THE HUNT INSTITUTE'S JAN 2014 | No. 04, Part 2

re:VISION

EVALUATING TEACHERS: OPPORTUNITIES AND BEST PRACTICES

By Stephen Jackson, Ph.D., Policy Analyst and Casey Remer, Interim Director of Research and Policy Analysis

Until recently, states focused on ensuring the presence of a "highly qualified teacher" in every classroom. Under the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), this title described a teacher holding at least a bachelor's degree and the appropriate state license and demonstrating subject matter competency.

But, research has shown that these "input" measures do little to explain differences in student performance. Thus, in recent years, the conversation has shifted to a focus on "outputs"—how effective a teacher is at improving student achievement.

Knowing how well, or how poorly, educators are performing is critical to drive improvement strategies for individual teachers, schools, districts, and states and to inform accountability systems. **High-quality teacher evaluation data can also be used to inform policies across the education system**, including measuring the effectiveness of teacher preparation programs, informing performance-based compensation, ensuring students have equal access to highly effective teachers, and identifying professional development needs.

EDUCATOR EFFECTIVENESS SERIES: SNAPSHOT

This is **part two** of a special series on improving the effectiveness of the nation's teachers and leaders. The other briefs in the series are:

- Overview
- Teacher Compensation
- Teacher Preparation
- School Leadership

Available at www.hunt-institute.org

THE hunt institute

Created by The University of North Carolina Board of Governors in 2001 to honor and continue the legacy of former North Carolina Governor Jim Hunt, The Hunt Institute collaborates with policymakers to secure America's future through quality education. Working at the intersection of policy and politics, The Hunt Institute connects leaders with the best strategies for developing and implementing policies and programs to improve public education. To that end, The Institute convenes governors, policymakers, and legislators, as well as business, education, and civic leaders across the nation to provide them with the best information to make informed policy decisions.

The need for quality data on teacher performance is clear, and the reform of teacher evaluation systems must be considered in the context of other activities designed to improve educator effectiveness. This issue of ***re:VISION***, part of a special series on teacher effectiveness, examines the evolution of teacher evaluation systems and the most commonly used evaluation measures and offers considerations for policymakers who are examining teacher evaluation in their states.

CURRENT CONTEXT

Until very recently, most teachers were evaluated only once every few years.[1] *The Widget Effect*, a 2009 report by The New Teacher Project (TNTP), found that under a binary rating system in which the choices were limited to "satisfactory" or "unsatisfactory," more than 99 percent of teachers received the satisfactory rating. For policymakers and education leaders who feel an urgent need to ensure that all of their state's students are taught by an effective teacher, reports such as *The Widget Effect* rang alarm bells.

In the last five years, however, **states have made substantial strides to improve teacher evaluation**. With incentives from the U.S. Department of Education, beginning with the *Race to the Top* competition in 2009 and followed by *ESEA* accountability waivers in 2011, states have moved towards systems that include multiple levels of performance classification, require more frequent evaluation for all teachers, and incorporate multiple measures, including student achievement.[2]

As a result of this rapid reform, **the number of states with annual evaluations of teachers increased from 15 states in 2009 to 28 states in 2013**. Forty-three states require more than two performance categories, up from 17 states in 2011. Perhaps most notably, 41 states now require teacher evaluations to include measures of student achievement, up from only 15 states in 2009.[3]

Common Student Achievement Measures: A Glossary

- **VALUE-ADDED MODELING (VAM)**
  A measure of student growth on standardized tests that can be attributed to the classroom teacher *(See pages 3-4 for more information.)*
- **STUDENT GROWTH PERCENTILES (SGP)**
  The rank—expressed as a percentile—of the growth of a student's standardized test score compared to students with similar scores on previous tests. *(See page 4 for more information.)*
- **STUDENT LEARNING OBJECTIVES (SLOS)**
  Measurable goals for student learning set by teachers and/or principals for both tested and non-tested subjects *(See pages 4-5 for more information.)*
- **PORTFOLIOS, PROJECTS, PERFORMANCE, PRODUCTS (THE FOUR P'S)**
  Measures of student output sometimes used for subjects that are not easily tested *(See page 5 for more information.)*

EVALUATION MEASURES

An effective evaluation system that can inform teacher development and accountability will **differentiate between the best-, worst-, and average-performing teachers by using multiple measures such as student achievement and teacher practice**.[4] Teaching is complex work, and no one measure could possibly capture all aspects. The *Measures of Effective Teaching Project* (MET) study found that student achievement, classroom observation ratings, and student survey responses in **combination** were better predictors of student performance than when each was used separately.[5] The multiple measures capture different aspects of teaching and learning and provide a more complete picture of what a teacher is doing and the effect on learning he or she has. Using a combination of measures can increase teacher confidence in the fairness of the system, as well as provide more fine-grained information that can be used to inform teacher growth and development.[6]

As states work on the design and/or redesign of their evaluation systems, the most common measures are student achievement, teacher observation, and student surveys.

Measures of Student Achievement

Of the 41 states and Washington, D.C. that now require the use of student achievement measures, 20 states require that student achievement be the most important measure contributing to the final teacher evaluation score; 16 states require that the student achievement measure be a significant contributor to the final evaluation score; and five states require, at a minimum, that only some measure of student learning be included.[7]

Thirty-one states require that the student achievement measure include the use of standardized state tests in tested grades and subjects. States that use standardized tests to produce measures of student growth must take care that the tests used are aligned to their curriculum standards. This alignment will ensure that the data generated is a valid measure of what is taught and learned.

Fewer states have developed requirements for non-tested subjects—those subjects not requiring testing under *ESEA*. This is a challenging area for states, as it is estimated that over two-thirds of teachers teach in these grades and subject areas.[8]

VALUE-ADDED MODELING

Value-added models (VAM) are designed to analyze student performance on standardized tests compared to an expected student growth trajectory. These often-complex statistical models typically attribute growth to an individual teacher by controlling for factors outside of the teacher's control such as student background. That growth, once other factors are considered, is a teacher's "value-add"— **an estimate of how much that teacher improved or depressed student achievement**. Proponents of VAM cite research which shows that a teacher's VAM score predicts the future performance of students taught by that teacher far better and far more reliably than any other variable or method, including years of experience and degrees.[9]

However, even proponents of VAM readily acknowledge that this method **should only be used in combination with other measures of teacher effectiveness**.[10] It is important to recognize that VAM is not as precise as one might hope: **a value-add score is only an estimate of a teacher's true impact**. The scores represent the mid-point of a range of probable scores for a teacher, rather than an exact point.[11] This means that while VAM is rather accurate at identifying the very best and very worst teachers, significant room for error exists regarding teachers in the middle.[12]

Using several years of data can increase the precision of VAM. This approach is the one adopted by **North Carolina**, for instance, which requires three years of data before calculating a score. **Louisiana** requires the use of additional data to parse teachers who score in the middle VAM range. Those rated between the 20th and 80th percentiles are assigned an evaluation rating after evaluators review student learning objective (SLO) data *(See pages 4-5 for more information on SLOs)*.

Given the complexity of VAM, states have confronted several practical challenges when trying to incorporate value-add scores into their evaluation systems for all teachers. **States need to determine how to measure student growth for teachers of untested subjects and grades**. Some states and districts, including **North Carolina**, **Michigan**, and **Hillsborough County (FL)** Public Schools, have developed additional state tests in all, or nearly all, subjects. Other states, such as **Ohio**, allow districts to choose approved tests from private vendors.

> Discussion of the **evaluation of school principals** can be found in the accompanying brief on *school leadership*.

It is also important to ensure that teachers are rated on the students they taught, not someone else's. This point has been a source of tension in states and districts where **school-wide VAM scores** have been given to teachers of non-tested subjects, but it is also a complication for schools that utilize team teaching. One study of a large urban district found that teachers team-taught around one-fifth of math and reading students. Approximately one in 14 teachers shared all their students with another teacher.[13] Addressing this situation requires data systems that include strong teacher-student data links that are informed by a parsimonious definition of the "teacher of record." Data needs to be collected several times a year, with a system in place for verifying rosters.

STUDENT GROWTH PERCENTILES

An alternative to VAM is student growth percentiles (SGP). This statistical method ranks students' academic growth compared to their peers who scored similarly on prior tests.[14] Districts in many states, including **Washington**, **Colorado**, **New Jersey**, and **Massachusetts**, are using student SGPs in their teacher evaluations. To date, states that have used this method have not controlled for student demographics; it is purely an assessment of relative growth.

Supporters of the method argue that SGPs provide a growth comparison between students with similar prior achievement or, when aggregated, schools. The results are rankings expressed as a percentile, and this expression means they may be easier to understand than ones from VAM models.[15] One significant practical advantage is that SGPs do not need data from tests that have similar scales, meaning states and districts can mix and match tests.[16] However, states using student growth percentiles share many of the same challenges as those using VAM: what to do about subjects and grades without tests and the correct attribution of scores in classes with team teaching.

Like VAM, SGPs also produce estimates of learning effects, not precise scores. Compared to VAM, an SGP score has relatively wider ranges of probable scores, and therefore, shares the need for multiple years of data to reduce this range. It shares with VAM, therefore, **the need to use additional evaluation measures**.[17]

STUDENT LEARNING OBJECTIVES

Student learning objectives (SLOs) are measurable goals for what students will learn over a set period of time and can be written for both tested and non-tested subjects.[18] When used in teacher-evaluation systems, SLOs are usually negotiated between a teacher—or group of teachers—and the principal. Fourteen states, including **Connecticut**, **Wisconsin**, and **Georgia**, **require the use of SLOs, while an additional six states explicitly allow districts to use SLOs**.

Rhode Island's state model uses SLOs as a measure of student achievement for all teachers and to supplement reading and math SGPs. Districts may adopt this model or adopt their own systems based on state principles. The basis of the system is that teachers, in consult with their principals and consistent with district and school growth goals, establish learning goals for their own students using curriculum and pacing guides, taking into account the previous performance of students. **Rhode Island educators cite the collaborative process around goal setting as a valuable development that improved their teaching and made them think concretely about improving student learning**.[19]

The main challenge around the SLO model is that it can be difficult to compare results across schools and districts. Some SLOs may be established using test scores (as in **Oregon**), but the learning goals chosen can still vary across schools and districts. Because of the time required, SLOs can also be more costly than

other student achievement measures. **Rhode Island** has expended a significant amount of time and *Race to the Top*-funded dollars establishing guidelines for their use.

PORTFOLIOS, PROJECTS, PERFORMANCE, PRODUCTS (THE FOUR P'S)

Some subjects are not easily tested, such as the fine arts and drama, and lend themselves to assessment based on student output during the school year. Some states have also decided that non-test measures of student achievement enable them to capture a more nuanced and detailed picture of student learning and use them in the tested subjects as well. **Connecticut** requires that 22.5 percent of a teacher's evaluation draw from non-standardized test measures of student learning, including portfolios and teacher-developed assessments. **New York** allows districts to incorporate additional evaluation factors that include non-test student growth measures, such as structured reviews of student work, portfolios, or evidence binders. The weight for these additional factors cannot exceed five percent.

The major caveats to the 'four Ps' echo those for SLOs. They render data that makes comparison between teachers difficult, and they take substantial time for supervisors and principals to assess. The ability to make fair comparisons between teachers becomes especially important when evaluations are tied to high-stakes decisions like performance-based pay.

Classroom Observation

Given that most states already included observation in their old evaluation systems, it is no surprise that 45 states now require districts to use teacher observations in their evaluation plans. Twenty-five states require multiple observations.[20] **Research shows that unbiased observations capture differences in teaching that lead to differences in student test performance.**[21]

A significant challenge for observation remains unchanged from the pre-reform era: ensuring that the observation is fair and unbiased. The key to valid and reliable observation of teachers is standards-based evaluator training and certification, backed by ongoing calibration checks.

Research suggests that principals and other observers need training in order for them to understand the difference between bias, interpretation, and evidence. This approach includes exposure to a variety of lessons — of varying quality — using video examples of teaching that have been reliably scored.

Research has shown that high-quality, meaningful evaluator certification requires 35 to 50 hours of training, at a minimum.[22] **Hillsborough County (FL) Public Schools**

The Measures of Effective Teaching Study

The *Measures of Effective Teaching Project* (MET), funded by the Bill & Melinda Gates Foundation, examined how effective teaching could be measured using classroom observation, measures of student achievement, and student surveys. Some of the major findings included:

- **Results from high-quality teacher observations were correlated with a teacher's value-added score.**
- **Having a second observer in a classroom made for more reliable scores than using multiple observations from one observer.**
- **There is a strong correlation between data from well-designed student surveys and student achievement growth.**
- **Combining observation scores with student survey and value-added scores produces a stable, reliable, and valid measure of educator effectiveness.**
- **Observation, VAM, and student survey data in combination were better predictors of student achievement than teaching experience and graduate degrees.**

Source: Bill & Melinda Gates Foundation (2013).

has a nationally recognized evaluator observation program that illustrates the resources and planning needed to implement observations effectively (*See box*: *Teacher Observation in Hillsborough County, Florida*).

There are various ways that states ensure principals and other observers receive necessary training. **Connecticut** has incorporated evaluation training, including classroom observation, into the state standards for principals. *The Connecticut Administrator Test*, required for principal certification and for other administrative positions, has a teacher observation component.[23] In **New York**, the state has set requirements for teacher and principal observations that districts must use when certifying lead evaluators.[24]

Despite training, the validity of scores from certified observers may decline over time. To ensure that certified observers continue to produce reliable scores, districts run regular calibration checks and spot-training for observers. Calibration checks compare the scoring of certified evaluators to benchmarked scoring, usually based on videotaped lessons. Some districts, such as **Guilford County, North Carolina**, calibrate observation ratings weekly. A growing number of online services offer relatively quick and inexpensive calibration checks and spot-training for observers. The main challenge for states and districts is the substantial amount of time required to train observers, calibrate ratings, and conduct observations.

Giving an honest and valid appraisal is a challenge for observers, especially when they are from the same school. Research suggests using multiple observers to improve the reliability of results, but very few states require or recommend them.[25] **Hillsborough County (FL) Public Schools** and **Denver Public Schools** require the use of multiple observers in their nationally recognized evaluation programs. In **Kentucky**, additional evaluations from peer evaluators can be requested by a teacher.

Teacher Observation in Hillsborough County, Florida

The attention to detail by the **Hillsborough County (FL) School District** is illustrative of how to ensure high-quality teacher observations.

Sixty percent of a teacher's annual evaluation rating is determined by peer (25 percent) and principal (35 percent) observations. Teachers with more than two years' experience are observed two-to-five times annually, depending on their prior performance, by their principal and peer teachers. Teachers rated "unsatisfactory" are also observed by a supervisor.

First- and second-year teachers receive weekly or bi-weekly support from a mentor and are observed six times annually by their administrator(s) and mentor. Mentors evaluate each other's mentees.

Formal observations are supplemented by shorter, informal, and unannounced observations. All observations are used for development, as well as evaluative purposes.

In order to ensure adequate capacity to complete these observations, master teachers are released from their teaching duties and serve as peer observers, rotating through all the schools in the district. Each peer observer serves approximately 100 teachers. Mentor evaluators work with 15 to 20 new teachers.

In Hillsborough County, principals and other teacher observers are trained together. In order to ensure reliability of scoring, an external organization is engaged annually to observe and review each trained and certified evaluator.

At the end of the year, peer and principal reviewers produce a summative annual observation rating. There is a reported 80 percent correlation in observation scores between peer and principal summative scores. The correlation between VAM and observation scores is equivalent to those obtained in the *MET* study (*See box: The Measures of Effective Teaching Project*). **A portion of a principal's annual evaluation is based on the school's correlation between VAM and observation scores.**

Source: Personal communication, Dr. David Steele, former Chief Information and Technology Officer, Hillsborough County Public Schools, Florida (2013).

Generating valid and reliable observation scores is very important because it is the major differentiator of final evaluation scores between good, average, and poor teachers. Early state adopters of new evaluation systems such as **Florida**, **Michigan**, and **Tennessee** are struggling with this: there is not yet significant differentiation in their summative evaluation scores. Early results from those three states show that 97 percent or more of teachers have been rated "effective."

It is also important to consider teacher evaluations in the context of other education reforms. The new career and college ready state standards for math and English Language Arts stress the importance of engagement with texts, critical thinking, and the sequencing of learning to build deep knowledge. It is far from clear that the most commonly used observation tools adequately reflect that re-orientation. A recent TNTP brief argues that many of the observation tools used by districts have not changed to reflect the emphasis in the new standards on textual analysis and logical sequencing of concepts and content, for instance. Instead, they continue to focus on the mechanics of teaching, such as behavior management or use of time.[27]

Student Surveys

Student survey data can capture aspects of a teacher's teaching style and effectiveness not apparent from test data. About half the states have evaluation guidelines that allow or recommend that districts use student surveys to collect feedback on teacher performance. **Connecticut's** state model allows districts to use student surveys to count towards five percent of the final rating for teachers.

One option for consideration is the *Tripod* survey. *Tripod* survey questions include inquiries about the teacher's attitude, ability to clarify concepts and challenge the student, and whether he or she had consolidated prior learning. The *MET* project studied whether the *Tripod* survey was correlated with student achievement and teacher effectiveness. The study ranked teachers based on the favorability of survey responses and found that ranking predicted achievement of other students taught by the same teacher. When math teachers were ranked based on their students' responses to the survey, those ranked in the top 25 percent had students who showed greater student achievement growth than students taught by teachers in the bottom 25 percent. The differences were quite substantial—the equivalent of over four and a half months of extra learning time.[28]

North Carolina has piloted student surveys using an adapted *Tripod* instrument and has found acceptable correlations between results and student achievement growth.[29]

PUTTING THE PIECES TOGETHER

As states and districts work to incorporate the measures described above into a coherent evaluation system, they must contend with two major issues: how to differentiate teacher performance and how to weight the measures.[30]

The number of final summative evaluation categories a state or district chooses to differentiate teachers will depend on the accountability, compensation, and development purposes for which the system is used, as well as the quality of the data expected from the evaluation components. The most typical system sorts teachers into four categories: top performers, two intermediate categories, and a struggling or ineffective category.[31]

The *MET* project findings on weighting are worth highlighting: **giving test score data a weight between 33 and 50 percent, and then combining that with survey results and observation scores, produces an evaluation formula that is highly predictive of student performance on other tests and varies little from year-to-year**.[32]

CONSIDERATIONS FOR POLICYMAKERS

As state policymakers consider the design and redesign of their evaluation systems, there are some broad sustainability issues to consider:

Ensure Teachers Support the System

States that have included teacher input when designing their evaluation systems are much less likely to experience pushback on their reforms. States and districts that have been committed to providing feedback to teachers, post-evaluation, have demonstrated that evaluation is both a development and accountability tool.

In **Colorado**, **New Mexico**, and **Iowa**, state-appointed commissions—which included policymakers, community members, teachers, and association leaders—considered how to measure and evaluate effective teaching. The commissions sought input from a broad audience representing traditional public schools and charters—including parents, administrators, and teachers—before delivering recommendations to the governor or the legislature.

When the initial rollout of **Tennessee's** new evaluation system was met with resistance from teachers, Governor Bill Haslam asked the **State Collaborative on Reforming Education (SCORE)** to conduct a statewide listening tour to gather feedback on the system. SCORE held meetings, interviewed stakeholders, and conducted focus groups with teachers, administrators, stakeholders, and partners across the state. This tour was accompanied by an online survey. The process provided feedback from more than 27,000 people in the form of a report to the Tennessee Department of Education and State Board of Education. **SCORE's recommendations focused on refinements to the teacher evaluation system and resulted in improvements to the quality of observer training and teacher observations, the use of student growth measures in non-tested subjects, and improved teacher development opportunities**, among others. The changes have improved acceptance among teachers for the new evaluation system.[33]

Monitoring and Developing the Evaluation System

Differences between individual district systems of evaluation can promote rancor among teachers because of perceptions of unequal treatment, but of the 39 states that offer some degree of flexibility to districts, only 19 require state review and approval of locally designed systems. **Colorado** and **Oregon**, both states with strong local control, are two of four states that review district plans. Fifteen states require explicit state approval of local plans, including **Texas**, **Florida**, **Kentucky**, and **Louisiana**.[34]

Data Systems

In all but a handful of states, the "bricks and mortar" for functioning data systems are more or less in place.[35] The challenge for states now is to ensure better use of the data. There has been progress but, according to the *Data Quality Campaign*, "the hardest work remains." **Areas requiring attention include training educators on how to access, analyze, and interpret the data; reaching out to non-education stakeholders so they can understand how to use and interpret data; and providing data access to local stakeholders, including parents**.[36]

Denver, CO, was a pioneer in the development of a readily useable interface for teachers to access evaluation data for the purpose of their own professional development and to use for instructional planning. Several states, including **Colorado**, **Louisiana**, and **North Carolina**, have developed Web pages that enable lesson exemplars to be shared, evaluation-related data to be inputted, and data summaries to be extracted by teachers, principals and administrators.

References

1 WEISBURG, DANIEL, SUSAN SEXTON, JENNIFER MULHERN, AND DAVID KEELING. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness, 2nd Edition*. New York, NY: The New Teacher Project. Available at: http://widgeteffect.org/downloads/TheWidgetEffect.pdf

2 US DEPARTMENT OF EDUCATION. (2012). *ESEA Flexibility, updated June 7, 2012.* Washington, DC: Author. Available at: http://www.ed.gov/esea/flexibility/documents/esea-flexibility-acc.doc

3 DOHERTY, KATHRYN, AND SANDI JACOBS. (2013). *State of the States 2013, Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice*. Washington, DC: National Council on Teacher Quality. Available at: http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

4 BILL & MELINDA GATES FOUNDATION. (2013a). Feedback for Better Teaching: Nine Principles for Using Measures of Effective Teaching. *MET Project*. Seattle, WA: Author. Available at: http://www.metproject.org/downloads/ MET_Feedback%20for%20Better%20Teaching_Principles%20Paper.pdf; Bill & Melinda Gates Foundation. (2013b). Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three Year Study. *MET Project Policy and Practice Brief.* Seattle, WA: Author. Available at: http://www.metproject .org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

5 BILL & MELINDA GATES FOUNDATION. (2013b).

6 GOE, LAURA, LYNN HOLDHEIDE, AND TRICIA MILLER. (2011). A Practical Guide to Designing Comprehensive Teacher Evaluation Systems: *A Tool to Assist in the Development of Teacher Evaluation Systems*. Washington, DC: National Comprehensive Center for Teacher Quality. Available at: http://www.lauragoe.com/lauragoe/practicalGuideEval Systems.pdf

7 DOHERTY AND JACOBS. (2013).

8 PRINCE, CYNTHIA, PATRICK SCHUERMANN, JAMES GUTHRIE, PETER WITHAM, ANTHONY MILANOWSKI, AND CHRISTOPHER THORN. (2009). *The Other 69 Percent: Fairly Rewarding the Performance of Teachers of Nontested Subjects and Grades. Guide to Implementation: Resources for Applied Practice*. Washington, DC: Center for Education Compensation Reform and US Department of Education, Office of Elementary and Secondary Education. Available at: http://cecr.ed.gov/guides/other69Percent.pdf

9 KANE, THOMAS, DANIEL MCCAFFREY, TREY MILLER, AND DOUGLAS STAIGER. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching using Random Assignment. *MET Project Research Paper*. Seattle Washington: Bill & Melinda Gates Foundation. Available at: http://www.metproject.org/downloads/ MET_Validating_Using_Random_Assignment_Research_Paper.pdf

10 GLAZERMAN, STEVEN, DAN GOLDHABER, SUSANNA LOEB, DOUGLAS STAIGER, STEPHEN RAUDENBUSH, AND GROVER WHITEHURST. (2010). Value Added: It's Not Perfect, But It Makes Sense. *Education Week* (12/15/10). Available at: http://www.edweek.org/ew/articles/2010/12/15/15whitehurst.h30.html

11 RAUDENBUSH, STEPHEN, AND MARSHALL JEAN. (2012). *How Should Educators Interpret Value-Added Scores?* Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Available at: http://www.carnegieknowledgenetwork .org/wp-content/uploads/2012/10/CKN_2012-10_Raudenbush.pdf

12 IBID.

13 HOCK, HEINRICH, AND ERIC ISENBERG. (2012). *Methods for Accounting for Co-Teaching in Value-Added Models*. Princeton, NJ: Mathematica Policy Research. Available at: http://www.mathematica-mpr.com/publications/pdfs/education/acctco-teaching_wp.pdf

14 RAND EDUCATION. (nd). *Student Growth Percentiles 101: Using Relative Ranks in Student Test Scores to Help Measure Teaching Effectiveness*. Santa Monica, CA: Author. Available at: http://www.rand.org/content/dam/ rand/pubs/corporate_pubs/2012/RAND_CP693z5-2012-09.pdf

15 BETEBENNER, DAMIAN. (2008). *A Primer on Student Growth Percentiles*. Dover, NH: National Center for the Improvement of Educational Assessment. Available at: http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Aprimeronstudentgrowthpercentiles.pdf

16 IBID.

17 US DEPARTMENT OF EDUCATION. (nd). *Measuring Teacher Effectiveness Using Growth Models: A Primer.* Washington, DC: Author. Available at: http://www.schools.utah.gov/charterschools/Training/Directors-Meetings/2011-Directors-Meetings/October-2011/Measuring-teacher-effectiveness-using-growth-model.aspx

References

18 COMMUNITY TRAINING AND ASSISTANCE CENTER. (nd). Student Learning Objectives. Boston, MA: Author. Available at: http://www.ctacusa.com/studentlearningobjectives.html

19 RHODE ISLAND DEPARTMENT OF EDUCATION. (nd). *Implementing Student Learning Objectives in Rhode Island* [Video file]. Available at: http://media.ride.ri.gov/PD/Eval/ImplementingSLOsRI.mp4

20 DOHERTY AND JACOBS. (2013).

21 KANE ET AL. (2013).

22 MCCLELLAN, CATHERINE, MARK ATKINSON, CHARLOTTE DANIELSON. (2013). Teacher Evaluator Training and Certification: Lessons Learned from the Measures of Effective Teaching Project. *Practitioner Series for Teacher Evaluation*. San Francisco, CA: Teachscape. Available at: http://www.teachscape.com/binaries/content/assets/teachscape-marketing-website/resources/march_13whitepaperteacherevaluatortraining.pdf

23 CONNECTICUT STATE DEPARTMENT OF EDUCATION. (nd). CAT: Connecticut Administrator Test. Hartford, CT: Author. Available at: http://www.sde.ct.gov/sde/cwp/view.asp?A=2641&Q=320422

24 REFORM SUPPORT NETWORK. (nd). *Race to the Top at a Glance, Evaluations of Teacher Effectiveness: State Requirements for Classroom Observations*. Available at: http://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/evaluations-teacher-effectiveness.pdf

25 KANE ET AL. (2013).

26 THE NEW TEACHER PROJECT AND STUDENT ACHIEVEMENT PARTNERS. (2013). Fixing Classroom Observations: How Common Core Will Change the Way We Look at Teaching. New York, NY: The New Teacher Project. Available at: http://tntp.org/assets/documents/TNTP_FixingClassroomObservations_2013.pdf; Anderson, Jenny. (2013). Curious Grade for Teachers: Nearly All Pass. *New York Times* (3/30/13). Available at: http://www.nytimes.com/ 2013/03/31/education/curious-grade-for-teachers-nearly-all-pass.html?_r=0

27 THE NEW TEACHER PROJECT AND STUDENT ACHIEVEMENT PARTNERS. (2013).

28 BILL & MELINDA GATES FOUNDATION. (2012). Asking Students about Teaching: Student Perception Surveys and Their Implementation. *Met Project Policy and Practice Brief*. Seattle, WA: Author. Available at: http://www.metproject.org/downloads/Asking_Students_Practitioner_Brief.pdf

29 PUBLIC SCHOOLS OF NORTH CAROLINA. (nd). Educator Effectiveness Model: Student Surveys. Raleigh, NC: NC Department of Public Instruction, Public Schools of North Carolina. Available at: http://www.ncpublicschools.org/ effectiveness-model/surveys/

30 THE ESEA WAIVER PROCESS REQUIRES THAT STATES HAVE SYSTEMS WITH AT LEAST THREE LEVELS OF EVALUATION. SEE US DEPARTMENT OF EDUCATION. (2012).

31 DOHERTY AND JACOBS. (2013).

32 BILL & MELINDA GATES FOUNDATION. (2013b).

33 STATE COLLABORATIVE ON REFORMING EDUCATION. (2012). *Supporting Effective Teaching in Tennessee: Listening and Gathering Feedback on Tennessee's Teacher Evaluations*. Nashville, TN: Author. Available at: http://www.tnscore. org/wp-content/uploads/2012/06/SCORE-Evaluation-Report.pdf

34 DOHERTY AND JACOBS. (2013).

35 DATA QUALITY CAMPAIGN. (2013). State Analysis by Essential Element: 10 Essential Elements. Washington, DC: Author. Available at: http://www.dataqualitycampaign.org/your-states-progress/10-essential-elements/

36 SEE: DATA QUALITY CAMPAIGN. (2013). State Analysis by Essential Element: 10 State Actions. Washington, DC: Author. Available at: http://www.dataqualitycampaign.org/your-states-progress/10-state-actions/

This is a special series on improving the effectiveness of the nation's teachers and leaders. The briefs in the series are:

PART ONE | IMPROVING THE QUALITY OF TEACHERS AND PRINCIPALS

PART TWO | EVALUATING TEACHERS: OPPORTUNITIES AND BEST PRACTICES

PART THREE | PAYING FOR IMPROVEMENT: TEACHER COMPENSATION REFORM

PART FOUR | GETTING TEACHERS LEARNER-READY: REFORMING TEACHER PREPARATION

PART FIVE | BUILDING LEADERSHIP IN SCHOOLS

An affiliate of The University of North Carolina at Chapel Hill, The Hunt Institute is a non-partisan, nonprofit service entity that does not lobby for, or take positions on, state and federal legislation.

JAMES B. HUNT, JR.
Foundation Board Chairman
Governor of NC (1977-1985 | 1993-2001)

JUDITH A. RIZZO, Ed.D.
Executive Director and CEO

APRIL D. WHITE
Publisher

www.hunt-institute.org
Twitter.com/Hunt Institute
facebook.com/thehuntinstitute
YouTube.com/TheHuntInstitute

For additional copies of ***re:VISION***, or to receive a copy via e-mail, please call 919-425-4160 or email awhite@hunt-institute.org.