidence of confounding compared value-added estimates when schools followed standard practices for assigning classes to teachers with value-added estimates when classes were randomly assigned to teachers. It used 78 pairs of teachers; both taught the same grade in the same school, and both remained at that school for two consecutive years. Value added cannot be confounded in the year that students were randomly assigned to classes since assignments resulted from the luck of the draw. The study found that value-added estimates on randomly assigned classes were statistically equivalent to the value-added estimates from classes that were assigned using standard practices. This result suggests that the measure from standard practice must not be confounded. However, the study tests for confounding only among teachers within a school. It does not provide evidence about confounding by differences among students from different schools.<sup>15</sup>

The Measure of Effective Teaching (MET) Project repeated this experiment with 1,181 teachers from six large school systems.<sup>16</sup> Teachers taught math or English language arts to students in grades 4 to 8. This study also found no evidence of confounding. Teachers' value-added with randomly assigned classes matched their value-added for classes assigned using standard practice for the full sample of math and English language arts teachers and for the sample of just math teachers.<sup>17</sup> For the sample of only English language arts teachers, the value-added for randomly assigned classes did not align as well with value-added for classes assigned using standard practice as it did for the math teachers. However, the study results were imprecise for the English language arts teachers, so they provide no conclusive evidence of confounding.

### STUDIES THAT SUGGEST CONFOUNDING EXISTS

As noted above, one set of studies finds compelling evidence that value-added measures from certain very simple models typically will be confounded. Two other studies have taken different approaches, and both demonstrate the potential for confounding, although neither proves it exists.

Studies that suggest that overly simple models will typically confound value-added measures do so by following process. First, they estimate value added from simple models, such as those that control for only a few prior test scores and don't make complicated adjustments for measurement error. Next, they estimate value added using more complex models. Finally, they show that the value added from

## DO VALUE-ADDED METHODS LEVEL THE PLAYING FIELD FOR TEACHERS?

simpler models is more strongly related to the background characteristics of the teachers' classes than the value added from more complex models. The studies argue that the relationship between value added and student variables can't be due to teachers since it doesn't exist for value-added estimates made with the complex models. It must result from the failure of simple models to fully control for differences among the students taught by different teachers.<sup>18</sup>

Another study compared estimates of the variability of value added in schools in which class assignments appeared to be random to that of schools in which assignments were distinctly non-random, that is, the distributions of student background variables were too different across classrooms to have occurred by chance.<sup>19</sup> Value-added estimates varied more among teachers in schools that did not appear to assign students at random than in schools that did. This finding is consistent with what we would expect to see if the estimates confounded student background variables with teacher effectiveness. The schools with non-random differences among classes have an additional source of variability in their value-added estimates because they conflate student variables with real teacher effectiveness. The other schools do not have this other factor because the student background variables do not vary among classes. However, we might also observe this empirical finding if (1) there were no confounding and (2) schools that did not appear to use random assignment had staffs that were more varied in their effectiveness. The study cannot rule out this alternative.

As we know, value-added estimates are based on statistical models. These models will not confound value added with student background variables under specified assumptions about how these variables relate to student achievement. If the assumptions do *not* hold, then confounding can occur. A test for confounding can then be constructed as a test of the model's assumptions. Tests of model assumptions are complicated, but one test<sup>20</sup> has an intuitive key component: it tests whether students' future teacher assignments are associated with students' current achievement scores given the control variables that are specified in the model. Since future teachers cannot directly affect students' current achievement, a finding that future teacher assignments are associated with those outcomes ("the future predicts the past") suggests a violation of the model assumptions that could lead to confounding. When applied to data from North Carolina, the test found strong evidence that students' future teacher assignments predict the students' current scores.<sup>21</sup>

This test has been replicated with other data, and those studies also found strong evidence that future teacher assignments predict current scores. The test of model assumptions is not the same as a test for bias. Recent theoretical findings show that if students are tracked on the basis of their current test scores, the test of assumptions might fail, but value added might not be confounded. Failing the test of assumptions says that confounding is possible, but it doesn't guarantee that it exists.

## WHAT MORE NEEDS TO BE KNOWN ON THIS ISSUE?

Value-added models clearly go a long way toward leveling the playing field by controlling for many student variables that differ among classrooms. Whether value added fully levels the playing field is a question that can't be answered without more evidence on confounding and the conditions under which it is likely to occur. Given the extensive controls used in value-added models, it is possible that even if confounding occurs, for many teachers it could lead to errors smaller than those produced by other means of teacher evaluation.<sup>22</sup>

## DO VALUE-ADDED METHODS LEVEL THE PLAYING FIELD FOR TEACHERS?

Most of the studies of confounding have not been replicated by other researchers in other places. Consider, for example, the study that shows that cohort-to-cohort changes in achievement are predicted by the value added of teachers who leave a school or grade. That study excluded many students in the district because their data were incomplete. It is not clear whether the results would have differed if all the students had been used. One way to alleviate concerns about the data would be to replicate the study. If the results of this study were replicated, they could provide fairly strong evidence that confounding is limited.

The studies described in the previous sections provide evidence about general patterns of confounding. They do not assess the confounding for individual teachers. Results for some teachers may differ from the general patterns, so districts should look carefully at the conditions of teachers who receive consistently low or high scores: What types of students do these teachers have? And how might they differ from other classes?

More generally, states and districts could better judge the merits of value-added measures if they knew more about the conditions under which there was strong or weak confounding. Studies across states and districts with different policies and populations could help expose these conditions. For instance, researchers could conduct studies in districts that do and do not track students in secondary schools, and they could determine if confounding appears stronger or more likely when tracking occurs.

Similarly, states and districts need guidance about which value-added model best mitigates confounding. The studies that find no evidence of confounding include classroom-level variables and sometimes school-level variables. The other studies do not include these variables. The relationship between value-added estimates with background variables is very sensitive to the inclusion of such aggregates,<sup>23</sup> and the results on confounding may be, too. Controlling for aggregate variables might even cause confounding for some value-added approaches.<sup>24</sup> On the other hand, *not* controlling for aggregate variables might also result in confounding. It would be valuable to know how much the difference in the models contributes to the differences in the results. An answer could come from replicating the studies on confounding using different models.

The studies discussed here focus only on elementary- and middle-school students and teachers; they do not include high schools, where results might be different. Standardized tests in high schools are less directly related to the coursework of students, and end-of-course tests may be taken by very selective samples of students. Moreover, the previous test scores of high school students, which are the most important variable used in value-added models, may be only weakly related to standardized or end-of-course tests. This is because of the time lapse between testing<sup>25</sup>, lack of alignment between material covered by the tests<sup>26</sup>, or both. We need empirical studies on high schools before we can draw any conclusions about how level the field is for these teachers.

An issue related to confounding is whether a teacher's effectiveness depends on the students she is teaching and the environment in which she teaches, and whether she would be more or less effective in a different context.<sup>27</sup> But the above studies consider confounding by student characteristics only in given schools and classrooms. If teacher effectiveness depends on context, this is another source of error that calls for further study.

### WHAT CAN'T BE RESOLVED BY EMPIRICAL EVIDENCE ON THIS ISSUE?

The question of whether value-added measures are truly confounded probably can never be resolved completely with empirical evidence. Empirical studies can determine whether the measures appear to be confounded for teachers in the grade levels, schools, districts, and states that were studied, but they can't rule out different conclusions in other settings. Studies from middle schools might not apply to high schools, for instance, and some teachers might teach students who are so different from other students that value-added measures fail to account for their achievement levels.

It is also very challenging to differentiate between a teacher's contribution to student achievement and the influences of the school, the classroom, and a student's peers. The separate contributions can only be studied when teachers move, and even then, we must assume that the effects of these other influences and the effectiveness of the teacher remain constant across time. Value-added models often control for variables such as school-wide average scores on previous tests, and, as noted above, controlling for such averages could introduce errors into the value-added measure. Or it could in the future if schools change how teachers are assigned to classes once the value-added estimates become available.

Even if we conclude that value-added measures are not confounded, we will never know the true effectiveness of a teacher working in all different classroom situations.

### PRACTICAL IMPLICATION

#### How Does This Issue Impact District Decision Making?

Because the evidence on confounding by student background variables is mixed, those who make decisions about teachers should allow for its potential to occur. Confounding can lead to serious errors with heavy consequences for teachers and society: ineffective teachers may be deemed effective and vice-versa. So, districts need accurate measures of these errors, and they must consider them if they plan to use value-added measures for high-stakes decisions.

Another implication of confounding is that these errors—and the resulting conclusions about teachers—will be associated with different groups of students. The errors may serve to discredit value-added modeling, precisely because the models purportedly remove such associations. Teachers might even be discouraged from teaching certain groups of students if confounding results in their getting consistently low value-added scores.

It is important for policymakers to remember that certain conditions, such as how teachers are assigned to schools and classes, can change over time and that these changes can affect confounding. The above studies used data that were collected before states and district reported value-added scores to teachers or used it for evaluations. A decision to use these scores for evaluations might in itself change the schooling environment in ways that could lead to different results. For example, teachers with higher value-added scores might transfer to certain kinds of schools, thus creating a new association between teacher effectiveness and student background variables that does not yet exist, but which could in the future.



## DO VALUE-ADDED METHODS LEVEL THE PLAYING FIELD FOR TEACHERS?

To reduce the risk of confounding in value-added estimates, states and districts should avoid the simple value-added models described above. They should control for multiple previous test scores and account for measurement error in those tests. They should use models advocated by studies that compare alternative value-added approaches,<sup>28</sup> and if they work with a vendor, use one that has given the potential for confounding careful consideration.

If decision-makers suspect confounding, they might be wise to limit the comparisons they make with value-added estimates. For instance, they might want to compare teachers only to their peers in similar schools, or compare only teachers within the same school. Districts might also study the relationship between value-added measures and student background variables. A strong relationship between value-added measures and background variables could indicate confounding or disparity in the assignment of teachers. Districts might pay particular attention to teachers of students with uncommon characteristics. They might monitor these teachers' value-added scores for consistent highs or lows and check how they relate to other measures of teaching, such as classroom observations. The relationship between value-added estimates and other measures should be the same for these teachers as it is for others. Districts could also track, over time, the average achievement of grade-level cohorts within schools to determine if performance changes as predicted by the value added by teachers who transfer into or out of schools and grades.

Finally, districts should monitor student achievement, along with scores for teacher observations, to determine whether the use of value-added measures in evaluations is doing what is most important – improving teaching and learning.

## ENDNOTES

---

- <sup>1</sup> Errors from chance assignments are also a serious concern for value-added and can contribute to year-to-year fluctuations in scores for a teacher. However, in this brief we focus on confounding. For a discussion of year-to-year instability in value-added and its implications see:  
Loeb, Susanna, and Christopher Candelaria. Carnegie Knowledge Network, "How Stable are Value-Added Estimates across Years, Subjects, and Student Groups?" Last modified October 2012. URL = <http://carnegieknowledge.org/briefs/value-added/value-added-stability/>.
- <sup>2</sup> J.R. Lockwood and Daniel F. McCaffrey, "Reducing Bias in Observational Analyses of Education Data by Accounting for Test Measurement Error," (Working Paper, 2012).
- <sup>3</sup> This result pertains to value-added models that control for students' prior achievement using a method known as regression. Regression is very commonly used for value-added models. This result will also hold for *median growth percentiles* based on student growth percentiles.
- <sup>4</sup> Damian W. Betebenner, "A Primer on Student Growth Percentiles," Colorado Department of Education, 2008. <http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudentgrowthpercentiles.pdf>
- <sup>5</sup> Ibid. For more on the impact of different value-added modeling choices on teachers' rankings see:  
Goldhaber, Dan, and Roddy Theobald. Carnegie Knowledge Network, "Do Different Value-Added Models Tell Us the Same Things?" Last modified April 2013. URL = <http://carnegieknowledge.org/briefs/value-added/different-growth-models/>.
- <sup>6</sup> Raj Chetty, John F. Friedman, and Jonah E. Rockoff, "Long-Term Impact of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," (National Bureau of Economic Research, Working Paper #w17699, 2011). <http://www.nber.org/papers/w17699>; and Thomas Kane and Douglas Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," National Bureau of Economic Research Working Paper #w14607 (2008). <http://www.nber.org/papers/w14607>

- 
- <sup>7</sup> Jun Ishii and Steven G. Rivkin, "Impediments to the Estimation of Teacher Value-Added," *Education Finance and Policy* 4 (2009): 520-36, Jesse Rothstein, "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables," *Education Finance and Policy* 4 (2009): 537-71. Jesse Rothstein, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics* 125 (2010): 175-214, Cory Koedel and Julian Betts, "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique," (University of Missouri, Department of Economics, Working Paper 09-02, 2009), and Douglas Harris, Tim Sass, and Anastasia Semykina, "Value-Added Models and the Measurement of Teacher Productivity" (Working Paper, 2010).
- <sup>8</sup> We can test whether or not value-added models do indeed control for all the troubling factors: we collect auxiliary data not available for states and districts for value-added modeling such as detailed family income data. We then check if value-added is associated with the auxiliary data. If it is, then we would have evidence that factors typically excluded from value-added appear to be confounded with it. If value-added is not related to the auxiliary data, then we would have evidence that value-added apparently is not confounded by those factors.
- <sup>9</sup> Raj Chetty, John F. Friedman, and Jonah E. Rockoff, "Long-Term Impact of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," (National Bureau of Economic Research Working Paper #w17699, 2011).
- <sup>10</sup> Other measures include parental marriage status, home ownership, and mother's age at the child's birth.
- <sup>11</sup> The study could not test for missing variables unrelated to socio-economic status such as whether or not students were neglected or abused, for instance. Social factors such as whether or not students are abused can vary significantly even among schools serving families with similar income levels. See: Anthony S. Bryk, Penny Bender Sebring, Elaine Allensworth, Stuart Luppescu, and John Q. Easton, *Organizing Schools for Improvement: Lessons from Chicago*, (Chicago: University of Chicago Press, 2010).
- <sup>12</sup> Ibid
- <sup>13</sup> Although this is a novel result, it is not the final word on confounding. The study removed many students due to incomplete data and the consequences of deleting these students are not investigated. Also, details on the study of cohort changes are not presented in the paper so they cannot be fully vetted. Finally, it is possible that just changing the groupings of students after teachers leave results in changes to average achievement.
- <sup>14</sup> Thomas Kane and Douglas Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," (National Bureau of Economic Research, Working Paper #w14607, 2008).
- <sup>15</sup> Technically, the study developed a novel way to test for statistical equivalence which involved testing that a statistic equals one. The test statistics for the study equaled 0.87 and was not statistically different from one. A finding that a statistic is not "significantly different from one" is not the same as demonstrating with a high degree of certainty that it was equal to one. Rather, this result means the study did not find strong evidence suggesting the coefficient is not equal to one. In other words, the test does not prove there is no confounding; it only fails to find evidence of it.
- <sup>16</sup> Thomas Kane, Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger, *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. The Bill and Melinda Gates Foundation, 2013.  
URL=[http://www.metproject.org/downloads/MET\\_Validating\\_Using\\_Random\\_Assignment\\_Research\\_Paper.pdf](http://www.metproject.org/downloads/MET_Validating_Using_Random_Assignment_Research_Paper.pdf).
- <sup>17</sup> As with the Kane and Staiger study, the test for confounding involved testing that a statistic equaled one. For the combined sample the statistic equaled 0.95 a 95% confidence interval of 0.71 to 1.20. For math teacher the statistic equaled 1.04 with a 95% confidence interval of 0.76 to 1.31 and for English language arts teachers it was 0.70 with a 95% confidence interval from 0.27 to 1.12.
- <sup>18</sup> J.R. Lockwood and Daniel F. McCaffrey, "Reducing Bias in Observational Analyses of Education Data by Accounting for Test Measurement Error," (Working Paper 2012). Daniel F. McCaffrey, "What Variables Belong in Your Value-Added Model?" (Paper presented at the annual meeting of the National Council on

Measurement in Education, New Orleans, Louisiana, April 9 -11 2011). Mark Ehlert, Cory Koedel, Eric Parsons, and Michael J. Podgursky, "The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence from School- and Teacher-Level Models in Missouri," (Working Paper 2012).

<sup>19</sup> Jun Ishii and Steven G. Rivkin, "Impediments to the Estimation of Teacher Value-Added" *Education Finance and Policy* 4 (2009): 520-36.

<sup>20</sup> Jesse Rothstein, "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables," *Education Finance and Policy* 4 (2009): 537-71. Jesse Rothstein, "Teacher Quality in *The Quarterly Journal of Economics*, 125, no.1, (2010): 175-214. doi: 10.1162/qjec.2010.125.1.175

<sup>21</sup> Ibid.

<sup>22</sup> Harris, Douglas N. Carnegie Knowledge Network, "How Do Value-Added Indicators Compare to Other Measures of Teacher Effectiveness?" Last modified May 2013. URL = <http://carnegieknowledge.org/briefs/value-added/value-added-other-measures/>.

<sup>23</sup> Daniel F. McCaffrey, "What Variables Belong in Your Value-Added Model?" (Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, Louisiana, April 9 -11 2011).

<sup>24</sup> Controlling for aggregates might cause confounding if the true effectiveness of teachers is associated with these aggregate control variables. For instance, if more effective teachers teach classes higher average prior achievement.

<sup>25</sup> In many states standardized testing occurs annually for students through grade but then not again until students are in grade 10 or 11.

<sup>26</sup> Scores on general achievement tests like the state grade 8 English language arts and mathematics tests may be weak predictors of student achievement in advanced mathematics, science, social studies, or high school English.

<sup>27</sup> For more details on teachers teaching in different contexts see:

Loeb, Susanna, and Christopher Candelaria. Carnegie Knowledge Network, "How Stable are Value-Added Estimates across Years, Subjects, and Student Groups?" Last modified October 2012. URL = <http://carnegieknowledge.org/briefs/value-added/value-added-stability/>.

<sup>28</sup> For example see Cassandra M. Guarino, Mark D. Reckase, and Jeffrey M. Wooldridge, "Can Value-Added Measures of Teacher Performance Be Trusted?" Michigan State University, 2012. <http://vam.educ.msu.edu/publications/> or Goldhaber, Dan, and Roddy Theobald. Carnegie Knowledge Network, "Multiple Value-Added Models: Do Different Models Tell Us the Same Thing?" Last modified April 2013. URL = <http://carnegieknowledge.org/briefs/value-added/different-growth-models/>.

### AUTHORS



**Daniel F. McCaffrey** is a Principal Research Scientist at Educational Testing Service. Previously, he was a Senior Statistician and PNC Chair in Policy Analysis at the RAND Corporation. He is a fellow of the American Statistical Association and is nationally recognized for his work on value-added modeling for estimating teacher performance. McCaffrey oversaw RAND's efforts as part of the Gates Foundation's Measures of Effective Teaching study to develop and validate sophisticated metrics to assess and improve teacher performance. He is currently working on additional studies comparing value-added measures to other measures of teaching, including classroom observations. He recently completed work on a four year project funded by the Institute of Education Sciences (IES) that developed alternative value-added models of teachers' effectiveness. McCaffrey is also the principal investigator of a National Institute on Drug Abuse-funded study, and recently led RAND's efforts as a major partner in the National Center on Performance Incentives, which conducted random control experiments to test the effects of using value-added to reward teachers with bonuses. He led an evaluation of the Pennsylvania Value-Added Assessment Pilot Program (PVAAS) and was the lead statistician on other randomized field trials of school-based interventions; including evaluations of the Cognitive Tutor geometry curriculum, the Project ALERT Plus middle and high school drug prevention program, and the teen dating violence prevention curriculum, Break the Cycle. McCaffrey received his Ph.D. in statistics from North Carolina State University.

### 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

McCaffrey, Daniel. Carnegie Knowledge Network, "Do Value-Added Methods Level the Playing Field for Teachers? " Last modified June 2013.

URL = <<http://carnegieknowledgenetwork.org/briefs/value-added/level-playing-field/>>



## **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/>.

**Knowledge Brief 2**  
**October 2012, Revised June 2013**  
**[carnegieknowledgenetwork.org](http://carnegieknowledgenetwork.org)**

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