

# Should Value-Added Measures Be Used for Performance Pay?

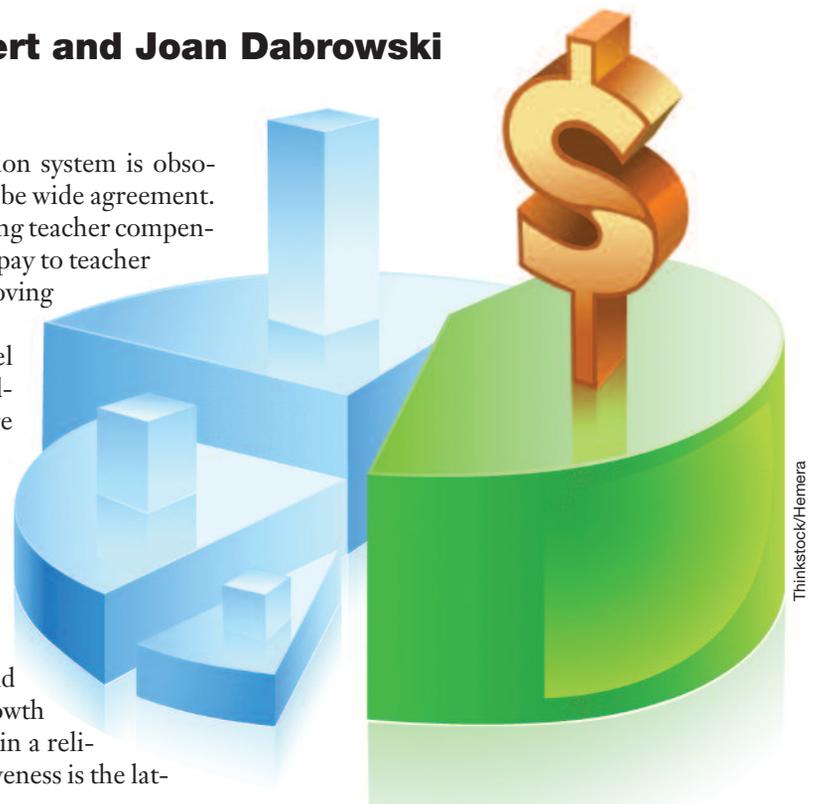
Value-added measures are becoming a critical piece of compensation reform, but buyers should be wary that any single number can express the effect teachers have on students.

**By Jonathan M. Eckert and Joan Dabrowski**

The current teacher compensation system is obsolete. On that point, there appears to be wide agreement. Virtually every new idea for improving teacher compensation focuses on how to tie teacher pay to teacher effectiveness as measured by improving student learning.

The so-called value-added model is the idea *du jour*. Value-added models vary in complexity, but they're generally based on an intuitively acceptable premise: By observing at least two data points at different times on vertically aligned tests, we can determine a student's growth over time. This, in turn, can be attributed to the effect of the teacher. The concept of pre- and posttesting students to determine growth is not new. Attributing that growth in a reliable and valid way to teacher effectiveness is the latest potential use of these data.

Evaluating teachers based on student performance makes sense. However, there's little evidence for the effect that this will have. Linking student performance with teacher pay before educators and data systems are ready is comparable to distributing an experimental drug before it's been adequately tested (Farmer 2009). But value-added measures should be included along with other measures of teacher effectiveness in an effort to reform the current approach to teacher compensation. Policy makers, administrators, and teach-



Thinkstock/Hemera

**JONATHAN M. ECKERT** is an assistant professor at Wheaton College, Wheaton, Ill. After teaching for 12 years in Illinois and Tennessee, he was a 2008-09 Teaching Ambassador Fellow at the U.S. Department of Education. **JOAN DABROWSKI** is a project manager at the Center for Applied Child Development at Tufts University, Medford, Mass.

ers must answer three questions in order to move forward on this:

- What are the current limits of value-added measures?
- What combination of measures could be used to determine teaching effectiveness?
- How can we move away from divisive rhetoric toward necessary teacher compensation reform?

### THE LIMITS

Tennessee has the longest history of using a value-added system tied to state assessments. The state began using Bill Sanders’ model in the early 1990s. Sanders’ Tennessee Value-Added Assessment System (TVAAS) is one of the most sophisticated and respected value-added models in use. Tennessee uses TVAAS in grades four through eight in math, reading, science, and social studies. In fact, in an attempt to receive Race to the Top funds from the U.S. Department of Education, Tennessee has passed legislation that requires 35% to 50% of a teacher’s evaluation to be based on TVAAS data if it is available. TVAAS is also used in Pennsylvania, North Carolina, and local districts across the country.

However, the model has been criticized for its lack of external review, lack of transparency, issues with missing data, and its lack of consideration of student background variables (Amrein-Beardsley 2008). Indeed, an important part of the model is cloaked in proprietary secrecy. In addition, recent analysis shows that value-added models such as TVAAS can retroactively affect a student’s prior year of growth data. In other words, a 5th-grade teacher can affect a student’s 4th-grade score. Thus, we need to question the validity of such measures (Rothstein 2008).

Value-added measures must have three elements to be viable:

1. Close, but not perfect alignment of assessments and curriculum;
2. Appropriate reliabilities; and
3. Sufficient “stretch” in the assessments to show growth (Sanders 2003).

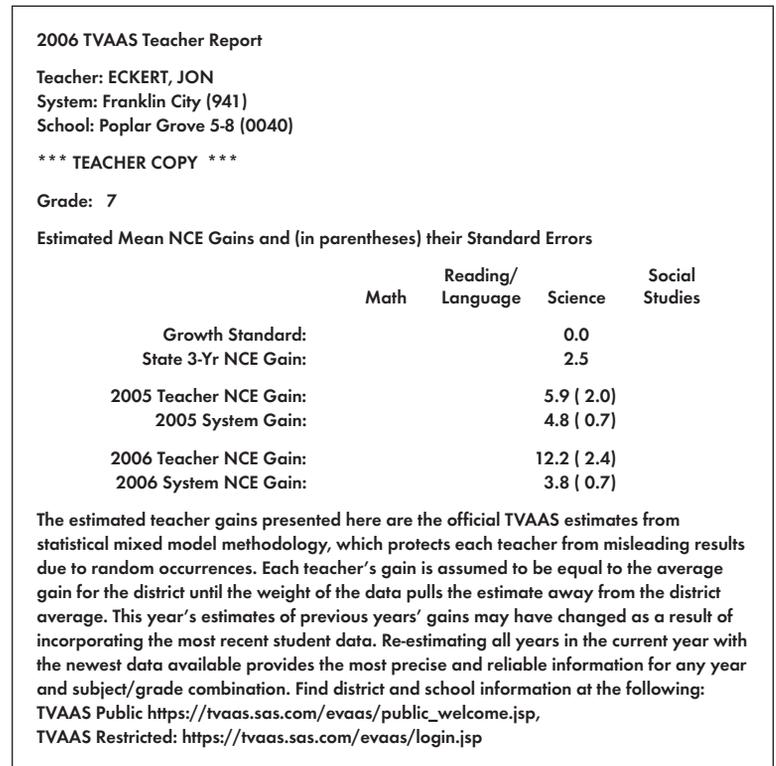
Sufficient stretch means assessments can’t have ceiling effects that prevent a student from showing the full extent of his or her knowledge. Koedel and Betts (2008) question the ability of current state assessment measures to provide sufficient stretch. “If different tests emphasize different kinds of material or have different effective maximums or ‘test ceilings,’ the measure effect of a teacher can vary de-

pending on the test instrument being used” (Lockwood, McCaffrey, and Sass 2008:14). Considering all three requirements, current assessment measures may not be sufficient or appropriate to tie to performance pay (Ballou 2002).

Sanders’ and TVAAS deserve credit for advancing thinking about assessment beyond the traditional “snapshot of a student” toward a measurement of growth over time. They do this by tracking student performance on the Tennessee Comprehensive Assessment Program (TCAP) tests from 3rd to 8th grade to determine their growth trajectory. The TVAAS system calculates how students should score to determine a year of growth. After the TCAP, the state releases scores that indicate the amount of growth students demonstrated. Newspapers report these gains by school as letter grades. An expected year’s growth is a C.

Each spring, the principal and teachers in grades 4 through 8 in the four tested subjects receive a computer printout with the “teacher-effect” score. In addition to the validity and reliability issues embedded in value-added measures — and beyond the fact that only teachers of four subjects in grades 4 through 8 receive value-added scores — the actual reports are particularly problematic. Figures 1 through 3 demonstrate the problems with the TVAAS reports — even for a teacher with outstanding teacher-effect scores.

FIG. 1. Teacher Effect Data for Jon Eckert 2006



**FIG. 2.**  
**Teacher Effect Data for Jon Eckert 2007**

<b>2007 TVAAS Teacher Report</b>				
Teacher: ECKERT, JON				
System: Franklin City (941)				
School: Poplar Grove 5-8 (0040)				
*** TEACHER COPY ***				
Grade: 7				
Estimated Mean NCE Gains and (in parentheses) their Standard Errors				
	Math	Reading/ Language	Science	Social Studies
Growth Standard:			0.0	
State 3-Yr NCE Gain:			3.9	
2005 Teacher NCE Gain:			5.9 ( 1.9)	
2005 System Gain:			4.8 ( 0.7)	
2006 Teacher NCE Gain:			8.6 ( 1.9)	
2006 System NCE Gain:			3.8 ( 0.7)	
2007 Teacher NCE Gain:			17.5 ( 1.0)	
2007 System NCE Gain:			9.1 ( 0.4)	
Teacher 3-Yr-Avg:			10.7 ( 1.0)	
System 3-Yr-Avg:			6.9 ( 0.4)	
<b>Teacher 3-Year-Average Gain Comparisons</b>				
Teacher vs Gr.Std.:			Above Gr.Std.	
Teacher vs State:			Above Mean	
Teacher vs System:			Above Mean	
<p>The estimated teacher gains presented here are the official TVAAS estimates from statistical mixed model methodology, which protects each teacher from misleading results due to random occurrences. Each teacher's gain is assumed to be equal to the average gain for the district until the weight of the data pulls the estimate away from the district average. This year's estimates of previous years' gains may have changed as a result of incorporating the most recent student data. Re-estimating all years in the current year with the newest data available provides the most precise and reliable information for any year and subject/grade combination. Find district and school information at the following: TVAAS Public <a href="https://tvaas.sas.com/evaas/public_welcome.jsp">https://tvaas.sas.com/evaas/public_welcome.jsp</a>, TVAAS Restricted: <a href="https://tvaas.sas.com/evaas/login.jsp">https://tvaas.sas.com/evaas/login.jsp</a></p>				

*First, there is no explanation of how schools can use these data to improve teaching.* How does a teacher act on these data? The principal's report is identical to the teacher's, so what improvement feedback can a principal give? High-stakes assessments are already time-consuming. If these assessments are linked with compensation, we can expect that even *more* time will be spent preparing for them. And for all of the time and energy spent to receive it, the data sheet offers no useful feedback for either the teacher or students.

*Second, the numbers fluctuate from year to year.* For example, the teacher-effect score for 2007 changes from 17.5 to 15.1 on the 2008 report. This is the same test, with the same students, but it changes because of some of the recalculations that the bottom of the form warns may occur. TVAAS made additional changes in summer 2009 in response to a decision by Tennessee authorities that too many teachers, schools, and districts were demonstrating growth

(Tennessee Department of Education 2009). A new benchmark for comparison was applied for scores going back to 2005.

What would happen if performance pay were attached to a system that continually revises its performance standards? That happened in Tennessee, for example, when Gov. Phil Bredesen pushed to increase standards. Teachers who had a good teacher-effect score one year, and thus received a bonus, might find after revision the following year that they

Many of the issues teachers have with teacher compensation reform could be resolved if they were a part of the planning and implementation process.

had actually dropped below the benchmark. Do they return their bonus? More important, what are the legal issues for a district when teachers' property rights, their salaries, are tied to fluctuating measures such as these? As Dan Goldhaber (2006) cautions, more research is needed before using student achievement tests to gauge teacher effectiveness. Without such analysis, legal issues will arise.

Interestingly, in a major study linking value-added and performance pay, the National Center on Performance Incentives at Vanderbilt University (NCPI 2009) chose not to use TVAAS to measure the teacher effectiveness of the middle school math teachers in its study. They used a simplified, transparent, value-added measure of teachers. Other work being done by Rob Meyer at the Value-Added Research Center at the University of Wisconsin and research being done by the CALDER Institute, a branch of the Urban Institute that focuses on longitudinal measures of student learning, will further inform the use of value-added measures.

#### **WHAT MEASURES CAN BE USED?**

Administrator evaluations are currently the primary measure of teacher effectiveness. However, the New Teacher Project's "Widget Effect Report" found that, in the 12 districts in four states in their study, 99% of teachers receive satisfactory ratings when the options are either "satisfactory" or "unsatisfactory." Even when a broader range of options is given, less than 1% of teachers receive a rating of unsatisfactory (Weisberg et al. 2009).

The TAP system, developed by the National Institute for Excellence in Teaching, addresses this issue of "inflated" evaluation scores by incorporating

multiple measures, including value-added measures. TAP sets the average teacher evaluation score at 3.5 on a 5-point scale and distributes the scores on a normal curve. The evaluations are based on multiple observations by administrators and master and mentor teachers, which makes it possible to determine inter-rater reliability. In addition to these observations, the TAP system measures teacher effectiveness using value-added measures (Eckert 2009). Many TAP schools use Sanders' TVAAS model, but several schools use Meyer's Wisconsin value-added model.

Local assessments also can provide data. Many school districts have moved to common assessments, and these assessments can be used if they have at least an acceptable level of validity and reliability. For example, Denver Public Schools use local data to show growth as one component of the ProComp system, which differentiates teacher pay based on contribution to student learning (<http://denverprocomp.dpsk12.org>). Teachers collaborate with principals to set two annual goals and document student progress toward these goals based on district, school, or class assessments. While not a perfect model, the emphasis on student growth is commendable. The strength of multiple measures is that no one measure needs to be a perfect indicator. Given the complexity of teaching and learning, no single measure can be a perfect indicator.

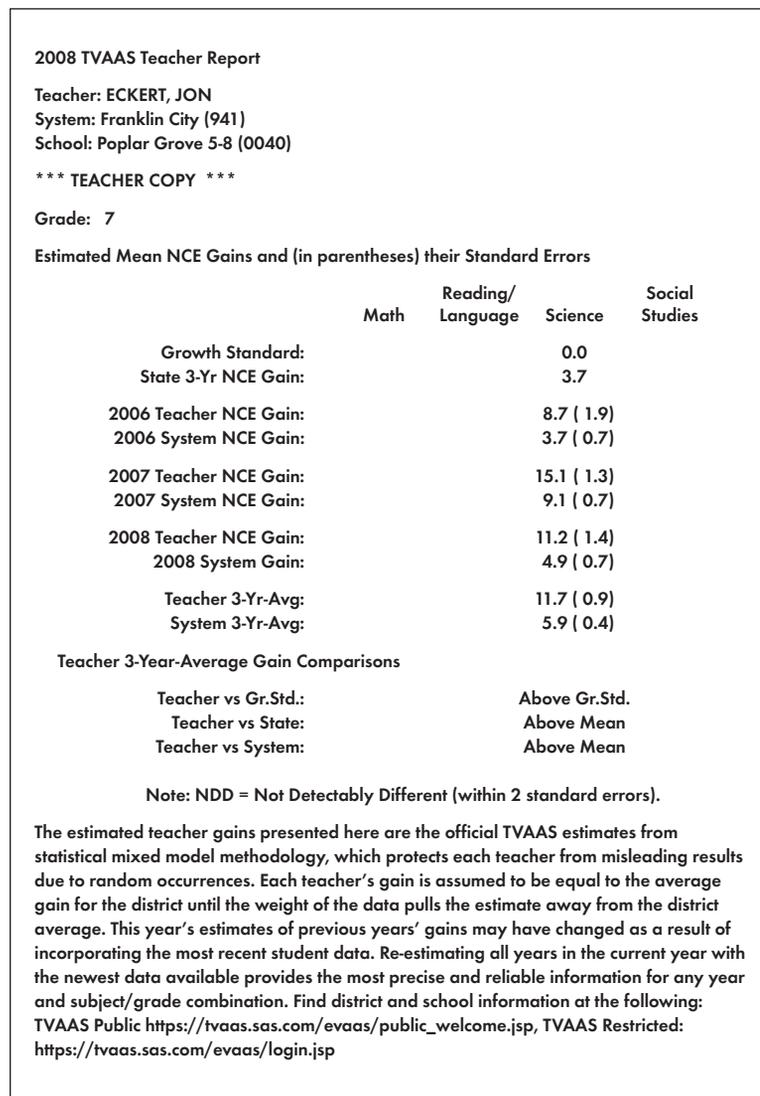
### AVOIDING DIVISIVE RHETORIC

Such terms as *performance pay* and *merit pay* often provoke highly charged responses. This combative dynamic hinders the thoughtful reform of teacher compensation. While very little is known about how to compensate teachers in a way that rewards excellence, we do have examples and ideas to help us move forward in this task.

The TAP system, for example, uses an array of measures to identify master teachers. These teachers then observe, mentor, and coach other teachers. Professional development becomes job-embedded and linked to student outcomes through the work of these teachers. Master and mentor teachers receive compensation for their additional work, and all teachers are eligible for bonuses based on measures of their students' growth, the growth of the school, and comprehensive teacher evaluations. TAP is just one possibility for how teacher compensation can be based on outcomes (student learning) instead of inputs (degree attainment and experience).

If teachers see value in these assessments beyond strict accountability and teacher compensation reform, their opposition to assessment might decrease. Moreover, if teachers feel that these assessments accurately represent the effect the teachers have on

**FIG. 3.**  
**Teacher Effect Data for Jon Eckert 2008**



their students' learning, their opposition would decrease even further. Indeed, many of the issues teachers have with teacher compensation reform could be resolved if they were a part of the planning and implementation process.

Many districts pay teachers significantly higher salaries if the teachers have earned master's degrees, though there is little evidence that a master's degree will improve student achievement. Districts across the United States spend over \$8.6 billion on the master's salary increase alone (Roza and Miller 2009). Reallocating these funds would allow states and districts some flexibility in how they compensate their most effective educators.

Value-added measures are a critical piece of compensation reform, but we should never expect a single number to encapsulate the tremendous impact teachers have on students. Districts should be en-

couraged to develop teacher compensation reform plans *with* teachers. Compensation reform based on teaching effectiveness is long overdue. **K**

## REFERENCES

Amrein-Beardsley, Audrey. "Methodological Concerns About the Education Value-Added Assessment System." *Educational Researcher* 37, no. 2 (2008): 65-75.

Ballou, Dale. "Sizing Up Test Scores." *Education Next* 2, no. 2 (Summer 2002): 10-15.

Eckert, Jonathan. *More Than Widgets: A Systematic Approach to Increased Teaching Effectiveness*. Santa Monica, Calif.: National Institute for Excellence in Teaching, 2009. [www.tapsystem.org/publications/ffo\\_rpts\\_eckert.pdf](http://www.tapsystem.org/publications/ffo_rpts_eckert.pdf).

Farmer, Blake. "VU Researcher Questions Direction of Race to the Top." *WPLN News*, Dec. 31, 2009. <http://wpln.org/?p=13606>.

Goldhaber, Dan. *Teacher Pay Reforms: The Political Implications of Recent Research*. Washington, D.C.: Center for American Progress, December 2006. [www.americanprogress.org/issues/2006/12/pdf/teacher\\_pay\\_report.pdf](http://www.americanprogress.org/issues/2006/12/pdf/teacher_pay_report.pdf).

Koedel, Cory, and Julian Betts. "Value-Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation." *NCPI Working Paper 2008-21*. Nashville, Tenn.: National Center on Performance Incentives, June 2008. [www.performanceincentives.org/data/files/news/PapersNews/KoedelBetts\\_200821\\_v2.pdf](http://www.performanceincentives.org/data/files/news/PapersNews/KoedelBetts_200821_v2.pdf).

Lockwood, J.R., Daniel F. McCaffrey, and Tim R. Sass. "The

Intertemporal Stability of Teacher Effect Estimates." Paper presented at the National Conference on Value-Added Modeling, Wisconsin Center for Education Research (WCER), Madison, Wisc., 2008.

National Center on Performance Incentives (NCPI). "Frequently Asked Questions: About Our Work in Metropolitan Nashville Public Schools." 2009. [www.performanceincentives.org/about\\_ncpi/faq\\_nash.asp](http://www.performanceincentives.org/about_ncpi/faq_nash.asp).

Roza, Marguerite, and Raegen Miller. *Separation of Degrees: State-by-State Analysis of Teacher Compensation for Master's Degrees*. Seattle, Wash.: Center on Reinventing Public Education, 2009. [www.crpe.org/cs/crpe/view/csr\\_pubs/289](http://www.crpe.org/cs/crpe/view/csr_pubs/289).

Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *NBER Working Paper no. 14442*. Cambridge, Mass.: National Bureau of Economic Research, October 2008.

Sanders, William L. "Beyond *No Child Left Behind*." Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 2003. [www.sas.com/govedu/edu/no-child.pdf](http://www.sas.com/govedu/edu/no-child.pdf).

Tennessee Department of Education. "Baseline Year Analyses: 2009 Value-Added Assessment Baseline 1998, Updated Baseline 2009." Slide presentation. 2009. [http://tennessee.gov/education/assessment/doc/TVAAS\\_new\\_baseline.pdf](http://tennessee.gov/education/assessment/doc/TVAAS_new_baseline.pdf).

Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. Brooklyn, N.Y.: New Teacher Project, 2009. <http://widgeteffect.org>.



"My term paper is almost finished. I updated my software, defragmented my hard drive, bookmarked an online dictionary, and installed new ink cartridges. Now all I need are some words and a topic!"

Copyright of Phi Delta Kappan is the property of Phi Delta Kappa International and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.