Carnegie Foundation for the Advancement of Teaching

Does Value-Added Work Better in Elementary Than in Secondary Grades?

DOUGLAS N. HARRIS
TULANE UNIVERSITY

ANDREW ANDERSON
UNIVERSITY OF WISCONSIN-MADISON

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

ATIL

ADVANCING TEACHING - IMPROVING LEARNING

SUMMARY

- The vast majority of research on value-added measures focuses on elementary schools; value-added measures for middle and high school teachers pose particular challenges.
- Middle and high schools often "track" students in ways that affect the validity of value-added.
- Student tracking in middle and high schools calls into question the validity of methods typically used to create value-added measures.
- The validity of secondary-level value-added measures can be improved by directly accounting for tracks and specific courses, although this may not completely solve the problem.
- Middle and high school teachers have more students, and this factor increases reliability, but it is offset by other factors that reduce reliability at those grade levels.
- End-of-course exams, which are becoming more common in high school, have both advantages and disadvantages for estimating value-added.

INTRODUCTION

There is a growing body of research on the validity and reliability of value-added measures, but most of this research has focused on elementary grades. This is because, in some respects, elementary grades represent the "best-case" scenario for using value-added. Value-added measures require annual testing and, in most states, students are tested every year in elementary and middle school (grades 3-8), but in only one year in high school. Also, a large share of elementary students spend almost all their instructional time with one teacher, so it is easier to attribute learning in math and reading to that teacher.[1]

Driven by several federal initiatives such as Race to the Top, Teacher Incentive Fund, and ESEA waivers, however, many states have incorporated value-added measures into the evaluations not only of elementary teachers but of middle and high school teachers as well. Almost all states have committed to one of the two Common Core assessments that will test annually in high school, and there is little doubt that value-added will be expanded to the grades in which the new assessments are introduced.[2] In order to assess value-added and the validity and reliability of value-added measures, it is important to consider the significant differences across grades in the ways teachers' work and students' time are organized.

As we describe below, the evidence shows that there are differences in the validity of value-added measures across grades for two primary reasons. First, middle and high schools "track" students; that is, students are assigned to courses based on prior academic performance or other student characteristics. Tracking not only changes our ability to account for differences in the students who teachers educate, but also the degree to which the curriculum aligns with the tests. Second, the structure of schooling and testing vary considerably by grade level in ways that affect reliability in sometimes unexpected ways. The problems are partly correctable, but, as we show, more research is necessary to understand how problematic existing measures are and how they might be improved.

WHAT DO WE KNOW ABOUT HOW TEACHER VALUE-ADDED MEASURES WORK IN DIFFERENT GRADES AND SUBJECTS?

We begin by discussing differences in *validity* across grades and follow with somewhat briefer discussions of *reliability* across grades. Validity refers to the degree to which something measures what it claims to measure, at least on average. Reliability refers to the degree to which the measure is consistent when repeated. A measure could be valid on average, but inconsistent when repeated, meaning it isn't very reliable. Conversely, a measure could be highly reliable but invalid—that is, it could consistently provide the same invalid information.

Validity of Value-Added Measures Across Grade Levels

Students and teachers are assigned to classrooms differently in elementary schools than in middle and high schools. In elementary schools, it is common for principals to create similar classrooms (e.g., with similar numbers of low-performing and special needs students).[3] Other elementary principals identify student needs and try to match them to teachers who have the skills to meet those needs. Principals may also take into account parental requests, so that students with more academically demanding parents get assigned to teachers with the best reputations. Either of these last two forms of assignment—those based on student needs and those based on parental requests—has the potential to reduce the validity of value-added measures.[4] Sometimes called "selection bias," the problem is that student needs and parental resources are never directly accounted for in value-added measures, even though they might affect student learning and therefore reduce validity of teacher value-added estimates.

Based on a series of experiments,[5] simulation studies,[6] and statistical tests,[7] elementary school value-added models do seem to address the selection bias problem well, on average. *This last caveat is important*. It is extremely difficult to provide strong evidence of validity for each teacher's value-added. Instead, prior studies are really examining whether selection bias averages out for whole groups of teachers.[8]

Students in *middle and high schools*, on the other hand, are not assigned to or "selected" for classes in the same way they usually are in elementary schools. Rather, students with low test scores and grades and certain other characteristics are generally tracked into remedial courses, and those with stronger academic backgrounds are tracked into advanced courses. Minority and low-income students are also more likely to end up in lower tracks. These decisions might not be driven by strict rules or requirements, but they reflect strong patterns. In our analyses of Florida data, 37 percent of the variation in students' middle school course tracks can be explained by a combination of their prior test scores, race/ethnicity, and family income.[9]

Tracking creates two potential problems for value-added. First, the academic content of the courses differs. This means that the material covered in each course aligns to the test in different ways. Tests are designed to align with state proficiency standards[10], which in many states require a fairly low level of academic skill.[11] For this reason, we would expect the test to align better with lower or middle tracks, implying that teachers in these tracks have an easier time showing achievement gains—and therefore higher value-added. This prediction is reinforced by evidence of "ceiling effects" in standardized tests; students in the upper tracks, as described above, are likely to have higher scores and to hit the ceiling with little growth.[12] The direction and magnitude of these influences depends, of course, on the test and no doubt varies

by state. Those states with low proficiency bars are probably more likely to have tests that align better with the remedial courses.

This disadvantage to teaching in the upper track, however, is apparently offset by an apparently larger advantage. That is, upper track students seem to have unobserved traits that make them likely to achieve *larger* achievement gains. This is what we would predict based on which parents tend to push hardest to get their children into upper track courses. Parents who press for more challenging academic courses probably also press their children to work harder, do their homework, and so on—generating higher achievement. We cannot observe these parental activities, so they could get falsely attributed to teachers in upper tracks.

The net effect is unclear. The curriculum-test misalignment places upper-track teachers at a *disadvantage* because of the misalignment between the test and course content, but this might be offset by the *advantage* of having students who are likely to make achievement gains for reasons having nothing to do with the teacher. Below, we report results of data analyses that shed more light on the issues created for value-added measures by tracking.

Analyses of Florida Secondary Schools. We estimated teacher value-added ignoring students' tracks and courses, as is typically done, and then we re-estimated with track/course effects.[13] In middle schools, our estimates suggest that for a teacher with all lower track courses, ignoring tracks would reduce measured value-added from the 50th to the 30th percentile. Only about 25-50 percent of teachers remain in the same performance quartile when we add information about the tracks.

One might wonder whether these effects exist because more effective teachers end up in upper-track courses. We addressed this possibility by analyzing teachers who taught both lower- *and* upper-track courses and comparing value-added in each course type for the same teacher.[14] Teachers had higher value-added when they taught the upper-track classes, compared with the same teachers teaching lower tracks. These results could actually understate the role of tracks because the information available about tracks might not always be accurate.[15] For this report, we extended the analysis to Florida high schools where a similar number, 33 to 45 percent, would be in the wrong group without tracks.

Analyses from North Carolina and End-of-Course Exams. If tracking is a problem in estimating value-added, we would expect the variation in high school teacher value-added to drop when we account for tracks; that is, when we ignore tracks and courses, some teachers end up with value-added that is too high because they teach many upper-track courses, and vice versa in the lower track. So accounting for tracks pulls these teachers back to the middle and reduces variation.

A recent report using data from North Carolina confirms this.[16] The variation in teacher value-added in high school is 33 percent lower when adding track coefficients. That is, some teachers have extremely low value-added simply because they teach more lower-track courses, and other teachers have high value-added because they teach upper-track courses. This does not prove that tracking is the problem, but the evidence is consistent with that interpretation.

The study goes further and considers how well current value-added predicts future value-added across grade levels. Several researchers have argued that this predictive validity of value-added

is an important sign of the measure's validity for making long-term employment decisions. We would hope, for example, that the measures used to make teacher tenure decisions are good predictors of how teachers will perform in the years after they receive tenure. The North Carolina study finds that even the course-adjusted value-added measure is a worse predictor of future value-added in high schools than in elementary schools, even after accounting for tracking.[17] This suggests that adjusting value-added measures in this way does not eliminate the concern that tracking reduces validity and/or that there might be other problems in estimating high school value-added.[18]

The differences in state testing regimes are also noteworthy. In theory, high school value-added measures should have higher validity in North Carolina because that state uses end-of-course exams, which should be better aligned to the curriculum than the single generic subject test in Florida. However, recall that the advantage for higher-track teachers from omitting tracks may offset the disadvantage to those same teachers from test misalignment (e.g., test ceilings). Paradoxically, this means better test alignment could actually make the validity or selection bias problem worse because one no longer offsets the other. This would give the upper-track teachers an even greater advantage. While it is in some ways helpful that the two problems cancel out, it places us in the awkward position of having to rely on one mistake to fix the other. As an analogy, it is like a golfer who accidentally aims too far to the left but still hits the ball in the fairway because of a slice to the right—the problems cancel out. In this case, fixing only the slice and not the aim would put the ball to the right of the fairway, making matters worse.

The use of end-of-course exams also raises the issue about how well prior achievement scores account for students' relevant prior achievement. The purpose of accounting for prior scores is that they can tell us where students started at the beginning of the school year, but the content of prior courses is so different in high school that it's unclear how informative the prior score really is. For example, few students have learned anything about physics before they take a physics course. However, accounting for prior math, science, and other scores is still important because those scores adjust for general cognitive and study skills that also influence subsequent scores.[19]

The ability of prior courses to account for sorting across grade levels is therefore unclear, but there are good reasons to think that having good alignment between this year's test and this year's content is more important than having good alignment between this year's test and last year's test. As further evidence of this, we estimated value-added to math scores in middle schools controlling only for prior *reading* scores—prior math scores were ignored. We then compared these new value-added estimates with the more typical ones where prior math is accounted for. The correlation between the two is high at 0.84.[20]

Other Evidence and Summary about Validity Issues. The well-known Measures of Effective Teaching (MET) project funded by the Gates Foundation reports results from experiments that also address the validity of value-added at the middle and high school levels. They randomly assigned teachers to classrooms in middle school as well as 9th grade. However, there was apparently no data about the tracks teachers taught or whether random assignment occurred only within tracks. Given the directions provided to principals, it seems likely that most assignments were within a track, but we cannot know for sure because tracking data was generally not available in MET, so this study is not informative about the role of tracks in value-added estimation.[21]

Overall, the evidence from the above studies suggests that ignoring tracks will reduce validity substantially in middle and high school, and even accounting for tracks may not solve the problem. This also reinforces the general problem of comparing teacher performance in different instructional contexts.[22]

Reliability of Value-Added Measures Across Grade Levels

There may be trade-offs between validity and reliability in evaluating value-added measures.[23] Below, I consider the reliability of value-added by grade and then illustrate those trade-offs.

There are many sources of random error in value-added estimates: standardized tests have measurement error, some students are sick at test-taking time, and the students assigned to teachers in any given year vary in essentially random ways. This helps to explain why teacher value-added measures are somewhat unstable over time. It also explains why researchers and value-added vendors typically report "confidence intervals" for value-added measures that help quantify the role of random error and the uncertainty this creates about teachers' "true" value-added.[24]

One of the key factors affecting confidence intervals is the sample size—the larger the number of students assigned to each teacher, the smaller the confidence interval. The fact that elementary students are assigned to only one teacher means that we can probably attribute that student's learning to that teacher, but the trade-off is that these elementary teachers have fewer students assigned to them, and this will tend to reduce reliability.

The larger number of students per teacher at the secondary level does not necessarily mean, however, that reliability is better. This is because reliability depends on error variance *relative* to the variance in the value-added estimates.[25] To take a sports analogy, suppose that we had very precise estimates of the performance of ten baseball players, but that every player was almost equally effective and therefore had almost identical batting averages. In this situation, the variance in true performance is very small, so even very precise estimates of batting averages (after lots of games) will make it hard to distinguish the best from the worst players— the estimates will be unreliable even after each player has hundreds of at-bats. Conversely, if half the players had high batting averages and the other half had no hits at all, then we could reliably identify the low-performers after a week's worth of games. The confidence intervals would be wide in that case, but it wouldn't matter because the differences in true performance are so large.

In this case, having more students reduces random error among middle school teachers, but, as the baseball analogy suggests, this does not increase reliability. Estimates by Daniel McCaffrey using MET project data show that there is almost no relationship between grade level and reliability—reliability may actually be worse in higher grades.

Why might that be? Is there greater variance in teacher effectiveness at the elementary level? Is random error lower at the elementary level? Or both? One plausible explanation is that middle school teachers each teach a wider range of students each year than do elementary teachers. This is plausible because our calculations in Florida suggest that most teachers work in multiple tracks. If value-added estimates do not fully account for unobservable differences in students, then we would expect to see this pattern—the variance in teacher value-added is greater at the elementary level perhaps because of biased estimates.[26] Differences in random

error could also explain lower reliability in middle schools if the reliability of the tests is lower, which would offset the advantage of having more students.

The calculations by McCaffrey provide some support for both interpretations. Compared with the elementary schools in his sample, the variance in teacher value-added is lower in middle schools and random error is higher. So, the advantage of having more students per teacher is offset by other factors that reduce reliability in middle school.

WHAT MORE NEEDS TO BE KNOWN ON THIS ISSUE?

The above evidence strongly suggests that accounting for course tracks is important to obtaining valid value-added estimates in middle and high school, but we do not yet know how well this solves the problem. We estimate value-added by accounting for prior achievement, but a key implication of our tracking argument is that prior achievement is affected by prior tracks. This creates a complex role for tracks that might not be easily captured by simply adding track variables to the value-added model.[27] We therefore cannot presume that accounting for the tracks is sufficient, and the North Carolina study reinforces this conclusion.[28]

It would also be useful to know how this issue affects teachers in schools that use non-standard courses. We focused on algebra and geometry and excluded courses like "Liberal Arts Mathematics" and "Applied Mathematics" that also showed up in the data. These courses are likely to align even less well with the tested content because, by definition, they are courses that are outside the norm. The teachers of these courses could be at a significant disadvantage in their performance ratings.

In addition, while we have learned a great deal about elementary teacher value-added from experiments and simulation evidence, we need to apply those same methods to address the particular threats to validity in middle and high schools. The MET project provides some experimental evidence in middle school, although simulation and other tests have been limited to elementary grades.

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

While the evidence described here provides a sense of the empirical problems that arise across grades and subjects, there is a larger question about how well the tests capture what we want students to learn and be able to do—and how this varies across grades. For example, creativity might be a skill that could be developed more easily in early grades, but creativity is hard to measure. So, in that case, the validity problems in early grades would be even worse than in later grades. The statistical issues are therefore intertwined with the philosophical ones about what we want students to learn.

HOW, AND UNDER WHAT CIRCUMSTANCES, DOES THIS ISSUE IMPACT THE DECISIONS AND ACTIONS THAT DISTRICTS MAKE ON TEACHER EVALUATIONS?

This evidence also informs the use of value-added measures in (potentially) high-stakes decisions. For the sake of simplicity and perceived fairness, it is desirable to have a common standard that applies to all grades—or really to all teachers. However, if the validity and reliability vary, not to mention the ways in which test scores align with the desired goals, then treating teachers *equitably* may require using value-added *unequally* across grades. As I have written elsewhere, the stakes attached to any measure should be inversely proportional to the measure's validity and reliability.[29] It appears that we may not be able to follow that rule and simultaneously use value-added the same way for all teachers, especially across elementary and secondary grades.

Given that the properties of value-added measures differ across grades and subjects, policymakers should consider using different methods for calculating and using value-added in different grades and subjects. In particular, in middle and high school, it is essential to account for the tracks and courses that teachers are assigned to when calculating value-added.

Since value-added seems to work differently across grades, this raises the question: How do we handle teachers who teach multiple grades? Fundamentally, the issues raised here do not change the answer to this question. Comparisons across grades have always been complicated by the fact that the tests differ across grades and the various approaches to combining them involve some sort of weighted average, or composite, that takes into account differences in the test scale across grades.[30] That basic solution is also reasonable for handling the additional complication of tracking. The key is to first get the estimation right at each grade level, perhaps by accounting for tracks. That is, we have to get the estimates right for each track and grade level before creating composite value-added measures for each teacher.[31]

It might also be tempting to reduce tracking or assign each teacher an equal mix of low- and high-track courses to easier accommodate value-added measures, but this is the proverbial "tail wagging the dog" problem. Changes in school organization and instruction should be made with caution and attention to effective instructional practice—not so that we can have better value-added measures.[32]

The implications of tracking are missed in the vast majority of value-added estimates now being used. This means that, even setting aside other issues with the measures, current standard value-added measures for teachers who concentrate their work in particular tracks in middle and high schools will suffer from validity concerns. As with many of the problems with value-added, this one can be addressed with better data collection efforts and careful attention to how the measures are created. Accounting for tracks would almost certainly improve the measures, but future research will be required to determine how well this solution works in practice.

ENDNOTES

[1] A third factor is that a somewhat narrower range of academic content is covered in elementary schools, so there may be better alignment between the curriculum and the test.

[2] I am referring to the new assessments from the Smarter Balanced Assessment Consortium (http://www.smarterbalanced.org/smarter-balanced-assessments/) and
and the Partnership for Assessment of Readiness for College and Careers (PARCC) (http://www.parcconline.org/about-parcc). All but three states have committed to adopting one of these two tests. Both systems plan to test annually through high school.
See: Educational Testing Service, *Coming Together to Raise Achievement: New Assessments for the Common Core State Standards*, 2012. Retrieved March 13, 2013 from:
http://www.k12center.org/rsc/pdf/Assessments_for_the_Common_Core_Standards.pdf

[3] There are two types of evidence on this. In a study that is now somewhat dated, David Monk (1987) asked principals in a sample of schools about how they create classrooms. Alternatively, the assignment process can be studied indirectly using statistical tests of measureable student characteristics. This evidence is discussed in the next paragraph.
See: David H. Monk, "Secondary school size and curriculum comprehensiveness," *Economics of Education Review* 6(2) (1987): 137-150.
http://www.sciencedirect.com/science/article/pii/0272775787900471

[4] Yet another possibility is that principals alternate students so that those with a less effective teacher in one year might be assigned to a more effective teacher the next year.

[5] Two studies have randomly assigned students to classrooms to test the validity of value-added. The MET recent study and a similar earlier experiment (Kane & Staiger, 2008) suggest that accounting for prior achievement is sufficient to establish internal validity. This is a very valuable study, though it does have several limitations. First, the study can only account for selection bias within schools. Comparisons across school contexts are considered more problematic (Reardon & Raudenbush, 2009). Second, the results apply only to a fairly narrow sample of schools that complied with the rules of the experiments. The same concerns arise with the earlier version of this experiment in Los Angeles.
See: Thomas J. Kane, and Douglas O. Staiger, "Estimating teacher impacts on student achievement: An experimental evaluation," (National Bureau of Economic Research, No. w14607, 2008). http://www.dartmouth.edu/~dstaiger/wp.html
Sean F. Reardon and Stephen W. Raudenbush, "Assumptions of value-added models for estimating school effects," *Education Finance and Policy* 4 (4) (2009): 492-519.
http://cepa.stanford.edu/sean-reardon/publications

[6] Cassandra M. Guarino, Mark D. Reckase, and Jeffrey M. Wooldridge, "Can value-added measures of teacher performance be trusted?" (Paper presented at the annual meeting of the Association for Education Finance and Policy, Seattle, WA, March 24-26, 2011).
http://education.msu.edu/epc/publications/

[7] While Rothstein's (2010) initial tests suggested that students are not approximately randomly assigned, subsequent studies have suggested that they generally are (Koedel & Betts, 2009; Chetty, Friedman, & Rockoff, 2012; Chaplin & Goldhaber, 2012).
See: Jesse Rothstein, "Teacher quality in educational production: Tracking, decay, and student achievement," *The Quarterly Journal of Economics* 125(1) (2010): 175-214.
Dan Goldhaber and Duncan Chaplin, "Assessing the Rothstein Falsification Test: Does It Really Show Teacher Value-Added Models Are Biased," (CEDR Working Paper 2011-5, 2011).
http://www.cedr.us/publications.html

Raj Chetty, John N. Friedman, and Jonah E. Rockoff, "The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood," (National Bureau of Economic Research, No. w17699., 2011). http://www.nber.org/papers/w17699

Cory Koedel and Julian R. Betts, "Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique," *Education Finance and Policy* 6(1) (2011): 18-42.

[8] This is a fairly complex issue, and there is not yet agreement among researchers about how well this type of experiment tests for bias. However, from a technical standpoint, note that researchers are typically trying to determine whether an estimate of a single parameter is valid, but in this case we are trying to test whether thousands of them (teacher value-added estimates) are valid.

[9] This is actually an under-estimate because the number of tracks is limited and there is considerable variation in scores within tracks. Also, I do not have access to course grades and this, too, influences tracks. It is difficult to make the same calculations in elementary schools because the tracks (if any) are not explicit.

[10] One of the nation's leading test publishers, Pearson, describes this alignment between the proficiency standards and the test as one of the "Fundamentals of Standardized Testing."
See: Sasha Zucker, "Fundamentals of standardized testing," (Pearson, Inc., 2003). http://www.pearsonassessments.com/hai/SearchResults.aspx?cx=011382766870466991018%3Ageowfzwxvog&cof=FORID%3A11&ie=UTF-8&q=sasha%20&as_sitesearch=www.PearsonAssessments.com

[11] Peterson & Hess (2008) show that state proficiency bars are set low relative to the National Assessment of Educational Progress (NAEP) though this varies considerably by state.
See: Paul E. Peterson and Frederick M. Hess, "Few states set world-class standards." *Education Next* 8(3) (2008): 70-73.

[12] Cory Koedel and Julian Betts, "Value added to what? How a ceiling in the testing instrument influences value-added estimation," *Education Finance and Policy* 5(1) (2010): 54-81.

[13] Douglas N. Harris and Andrew Anderson "The Difficulty of Worker Monitoring in the Service Sector: A Formal Sorting Model and Empirical Evidence of the Value-Added of Middle School Students and Teachers," (Paper presented at the Association for Public Policy and Management, 2012).

[14] The fact that teachers in upper-track courses are at an overall advantage suggests that the student selection bias is larger than the influence of curriculum-test misalignment. We confirmed this by taking additional steps in the value-added estimation to account for student sorting and therefore isolate the curriculum-test misalignment. When we did this, the advantage of teaching upper-track courses turned into a disadvantage, just as we predicted.

[15] This is an example of the standard "attenuation bias" problem. Whenever an independent variable in a regression is measured with error, the role of that variable in explaining the dependent variable (in this case, student test scores) is under-estimated.

[16] C. Kirabo Jackson. Teacher Quality at the High-School Level: The Importance of Accounting for Tracks. Working Paper (2012).

[17] The Jackson (2012) study also focuses on within-teacher estimation to avoid conflating the results with teacher sorting.

[18] One additional potential problem is that courses are often split up during the school year, so the same student might have two different math teachers. This not only reduces reliability but calls into question validity since the pairings of teachers are unlikely to be even approximately random.

[19] This may also be true in elementary school. Even though we think of the content in one year building on the prior year, the learning process is not really so linear.

[20] This is a Spearman rank order correlation.

[21] We thank Dan McCaffrey for a useful conversation about the potential role of tracking at the secondary level in the MET study.

[22] Raudenbush and Reardon, 2009, ibid.

[23] See: Douglas N. Harris, *Value-Added Measures in Education: What Every Educator Needs to Know.* (Cambridge, MA, Harvard Education Press, 2011).

[24] See Douglas N. Harris, 2011, ibid. More formally, the confidence intervals that some value-added reports are based on the notion that the students assigned to teachers are a sample from the population of students who could have potentially been in those classes. Measurement error in the tests themselves though this is often not accounted for in the confidence intervals.

[25] Specifically, reliability is the ratio of the standard deviation in the estimates to the standard error of the estimate, which is the basis for the confidence interval.

[26] It is also possible that the true variance in teacher performance is greater at the elementary school level, but there is no reason to expect this and we find it unlikely.

[27] Harris & Anderson, 2012, ibid.

[28] It is also worth noting, however, that some aspects of the curriculum-test alignment is partly within the control of teachers and therefore perhaps something that should be attributed to teachers. Control over the curriculum varies across schools and districts, however.

[29] Douglas N. Harris, 2011, ibid.

[30] One approach is to estimate the model separately by grade level and then create a weighted average based on the number of teachers taught in different grades. Another similar approach is to estimate teacher value-added across grades but in a single model that takes grade effects into account.

[31] Whatever approach is used, it is important to identify the track effects from variation in performance by teachers who teach in multiple tracks. This would not necessarily arise in typical value-added measures. To address this, value-added estimation could be carried out in multiple stages where the effects of tracks would be identified first using only those teachers who teach in multiple tracks then that information would be used in a second stage to estimate value-added for other teachers.

[32] Douglas N. Harris, 2011, ibid.

AUTHORS

Douglas N. Harris is Associate Professor of Economics and University Endowed Chair in Public Education at Tulane University. His research explores how students' educational outcomes are influenced by school choice, standards, teacher evaluation, test-based accountability, college financial aid, and college access programs. A former school board member, his research marries theory and rigorous research with the practical realities of schooling with publications ranging from the general interest journal *Science* to the *Journal of Public Economics*, *Journal of Policy Analysis and Management*, and others. His recent book, *Value-Added Measures in Education* (Harvard Education Press, 2011) was nominated for the national Grawemeyer Award. *Washington Monthly* magazine has used his research on college performance measures in its college ratings and David Brooks has cited related work in his *New York Times* column. He has advised eight state departments of education, elected officials at all levels of government, and groups such as the National Academy of Sciences, National Council of State Legislatures, National Governors Association, and National School Boards Association. His work is frequently cited in the national media, including CNN, *Education Week*, *The New York Times*, and *The Washington Post*.

Andrew Anderson is completing a Ph.D. program in economics at the University of Wisconsin - Madison. His work investigates heterogeneity across postsecondary schooling options. He is interested in how vocational certificate programs differ from degree granting programs, how two-year colleges differ from four-year colleges, and how career-oriented fields of study differ from liberal arts education. His goal is to understand how these educational pathways impact students as well as the consequences for lifetime income, occupational stability, labor market participation and welfare. Andrew finds these topics interesting in themselves as well as due to their policy implications.

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**.

CITATION

Harris, Douglas, and Andrew Anderson. Carnegie Knowledge Network, "Does Value-Added Work Better in Elementary Than in Secondary Grades?" Last modified May 2013. URL = <http://carnegieknowledgenetwork.org/briefs/value-added/grades/>

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

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.

Knowledge Brief 7
May 2013
carnegieknowledgenetwork.org