



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Evaluating Teacher Effectiveness

Laura Goe, Ph.D.

Presentation to the Hawaii Department of Education

July 20, 2011 ♦ Honolulu, HI

Today's presentation available online

- To download a copy of this presentation or look at it on your iPad, smart phone or laptop, go to www.lauragoe.com
 - Go to Publications and Presentations page.
 - Today's presentation is at the bottom of the page

Laura Goe, Ph.D.

- Former teacher in rural & urban schools
 - Special education (7th & 8th grade, Tunica, MS)
 - Language arts (7th grade, Memphis, TN)
- Graduate of UC Berkeley's Policy, Organizations, Measurement & Evaluation doctoral program
- Principal Investigator for the National Comprehensive Center for Teacher Quality
- Research Scientist in the Performance Research Group at ETS

The National Comprehensive Center for Teacher Quality

- A federally-funded partnership whose mission is to help states carry out the teacher quality mandates of ESEA
 - Vanderbilt University
 - Learning Point Associates, an affiliate of American Institutes for Research
 - Educational Testing Service

The goal of teacher evaluation

*The **ultimate** goal of all teacher evaluation should be...*

**TO IMPROVE
TEACHING AND
LEARNING**

Trends in teacher evaluation

- **Policy is way ahead of the research in teacher evaluation measures and models**
 - Though we don't yet know which model and combination of measures will identify effective teachers, many states and districts are compelled to move forward at a rapid pace
 - **Inclusion of student achievement growth data represents a huge “culture shift” in evaluation**
 - Communication and teacher/administrator participation and buy-in are crucial to ensure change
 - **The implementation challenges are enormous**
 - Few models exist for states and districts to adopt or adapt
 - Many districts have limited capacity to implement comprehensive systems, and states have limited resources to help them
-

The focus on teacher effectiveness is changing policy

- Impacting seniority and tenure rules
 - New legislation is changing “Last hired, first fired” policies in many states and cities, including Los Angeles, New York City, Washington, DC, Illinois, Florida, Colorado, Tennessee
- Impacting privacy and confidentiality
 - Los Angeles has already published teachers’ valued-added scores and New York City will likely follow suit

The stakes have changed

- Many of the current evaluation measures and models being used or considered have been around for years, *but the consequences are changing*
 - Austin's student learning objectives model could earn a teacher a monetary reward but could not get her fired
 - Tennessee's value-added results could be considered in teacher evaluation but poor TVAAS results did not necessarily lead to dismissal

How did we get here?

- Value-added research shows that teachers vary greatly in their contributions to student achievement (Rivkin, Hanushek, & Kain, 2005).
- The Widget Effect report (Weisberg et al., 2009) “...examines our pervasive and longstanding failure to recognize and respond to variations in the effectiveness of our teachers.” (from Executive Summary)

Definitions in the research & policy worlds

- Anderson (1991) stated that “... an effective teacher is one who quite consistently achieves goals which either directly or indirectly focus on the learning of their students” (p. 18).

Race to the Top definition of effective & highly effective teacher

Effective teacher: students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth (as defined in this notice). States, LEAs, or schools must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, by student growth (as defined in this notice). Supplemental measures may include, for example, multiple observation-based assessments of teacher performance. (pg 7)

Highly effective teacher students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth (as defined in this notice).

Measures and models: Definitions

- Measures are the instruments, assessments, protocols, rubrics, and tools that are used in determining teacher effectiveness
- Models are the state or district systems of teacher evaluation including all of the inputs and decision points (measures, instruments, processes, training, and scoring, etc.) that result in determinations about individual teachers' effectiveness

Multiple measures of teacher effectiveness

- **Evidence of *growth in student learning and competency***
 - Standardized tests, pre/post tests in untested subjects
 - Student performance (art, music, etc.)
 - Curriculum-based tests given in a standardized manner
 - Classroom-based tests such as DIBELS
 - **Evidence of *instructional quality***
 - Classroom observations
 - Lesson plans, assignments, and student work
 - Student surveys such as Harvard's Tripod
 - Evidence binder (next generation of portfolio)
 - **Evidence of *professional responsibility***
 - Administrator/supervisor reports, parent surveys
 - Teacher reflection and self-reports, records of contributions
-

Using multiple measures

- Lots of questions about multiple measures
 - What is the right combination of measures?
 - How do we “weight” measures?
 - Are student growth measures fair and valid for measuring teacher performance?
- Need more thinking around how to create systems that turn *evidence* from multiple measures into *strategies* for continuous improvement

Measures that help teachers grow

- Measures that motivate teachers to examine their own practice against specific standards
 - Measures that allow teachers to participate in or co-construct the evaluation (such as “evidence binders”)
 - Measures that give teachers opportunities to discuss the results with evaluators, administrators, colleagues, teacher learning communities, mentors, coaches, etc.
 - Measures that are directly and explicitly aligned with teaching standards
 - Measures that are aligned with professional development offerings
 - Measures which include protocols and processes that teachers can examine and comprehend
-

Keep in mind...

All teachers want to be effective, and supporting them to be effective is perhaps the most powerful talent management strategy we have

Considerations

- Consider whether human resources and capacity are sufficient to ensure fidelity of implementation
 - Poor implementation threatens validity of results
 - Establish a plan to evaluate measures to determine if they can effectively differentiate among teacher performance
 - Need to identify potential “widget effects” in measures
 - If measure is not differentiating among teachers, may be faulty training or poor implementation, not the measure itself
 - Examine correlations among results from measures
 - Evaluate processes and data each year and make needed adjustments
-

Validity of classroom observations is highly dependent on training

- Even with a terrific observation instrument, the results are meaningless if observers are not trained to agree on evidence and scoring
 - *A teacher should get the same score no matter who observes him*
 - This requires that all observers be trained on the instruments and processes
 - Occasional “calibrating” should be done; more often if there are discrepancies or new observers
 - Who the evaluators are matters less than that they are adequate trained and calibrated
 - Teachers should also be trained on the observation forms and processes to improve validity of results
-

Most popular growth models: Value-added and Colorado Growth Model

- EVAAS uses prior test scores to predict the next score for a student
 - Teachers' value-added is the difference between actual and predicted scores for a set of students
 - <http://www.sas.com/govedu/edu/k12/evaas/index.html>
- Colorado Growth model
 - Betebenner 2008: Focus on “growth to proficiency”
 - Measures students against “academic peers”
 - www.nciea.org

What nearly all state and district models have in common

- Value-added or Colorado Growth Model will be used for those teachers in tested grades and subjects (4-8 ELA & Math in most states)
- States want to increase the number of tested subjects and grades so that more teachers can be evaluated with growth models
- States are generally at a loss when it comes to measuring teachers' contribution to student growth in non-tested subjects and grades

Measuring teachers' contributions to student learning growth: A summary of current models

Model	Description
Student learning objectives	Teachers assess students at beginning of year and set objectives then assesses again at end of year; principal or designee works with teacher, determines success
Subject & grade alike team models (“Ask a Teacher”)	Teachers meet in grade-specific and/or subject-specific teams to consider and agree on appropriate measures that they will all use to determine their individual contributions to student learning growth
Pre-and post-tests model	Identify or create pre- and post-tests for every grade and subject
School-wide value-added	Teachers in tested subjects & grades receive their own value-added score; <i>all other teachers get the school-wide average</i>

SLOs + “Ask a Teacher” (Hybrid model)

- Concerns about SLOs are 1) rigor, 2) comparability, and 3) administrator burden
- A “rigor rubric” helps with first concern
- Combining SLOs with aspects of the “Ask A Teacher” model will help with all 3 concerns
 - Teachers discuss and agree to use particular assessments and measures of student learning growth, ensuring great rigor and comparability
 - Teachers work together on aspects of scoring which improves validity and comparability and lightens the administrator burden

What's next for Hawaii?



Next steps

- Ensure that evaluation systems allow you to differentiate between effective and less effective teachers
- Focus on improving effectiveness of teachers you already have
- Develop strategies for retaining effective and potentially effective teachers
- Recruit effective teachers through multiple, coordinated strategies (not one time bonuses)

Final thoughts

- The limitations:
 - There are no perfect measures
 - There are no perfect models
 - Changing the culture of evaluation is hard work
- The opportunities:
 - Evidence can be used to trigger support for struggling teachers and acknowledge effective ones
 - Multiple sources of evidence can provide powerful information to improve teaching and learning
 - Evidence is more valid than “judgment” and provides better information for teachers to improve practice

Evaluation System Models

Austin (Student learning objectives with pay-for-performance, group and individual SLOs assess with comprehensive rubric)

<http://archive.austinisd.org/inside/initiatives/compensation/slos.phtml> **Delaware**

Model (Teacher participation in identifying grade/subject measures which then must be approved by state)

http://www.doe.k12.de.us/csa/dpasii/student_growth/default.shtml

Georgia CLASS Keys (Comprehensive rubric, includes student achievement—see last few pages)

System: http://www.gadoe.org/tss_teacher.aspx

Rubric:

<http://www.gadoe.org/DMGetDocument.aspx/CK%20Standards%2010-18-2010.pdf?p=6CC6799F8C1371F6B59CF81E4ECD54E63F615CF1D9441A92E28BFA2A0AB27E3E&Type=D>

Hillsborough, Florida (Creating assessments/tests for all subjects)

<http://communication.sdhc.k12.fl.us/empoweringteachers/>

Evaluation System Models (cont'd)

New Haven, CT (SLO model with strong teacher development component and matrix scoring; see Teacher Evaluation & Development System)

<http://www.nhps.net/scc/index>

Rhode Island DOE Model (Student learning objectives combined with teacher observations and professionalism)

http://www.ride.ri.gov/assessment/DOCS/Asst.Sups_CurriculumDir.Network/Asst_Sup_August_24_rev.ppt

Teacher Advancement Program (TAP) (Value-added for tested grades only, no info on other subjects/grades, multiple observations for all teachers)

<http://www.tapsystem.org/>

Washington DC IMPACT Guidebooks (Variation in how groups of teachers are measured—50% standardized tests for some groups, 10% other assessments for non-tested subjects and grades)

[http://www.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+\(Performance+Assessment\)/IMPACT+Guidebooks](http://www.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+(Performance+Assessment)/IMPACT+Guidebooks)

References

Betebenner, D. W. (2008). *A primer on student growth percentiles*. Dover, NH: National Center for the Improvement of Educational Assessment (NCIEA).

<http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudentgrowthpercentiles.pdf>

Braun, H., Chudowsky, N., & Koenig, J. A. (2010). *Getting value out of value-added: Report of a workshop*. Washington, DC: National Academies Press.

http://www.nap.edu/catalog.php?record_id=12820

Finn, Chester. (July 12, 2010). *Blog response to topic "Defining Effective Teachers."* National Journal Expert Blogs: Education.

<http://education.nationaljournal.com/2010/07/defining-effective-teachers.php>

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011). *Passing muster: Evaluating evaluation systems*. Washington, DC: Brown Center on Education Policy at Brookings.

http://www.brookings.edu/reports/2011/0426_evaluating_teachers.aspx#

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy at Brookings.

http://www.brookings.edu/reports/2010/1117_evaluating_teachers.aspx

References (continued)

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

<http://www.tqsource.org/publications/LinkBetweenTQandStudentOutcomes.pdf>

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

<http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>

Hassel, B. (Oct 30, 2009). *How should states define teacher effectiveness?* Presentation at the Center for American Progress, Washington, DC.

<http://www.publicimpact.com/component/content/article/70-evaluate-teacher-leader-performance/210-how-should-states-define-teacher-effectiveness>

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). *Ready to learn? Children's pre-academic achievement in pre-kindergarten programs*. *Early Childhood Research Quarterly*, 23(1), 27-50.

<http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ783140>

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data*. Cambridge, MA: National Bureau of Economic Research.

<http://www.nber.org/papers/w15803>

References (continued)

Koedel, C., & Betts, J. R. (2009). *Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique*. Cambridge, MA: National Bureau of Economic Research.

http://economics.missouri.edu/working-papers/2009/WP0902_koedel.pdf McCaffrey, D., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). *The intertemporal stability of teacher effect estimates*. *Education Finance and Policy*, 4(4), 572-606.

<http://www.mitpressjournals.org/doi/abs/10.1162/edfp.2009.4.4.572>

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. [Education Forum]. *Science*, 315, 1795-1796.

<http://www.sciencemag.org/cgi/content/summary/315/5820/1795>

Prince, C. D., Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. T., & Thorn, C. A. (2006). *The other 69 percent: Fairly rewarding the performance of teachers of non-tested subjects and grades*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education.

<http://www.cecr.ed.gov/guides/other69Percent.pdf>

Race to the Top Application

<http://www2.ed.gov/programs/racetothetop/resources.html>

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417 - 458.

<http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf>.....

References (continued)

Sartain, L., Stoelinga, S. R., & Krone, E. (2010). *Rethinking teacher evaluation: Findings from the first year of the Excellence in Teacher Project in Chicago public schools*. Chicago, IL: Consortium on Chicago Public Schools Research at the University of Chicago.

<http://ccsr.uchicago.edu/publications/Teacher%20Eval%20Final.pdf>

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

<http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf>

Redding, S., Langdon, J., Meyer, J., & Sheley, P. (2004). *The effects of comprehensive parent engagement on student learning outcomes*. Paper presented at the American Educational Research Association

<http://www.adi.org/solidfoundation/resources/Harvard.pdf>

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.

<http://widgeteffect.org/downloads/TheWidgetEffect.pdf>

References (continued)

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement (No. REL 2007-No. 033)*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf

Questions?





NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Laura Goe, Ph.D.

609-734-1076

lgoe@ets.org

**National Comprehensive Center for
Teacher Quality**

1100 17th Street NW, Suite 500

Washington, DC 20036-4632

877-322-8700 > www.tqsource.org



NATIONAL COMPREHENSIVE CENTER
FOR TEACHER QUALITY

A Practical Guide to Designing Comprehensive Teacher Evaluation Systems

A Tool to Assist in the
Development of Teacher
Evaluation Systems

.....
MAY 2011



A Practical Guide to Designing Comprehensive Teacher Evaluation Systems

Laura Goe, Ph.D.

ETS

Lynn Holdheide

Vanderbilt University

Tricia Miller, Ph.D.

American Institutes for Research

Contents

- Rationale and Structure 1
- Introduction 2
- State Accountability and District Responsibility in Teacher Evaluation Systems 3
 - Key State Roles 4
 - Models for State and District Evaluation Systems 5
 - Factors for Stakeholder Consideration. 8
- Development and Implementation of Comprehensive Teacher Evaluation Systems 9
 - Component 1: Specifying Evaluation System Goals 9
 - Component 2: Securing and Sustaining Stakeholder Investment and
Cultivating a Strategic Communication Plan. 14
 - Component 3: Selecting Measures 19
 - Component 4: Determining the Structure of the Evaluation System. 34
 - Component 5: Selecting and Training Evaluators 37
 - Component 6: Ensuring Data Integrity and Transparency. 40
 - Component 7: Using Teacher Evaluation Results 43
 - Component 8: Evaluating the System 47
- Conclusion and Recommendations 50
- References 51
- Appendix A. Summary of Measures 53

Rationale and Structure

Across the nation, states and districts are in the process of building better teacher evaluation systems that not only identify highly effective teachers but also systematically provide data and feedback that can be used to improve teacher practice. *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems* is a tool designed to assist states and districts in constructing high-quality teacher evaluation systems in an effort to improve teaching and learning.

This tool is not a step-by-step guide to devising a teacher evaluation system. Rather, it is intended to facilitate discussion and promote coherence in the development process. The following assumptions have guided its construction:

- In response to federal initiatives and priorities as well as state legislation, states are motivated to improve their current evaluation systems to better identify successful teachers, assist less successful teachers, and help all teachers improve their practice.
- Most current definitions of teacher effectiveness (e.g., the Race to the Top definition) include teachers' contributions to student learning growth, and states need to consider measuring these contributions for all teachers.

- States are interested in systems that use multiple measures to assess various aspects of teachers' performance and instructional practice.
- States may be in various stages in terms of creating or revising teacher evaluation systems. This tool allows states to focus on the specific components of the system that are most relevant for them.
- In states where districts have substantial control over teacher evaluation systems, this tool may be used by districts or consortiums of districts for discussion and guidance.
- Teachers play a critical role in ensuring that the evaluation system is fair, valid, and successful, and they should be active participants in designing, developing, implementing, and evaluating the system.

The guide begins with an overview of the factors influencing teacher evaluation reform today and continues with a discussion of approaches to balancing state accountability and district autonomy. The next section of the guide is structured around the following essential components of the design process as supported through research:

- Component 1: Specifying Evaluation System Goals
- Component 2: Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan

- Component 3: Selecting Measures
- Component 4: Determining the Structure of the Evaluation System
- Component 5: Selecting and Training Evaluators
- Component 6: Ensuring Data Integrity and Transparency
- Component 7: Using Teacher Evaluation Results
- Component 8: Evaluating the System

Each subsection includes an overview of the component, resources and practical examples, and a series of guiding questions designed to help states organize their work and move strategically toward an evaluation system that functions to improve student learning and teacher performance.

Introduction

The research community has long recognized the importance of teachers to student achievement. Although research has shown that teachers are the most significant school-based factor in student achievement, traditional methods of evaluating teachers have not been able to capture or explain differences between effective and ineffective teachers.

Initial efforts to ensure quality education focused on teacher qualifications and degrees; however, research does not indicate that these factors significantly influence teacher effectiveness. For example, Rivkin, Hanushek, and Kain (2005) analyzed results from thousands of teachers and their students in Texas and determined that there were strong teacher effects on academic achievement, but variation in these effects could not be explained by education or experience.

Further, mounting evidence indicates that the United States is losing ground in comparison to other countries in terms of educational outcomes. One international study showed that U.S. students were outperformed in mathematics by students in 20 of the other 28 industrialized countries studied (Lemke et al., 2004). In addition, a recent Program for International Student Assessment study found that only 5 of the

other 33 participating countries had lower scores in mathematics literacy than the United States (Fleischman, Hopstock, Pelczar, & Shelley, 2010). These types of findings resulted in increased concern about determining the best way to improve student learning through teacher performance and a shift in focus from analyzing teacher inputs (e.g., education, certification, and experience) to measuring teacher effects (e.g., student achievement and classroom practice).

Improving teacher quality and effectiveness is a complex issue, and the ability to identify high-performing and low-performing teachers is a necessary step toward pinpointing instructional strategies and pedagogy that result in improved student growth (e.g., evidence-based instructional strategies, strong student-teacher relationships). Unfortunately, traditional evaluation methods have not proven to be useful in meeting this challenge. In the past, teacher evaluation systems have varied widely in their rigor and utility. Most systems were based on classroom observations, usually conducted by principals but sometimes conducted by trained evaluators (see Practical Example: “Cincinnati Public Schools Evaluation System”). The steps taken after the observations differed considerably across states, districts, and even schools, with some schools linking results to professional growth plans for teachers and others filing the results away

with little or no follow-up. The perfunctory, compliance-oriented approach to teacher evaluation in some districts likely did not contribute to tangible improvement in teaching and learning. Unfortunately, there has been little research on how these different approaches to classroom observation influenced teacher performance.

PRACTICAL EXAMPLE

Cincinnati Public Schools Evaluation System

Cincinnati teachers participate in a “comprehensive evaluation” during their first and fourth years of teaching, after which, they are evaluated every five years. Teachers are observed four times by teacher evaluators and once by a school administrator. Before they can become teacher evaluators, teachers must complete a three-step application process to become lead teachers. Lead teachers may then apply for positions such as teacher evaluators, consulting teachers, and program facilitators. Those selected for teacher evaluator positions are required to undergo extensive training in collecting and scoring evidence. Using videos, they are certified through a process of verifying the agreement of their scores with those of “master raters.” Through this process, a high degree of reliability is ensured, meaning that a teacher’s observation score is likely to be the same or nearly the same, regardless of which trained evaluator conducts the observation.

Source: Cincinnati Public Schools (n.d.)

In 2009, an investigation into the compliance-oriented approaches of evaluation systems conducted by The New Teacher Project sent shockwaves through the policy world. The study, titled *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, examined large and small districts across several states where evaluation consisted primarily of classroom observations (Weisberg, Sexton, Mulhern, & Keeling, 2009). The following conclusions emerged from the study:

- Nearly all teachers received high ratings (*good* or *great*).
- Districts failed to recognize and reward excellence.
- Professional development was rarely tied to results and when it was, little support was offered to teachers.
- New teachers generally were rated above satisfactory, and tenure was seldom denied to teachers based on observation results.
- Poor performance rarely led to teacher dismissal.

The inability of evaluation systems to differentiate factors contributing to teacher effectiveness suggests that classroom observations, at least as they were used in most districts in the study, are of little use for improving and rewarding performance or identifying teachers who need support and training and those who should be dismissed.

Through funding opportunities including the American Recovery and Reinvestment Act (ARRA) and the Race to the Top competition, the federal government has encouraged states and districts to develop rigorous evaluation systems for use in high-stakes decisions including teacher advancement, compensation, distribution, and retention. These opportunities, coupled with the evidence of poorly functioning teacher evaluation systems, have resulted in a national urgency to create and implement comprehensive, strategic systems for evaluating teacher performance that identify, support, and develop teacher effectiveness and student growth.

In response to this urgency, many states have passed legislation mandating the development of rigorous, high-quality evaluation systems for use in high-stakes situations related to teacher employment and advancement. Advisory boards, committees, and multistate consortia are meeting to gather information on research and best practices related to the development, implementation, and use of these evaluation systems. This Practical Guide provides education policymakers and stakeholders with guidance on the key areas that should be addressed during the development and implementation of a new evaluation system.

State Accountability and District Responsibility in Teacher Evaluation Systems

Until recently, teacher evaluation has largely been considered the purview of districts or schools without much, if any, involvement from states. Starting with the highly qualified teacher requirements as codified in the Elementary and Secondary Education Act (ESEA), as reauthorized by the No Child Left Behind Act (NCLB), and continuing through the ARRA reform goals and assurances, states are expected to play an increasingly larger role in ensuring the quality and effectiveness of the nation's teaching force. This expectation creates a challenge for many states with a long history of local autonomy in most education matters. Specific to teacher evaluation systems, states now must decide the extent to which the teacher evaluation model will make allowances for local flexibility and provide a balance between local and state control that encourages collective responsibility and accountability. This section includes an overview of several key roles states may play in assisting districts in the implementation of new evaluation requirements and descriptions of several models that balance state accountability with local autonomy.

Key State Roles

Interpreting Federal and State Regulations

Spurred on by the Race to the Top competition, many states have either recently passed new legislation or pointed to existing legislation concerning teacher evaluation, most of which is directly related to the four ARRA reform goals or assurances: the quality of standards and assessments, improving the collection and use of data, increasing teacher effectiveness and equitable distribution, and supporting struggling schools (Learning Point Associates, 2010). The language of this federal and state legislation permits varying degrees of flexibility in terms of how the evaluation system is to be implemented. As such, the responsibility for interpretation and implementation falls primarily to states. Implementing new teacher evaluation laws and policies usually involves interpretation and overcoming challenges that may not have been anticipated by the policymakers. Recognition of these challenges and their potential variations according to local contexts should inform the training needs of personnel and contribute to the development of the evaluation model. Accordingly, states need to take proactive steps in helping districts interpret the new legislation and determining the best course of action toward full implementation.

For example, the Race to the Top application indicated that student achievement was to be a “significant” component of teacher evaluation. However, the federal government did not define *significant*, and there is not currently a research base to support

differential weighting of the components in an evaluation system. As a result, many states looked to their own legislation to resolve any discrepancies in interpretation or implementation.



TQ CENTER RESOURCES

Approaches to Evaluating Teacher Effectiveness: A Research Synthesis
<http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>

This research synthesis examines how teacher effectiveness is measured and provides practical guidance for evaluating teacher effectiveness. It evaluates the research on teacher effectiveness and various measures. In addition, it defines components and indicators that characterize effective teachers, extending this definition beyond teachers' contributions to student achievement gains to include how teachers affect classrooms, schools, and colleagues as well as how teachers contribute to other important outcomes for students.

Methods of Evaluating Teacher Effectiveness (Research-to-Practice Brief)
http://www.tqsource.org/publications/RestoPractice_EvaluatingTeacherEffectiveness.pdf

This brief is intended to help regional comprehensive centers and state policymakers as they consider evaluation methods to clarify policy, develop new strategies, identify effective teachers, or guide and support districts in selecting and using appropriate evaluation methods for various purposes. Included in this brief is a five-point definition of teacher effectiveness the authors developed by analyzing research, policy, and standards that address teacher effectiveness and by consulting experts in the field.

A Practical Guide to Evaluating Teacher Effectiveness
<http://www.tqsource.org/publications/practicalGuide.pdf>

This guide offers a definition of teacher effectiveness that states and districts may adapt to meet local requirements, provides an overview of the many purposes for evaluating teacher effectiveness, and indicates which measures are most suitable to use under different circumstances. The guide also includes summaries of various measures, such as value-added models, classroom observations, analysis of classroom artifacts, and portfolios. The summaries include descriptions of the measures, along with a note about the research base and strengths and cautions to consider for each measure.

For example, several states codified the weight (percentage) of student achievement in the teacher evaluation system (e.g., Tennessee specified 50 percent, Rhode Island specified 51 percent, and Colorado specified 50 percent). Such state legislation was intended to drive changes in evaluation systems and provide better information about teachers' contributions to student learning growth. However, the legislation often did not address the other logistical and procedural aspects of teacher evaluation (e.g., how growth would be measured, the frequency of the evaluation, who would conduct the evaluations, and how evaluators would be trained). States should play a critical role in interpreting such legislation and be prepared to help districts address specific challenges, unintended consequences, and implementation considerations at the district level

Interpreting/Conducting Research

The dearth of available research-based methods and models of comprehensive teacher evaluation hinders states' abilities to offer assistance to districts. Although some research on the utility of specific measures of teacher performance exists (Goe, Bell, & Little, 2008), albeit limited, most has been conducted in low-stakes environments. Therefore, many states have chosen to assemble task forces and engage national experts in evaluation and

measurement to secure recommendations and inform the conversation concerning teacher effectiveness and evaluation policy.

In many cases, states and districts may need to identify measures and conduct research during and after implementation. Given potential resource and human capacity limitations at the district level, states may need to play an active role in conducting research to ensure that the evaluation model is technically sound and therefore defensible, especially in situations in which teacher evaluation results will be used to make personnel and compensation decisions. These conversations and preliminary research could be instrumental in ensuring the validity of the results from comprehensive teacher evaluation systems and gaining educator and stakeholder support.

Models for State and District Evaluation Systems

Historically, models of teacher evaluation varied among schools, districts, and states and were largely dependent on the context in which the model evolved and was implemented. However, as federal guidance and policy lean toward more state responsibility for ensuring teacher quality and monitoring district teacher evaluation, states must determine their role and level of involvement—from providing limited guidance to taking a more directive approach.

For example, some states may elect to mandate a particular evaluation model, governing logistics (e.g., how often teachers are evaluated), format (e.g., selection of measures), and personnel decisions (e.g., what a rating means in terms of teacher tenure). Others may provide specific guidelines for the evaluation model while allowing the district flexibility in adapting those guidelines locally and in the implementation of the system. The level of flexibility will likely vary according to many factors (e.g., political context, local bargaining agreements, state size, and district capacity) and the state's goals. Some states, like Delaware, have mandated a statewide evaluation system. Other states allow every district to determine its own model for teacher evaluation as long as

stated requirements are met. States also may create or facilitate consortiums among districts in the same region or those that share similar challenges (e.g., a rural consortium encompassing several contiguous districts or a statewide consortium of urban districts).

Various state options and accompanying stakeholder considerations are discussed in the following subsections. Note that this is not an exhaustive listing of options.

State-Level Evaluation System

Within a state-level evaluation system, the state provides a strict interpretation of legislation and prescribes the requirements for the teacher evaluation model. The state determines the components of the teacher evaluation model, which measures are to be used, how often evaluations are to be conducted, and by whom. Therefore, the state is instrumental in the design, implementation, and evaluation of the teacher evaluation model. Delaware is currently in the process of implementing this type of system (See Practical Example: “Delaware’s Evaluation System”). With significant contribution from local practitioners, the state has led the efforts related to the development of a comprehensive teacher evaluation model.

After the model is finalized, all Delaware districts will be required to implement the model with little flexibility.

Elective State-Level Evaluation System

Within an elective state-level evaluation system, states may elect to provide a strict interpretation of state or federal legislation and dictate certain aspects of the evaluation model but allow flexibility in others. For example, the state may have legislation that mandates the use of student achievement as a significant factor, and district models would have to include measures of student achievement. Or the state may have specific language about which aspects of teacher evaluation are subject to local decision making and which aspects are state mandates that are not open to negotiation. The state may mandate the type of growth model and other measures the districts use to determine student growth, the attribution of growth to teachers, and the weight (percentage) of the components of the teacher evaluation system. The possible components of the evaluation model (e.g., observation protocols, portfolios, student/parent surveys) and processes for using them would be determined by the district. For instance, the state might offer Charlotte Danielson’s (2007) *Framework for Teaching*

as an option that districts could elect to use but allow districts to choose different observation models as long as certain criteria are maintained. Thus, the state plays a major role in ensuring that certain components are part of the district models but allows for local flexibility in other aspects of the system. This option allows a continuance of established district models, provides flexibility for bargaining agreements, and continues the tradition of local control over teacher evaluation that exists in many states (See Practical Example: “New York’s Evaluation System”).

District Evaluation System With Required Parameters

States that find it impractical to adopt a single statewide evaluation system may still deem it necessary to provide guidance to districts in implementing regulations and state priorities for teacher evaluation. Within a district evaluation system with required parameters, states play a much smaller role in the design and implementation of the teacher evaluation system at the district level. Guidance may be somewhat general, such as requiring states to implement an evaluation system that includes several components (e.g., observations, evidence of professional responsibilities, and evidence of teachers’ contributions to student achievement).

The guidance also may be more restrictive, particularly if some aspects of the evaluation system are already in place. In this case, the state provides some level of guidance to districts and specifies the parameters for the district models. The state, therefore, does not play a major role in the evaluation

process but provides some type of screening/approval to ensure district compliance in selecting models as well as an audit or follow-up mechanism to ensure that districts are working within the defined parameters. For instance, the state may indicate that districts can select their own

evaluation model but require that new teachers be observed three times per year for at least 20 minutes per visit. The district has flexibility in the model selection, but the logistical parameters need to be met (See Practical Example: “Ohio’s Guidelines to Evaluation”).

PRACTICAL EXAMPLES

Delaware’s Evaluation System

Delaware already had a statewide evaluation system in existence prior to being awarded Race to the Top funds. This system included classroom observation and opportunities for professional growth. However, Delaware’s existing system lacked a mechanism to measure student growth. Therefore, an external evaluation was conducted that included soliciting feedback from teachers and administrators through surveys, interviews, and focus groups. The state department collaborated with union representatives to ensure that the system would be accepted as comprehensive, valid, and fair. These results contributed to the design of a statewide model. However, given the timelines and implementation demands, it is not clear whether Delaware will ultimately use this model; alternatives are still being considered.

Adapted from *Measuring Teachers’ Contributions to Student Learning Growth for Nontested Grades and Subjects* by L. Goe and L. Holdheide. Copyright © 2011 National Comprehensive Center for Teacher Quality, p. 12

New York’s Evaluation System

New York’s system is an example of an elective state model, providing clear guidance about how new evaluation requirements will be phased in over several years. The system is based on a 100-point scale; 60 percent of the evaluation score will be based on locally negotiated processes (e.g., classroom observations by trained evaluators), and 40 percent will be based on a combination of state standardized tests and local assessments and measures, which will have to be developed by each school system.

Year 1: 20 percent of student growth is based on state assessments or comparable measures for teachers in the common branch subjects or ELA and mathematics in Grades 4–8 only, and 20 percent is based on other *locally selected* measures that are rigorous and comparable across classrooms.

Year 2: After two years, 25 percent will be based on standardized tests, and 15 percent will be based on locally selected measures.

Source: New York State Education Department (2011)

Ohio’s Guidelines to Evaluation

Ohio has developed state teacher evaluation guidelines as follows.

The teacher evaluation system adopted by the district should:

- “Align to the *Standards for Ohio Educators*.”
- “Be systematic and ongoing in order to promote professional development and student learning.”
- “Take into account experience, skill, longevity, and responsibility.”
- “Use a variety of measures to collect evidence.”
- “Include differentiated effectiveness levels of performance using multiple rating categories of teachers and encourage continuous professional growth.”

Note: Components of this system are being piloted this year. Next year, the guidelines will be piloted.

Source: (Ohio Department of Education, n.d., Slides 16–17)

Factors for Stakeholder Consideration

Stakeholders might consider the following factors in selecting a particular model:

- Grant requirements as applicable (e.g., Race to the Top, School Improvement Grants, Teacher Incentive Fund)
- Existing or impending state legislation that affects the evaluation process
- Goals and priorities at the state and district levels
- State-level role in district practice
- The role of unions and bargaining agreements in local and state decisions
- The number and diversity of districts within a state as well as geographical distance between them
- The human and resource capacity at the state and local levels
- The training needed to implement the system with fidelity
- Stakeholder support for changes in teacher evaluation
- Technological capacity, including the ability to link teachers with students
- District models already in use and their level of acceptance at the local level
- Teachers' and administrators' preferences for certain types of measures

Note: Race to the Top and ARRA indicate that total district-level control with no state-level involvement or accountability is not supported at the federal level.

Table 1 lists some potential strengths and weaknesses within the various models.

Table 1. Evaluation Model Strengths and Weaknesses

Model	Strengths	Weaknesses
State-Level Evaluation System	<ul style="list-style-type: none"> • Measures and dimensions are the same statewide. • Data collection can be standardized. • Districts can be directly compared. • Evaluating the system and results will be easier. • System is perceived as fair because all districts are held to the same standards. • There is increased system reliability because changes from year to year affect all districts. 	<ul style="list-style-type: none"> • Local flexibility and ownership is diminished. • The system fails to consider local context. • It is difficult to obtain statewide support. • There is variance in district resources. • The system may be subject to local bargaining agreements. • The system may be seen as unfair by low-capacity districts forced to implement the same model as districts with greater capacity. • Local variations in school year and testing times may result.
Elective State-Level Evaluation System	<ul style="list-style-type: none"> • The system allows for some local flexibility. • Data collection can still be standardized for certain components. • Districts can be directly compared in certain areas. • Reliability is strong in required components. • The system allows for continuance of locally developed models. 	<ul style="list-style-type: none"> • Local flexibility in certain areas is diminished. • The system presents more challenges for state oversight. • Data aggregation of teacher results may be more difficult.
District Evaluation System With Required Parameters	<ul style="list-style-type: none"> • Local ownership and buy-in is increased. • Districts have the ability to address local priorities within the model. • The system allows for continuance of locally developed models. 	<ul style="list-style-type: none"> • It is difficult to compare progress/results. • Data aggregation may present considerable challenges. • Reliability is vulnerable across districts. • Training to ensure fidelity would likely be conducted at the district level, meaning more district resources are required. • Resources may be limited.

Development and Implementation of Comprehensive Teacher Evaluation Systems

This section is divided into eight subsections describing the critical components of designing a comprehensive teacher evaluation system:

- Component 1: Specifying Evaluation System Goals
- Component 2: Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan
- Component 3: Selecting Measures
- Component 4: Determining the Structure of the Evaluation System
- Component 5: Selecting and Training Evaluators
- Component 6: Ensuring Data Integrity and Transparency
- Component 7: Using Teacher Evaluation Results
- Component 8: Evaluating the System

Each subsection discusses the relevance of the component in the design process and concludes with a series of questions to guide the development process.

COMPONENT 1

Specifying Evaluation System Goals

Goal-setting is an imperative, and often challenging, first step in designing a teacher evaluation system. The establishment of explicit, well-defined goals lays the foundation for a comprehensive, sustainable evaluation system. Some states have defined teacher evaluation goals and purposes in recent legislation and/or policy. In most scenarios, however, stakeholders are left to define effective teaching and achieve consensus on the evaluation system's purpose. There is a general tendency to oversimplify this crucial step; however, agreement about goal selection focuses and guides all decisions throughout the design process. The methods and weighting used for various components of the resulting system and any actions informed by evaluation results (e.g., professional development targeted to a challenge area) should reflect the evaluation system goals.

Stakeholders should exercise caution when selecting goals, keeping in mind that the ultimate objective of teacher evaluation is to improve teaching and learning. Systems designed exclusively for accountability are less likely to have an impact on teacher practice than those tied to professional learning opportunities and growth.

At the same time, if a goal of the teacher evaluation system is to make personnel and compensation decisions, there is an increased need to ensure that measures are technically defensible. The higher the stakes, the greater the need to establish reliable and valid measures that can accurately differentiate among more and less effective teachers. Likewise, if the goal of the evaluation system is to improve teacher practice, ensuring a link to professional learning within the evaluation cycle is crucial.

Reviewing current state and district initiatives is another important step of the goal selection process. Gaining clarity in state and district reform initiatives enables consistency among programs and prevents fragmentation in which human resource capacity and professional development decisions are made in isolation. Integrating and embedding the evaluation system goals into large state and district reform initiatives will facilitate coherence and strengthen the system's credibility and implementation.

Stakeholders might consider the guiding questions for Component 1 as they work to develop the overall vision and goal of the evaluation system.

Specifying Evaluation System Goals

		NOTES
SYSTEM GOALS AND PURPOSES 1. Have the goals and purposes of the evaluation system been determined?	GUIDING QUESTIONS <ul style="list-style-type: none">What type of impact do stakeholders hope to achieve (e.g., better teacher retention, improved student test scores, increased teacher capacity)?Will teacher evaluation results be used for personnel and compensation decisions?Will teacher evaluation results be used to improve teacher practice?Will teachers be held accountable for student academic growth?What type of reform efforts are most important to the teachers union? (if applicable)Will incentives be offered to teachers according to performance?Will support be available for teachers identified in need?What financial and human capital resources are available?Are state teacher performance standards established?	
GOAL DEFINITION 2. Are the goals explicit, well-defined, and clearly articulated for stakeholders?	GUIDING QUESTIONS <ul style="list-style-type: none">Are the goals stated in measurable terms?Can a model of teacher evaluation conceivably meet these goals?Do all the training and explanatory materials portray a consistent message?	
GOAL ALIGNMENT 3. Have the evaluation system goals been aligned to the state strategic plan or other teacher reform initiatives?	GUIDING QUESTIONS <ul style="list-style-type: none">Are there other teacher quality initiatives occurring within the state?How will the efforts in teacher evaluation affect other quality initiatives (e.g., curriculum, professional learning, certification)?How can reform efforts be aligned to create a coherent system?Is there flexibility for district input/alignment with district initiatives?	

Establishing Standards

After the goals and purposes of an evaluation system are determined, the state or district needs to ensure alignment between these goals and teacher standards. This task often begins with defining the term *effective teacher*, then breaking that definition into teacher standards, competencies, and achievement-related outcomes (See “Defining Teacher Effectiveness”). Most states already have teacher standards in place, for use in hiring and traditional evaluation processes. However, as outlined previously, Race to the Top requirements and potential forthcoming mandates in the reauthorization of ESEA demand evaluation systems with the capacity to determine teacher effectiveness through measures of teacher performance and student growth. It is important, therefore, that teacher standards not only define what is valued in a teacher but indicate

knowledge, skills, or practices that can be measured reliably, correlated to student growth, and aligned with professional learning opportunities. Finally, the standards or criteria should include concise descriptions so that the broad statements are clearly articulated in a meaningful way at the practitioner level for shared understanding. These standards will form the basis from which definitions of desired behaviors can be created—the rating scale for the evaluation system.

Many states have referred to the National Board for Professional Teaching Standards (2011) for reference in standard development. In addition, the Council of Chief State School Officers Interstate Teacher Assessment and Support Consortium (InTASC) released its new *Model Core Teaching Standards: A Resource for State Dialogue* in April 2011. These standards are an update of the 1992

InTASC Standards that were primarily designed for licensing new teachers (See Council of Chief State School Officers, 2011). The professional practice standards can be used for all stages of a teacher’s career. Both sets of standards have either been adopted or used as the basis for standards development.

State stakeholders might consider the guiding questions on p. 13 as they develop or revise teacher standards.

DEFINING TEACHER EFFECTIVENESS

TQ Center Definition

- “Effective teachers have high expectations for all students and help students learn, as measured by value-added or other test-based growth measures, or by alternative measures.
- “Effective teachers contribute to positive academic, attitudinal, and social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior.
- “Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.
- “Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.
- “Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure” (Goe et al., 2008, p. 8).

Federal Definition

“*Effective teacher* means a teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth (as defined in this notice). A method for determining if a teacher is effective must include multiple measures, and effectiveness must be evaluated, in significant part, on the basis of student growth (as defined in this notice). Supplemental measures may include, for example, high school graduation rates (as defined in this notice) and college enrollment rates, as well as evidence of providing supportive teaching and learning conditions, strong instructional leadership, and positive family and community engagement” (Secretary’s Priorities for Discretionary Grant Priorities, 2010, p. 47288).

“Student growth means the change in student achievement (as defined in this notice) for an individual student between two or more points in time. A State may also include other measures that are rigorous and comparable across classrooms” (Secretary’s Priorities for Discretionary Grant Priorities, 2010, p. 47290).

“*Student achievement* means—

“(a) *For tested grades and subjects*: (1) A student’s score on the State’s assessments under the ESEA; and, as appropriate, (2) other measures of student learning, such as those described in paragraph (b) of this definition, provided they are rigorous and comparable across schools.

“(b) *For non-tested grades and subjects*: Alternative measures of student learning and performance, such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments; and other measures of student achievement that are rigorous and comparable across schools” (Secretary’s Priorities for Discretionary Grant Priorities, 2010, p. 47290).

Guiding Questions

Establishing Standards

DEFINITION OF EFFECTIVE TEACHER

1. Has the state defined what constitutes an effective teacher?

GUIDING QUESTIONS

- Will the state or district go beyond a teacher's ability to improve student learning in its definition of an effective teacher?
- Will the use of evidence-based teaching practices be a factor in identifying an effective teacher?
- Will behavioral and social outcomes be a factor in identifying an effective teacher?
- Will effective collaboration be a contributing factor in identifying an effective teacher?
- Will a teacher's professionalism be a factor in identifying an effective teacher?
- What characteristics, behaviors, and values should a highly effective teacher demonstrate?
- What type of classroom environment should a teacher create in his or her classroom?
- Should a highly effective teacher demonstrate leadership? If so, what might that look like?
- What content knowledge do the teachers need to translate to their students?

NOTES

TEACHING STANDARDS

2. Has the state established teaching standards?

GUIDING QUESTIONS

- Are there existing state teaching standards that can be used to guide system development?
- Are the standards written in a manner that reflects measures of teacher performance and student growth?
- Do the standards explicitly define desired teaching competencies?
- Have levels of teaching performance been established within the standards?
- Have the standards been written in a manner in which evaluation system results will yield reliable information on teacher performance according to the identified standards?
- Have sample performance indicators been developed to provide examples of observable behavior?
- Was public comment a step in developing teaching standards?

COMPONENT 2

Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan

Stakeholder Investment

Evaluation systems are much more likely to be accepted, successfully implemented, and sustained if stakeholders are included in the design process. Stakeholder involvement throughout the design, implementation, assessment, and revision of teacher evaluation systems increases the likelihood that the system is perceived as responsive, useful, and fair. Potential stakeholder representation could include the following:

- Teachers (including various levels, content areas, and specialists)
- Teacher union representatives
- Related services personnel
- School board members
- Superintendents
- School principals
- Teacher preparation programs, parents
- Students
- Business and community leaders

Involving teachers in the initial stages of development and throughout the implementation process will likely increase

teachers' collaboration, support, and promotion of state and district efforts. Teachers are in the best position to inform this process, as they can discern what will work in their classrooms.

Clarifying expectations in terms of stakeholder purpose, level and duration of commitment, and authority in decisions will assist in sustaining stakeholder investment throughout the process. Individual members bring to bear unique sets of skills, experiences, and interests, and the level of involvement of each stakeholder may shift during the process of designing and implementing the evaluation system. Defining stakeholders' roles and responsibilities, while capitalizing on their expertise, may cultivate a high level of active participation. Stakeholders could play an integral role in the following tasks:

- Determining the standards and criteria for the system
- Mobilizing administrator, teacher, and community support
- Facilitating practitioner groups to obtain input and feedback
- Marketing the system and publicizing the findings
- Interpreting policy implications
- Investigating and/or securing federal, state, or private sector funding



TQ CENTER RESOURCE

Communication Framework for Measuring Teacher Quality and Effectiveness: Bringing Coherence to the Conversation
(<http://www.tqsource.org/publications/NCCTQCommFramework.pdf>)

This framework can be used by regional comprehensive center staff, state education agency personnel, and local education agency personnel to promote effective dialogue about the measurement of teacher quality and effectiveness. The framework consists of the following four components: communication planning, goals clarification, teacher quality terms, and measurement tools and resources.

Communication Plan

Early on in the process, stakeholders should consider communication needs. A strategic communication plan detailing steps to inform the broader school community about implementation efforts, results, and future plans may increase the potential for statewide adoption. Misperceptions and opposition can be minimized if the state and districts communicate a clear and consistent message. A strategic communication plan first identifies the essential messages and audiences. Potential key audiences could include pilot participants, school personnel, families, and the external community.

Stakeholders would then determine the most effective channel of communication for its purpose and target audience. Written, spoken, and/or electronic communication strategies may include the following:

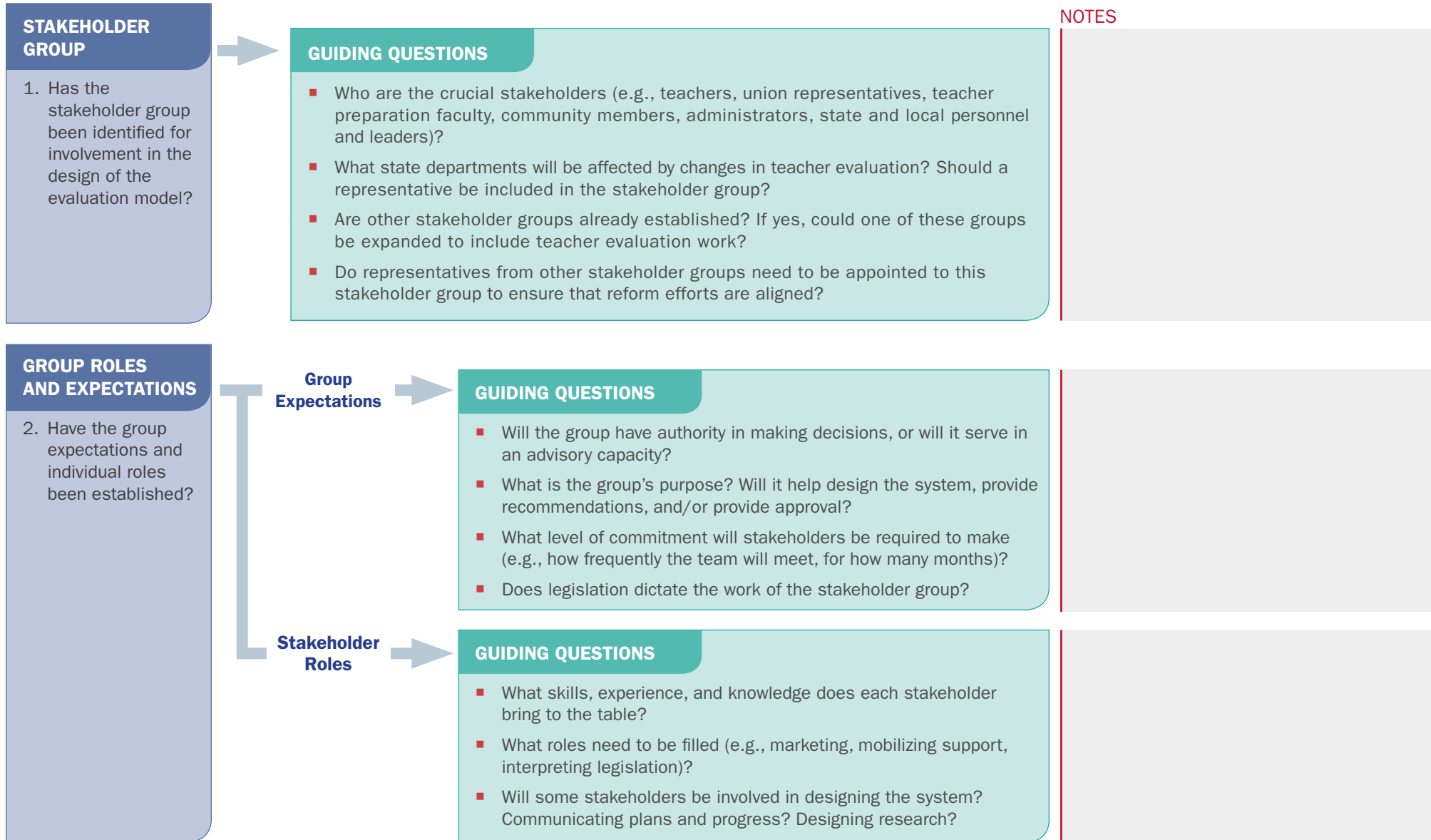
- Online communications
- Community information nights
- Quarterly memos
- Weekly e-mail updates
- Media relations materials
- Word of mouth
- Events
- Workshops
- Videos
- CDs
- Press releases
- Newsletters

Communication plans should take into account the duration of the process of improving the evaluation system including its initiation and all implementation phases. For example, communication needs during the design of the system will be different from those during implementation and the process of gathering feedback. Plans should include updates on efforts to build the evaluation system, celebrations of successes as the work moves forward, and recognition of stakeholder contributions. Communicating success in terms of implementation efforts, changes in teacher practice, and student outcomes can be a powerful way to ensure buy-in and secure stakeholder investment. Highlighting successes also reinforces, inspires, and energizes teachers.

Stakeholders might consider the guiding questions for Component 2 as they develop a strategic communication plan.

Guiding Questions

Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan



COMMUNICATION PLAN

3. Does the group have a strategic communication plan to keep the broader school community informed?

Content

GUIDING QUESTIONS

- What information needs to be communicated to stakeholders?
- Will pilot results be communicated?
- Will progress on the design, implementation, and success of the evaluation system be shared?
- Will teacher evaluation results be reported?

Target Audience

GUIDING QUESTIONS

- Which stakeholders should be kept informed about the development, implementation, and results of efforts related to teacher evaluation?
- Who will be the target audience (e.g., pilot participants, teachers, administrators, students, parents, community)?
- Will communication efforts be varied according to audience (e.g., board members require more detailed updates than community members)?
- How will personnel outside of the stakeholder group be kept informed?

Mode

GUIDING QUESTIONS

- Do channels of communication with stakeholders already exist?
- Does the state have a public communications department that could assist in marketing?
- What forms of communication will be utilized (e.g., website, e-mails, newsletters, public announcements)?

Timing

GUIDING QUESTIONS

- Does the plan include communication strategies throughout the development process (e.g., in the beginning, during, and after each phase)?
- Has the plan considered optimal timing for communicating evaluation efforts and results?

FEEDBACK

4. Has the stakeholder group determined a process to ensure that constituent feedback is integrated into the systems' redesign efforts?

Who

GUIDING QUESTIONS

- From whom does the group wish to solicit feedback (e.g., pilot participants, teachers, legislators, administrators, parents)?

Methods

GUIDING QUESTIONS

- What methods will the state use to obtain feedback from affected school personnel during the design process (e.g., surveys, focus groups)?
- Are there teacher groups or electronic mailing lists that could be accessed to obtain stakeholder feedback?
- Are there teachers of certain student populations and content areas in which focus groups should be considered?
- Has the group considered an internal or external evaluation to determine the effectiveness of the system (from a teacher/principal perspective) during implementation?

Response

GUIDING QUESTIONS

- Who will consolidate the stakeholder feedback? How will it be incorporated into the redesign process?
- How will the group respond to stakeholder feedback (e.g., Q&A document, FAQ newsletter)?
- What weight will constituent feedback hold?
- Will student outcomes be considered before changes are considered?

NOTES

COMPONENT 3

Selecting Measures

The evaluation system's purpose and teacher standards should inform the types of outcomes and practices that will be assessed through the evaluation system, which in turn, will inform the methods and measures to be used. Selecting appropriate measures is a critical component of the design process. Measures should yield reliable information on whether teaching standards have been demonstrated and evaluation system goals have been realized.

Current federal definitions of teacher effectiveness have focused strongly on student growth. This focus was made clear in the Race to the Top competition, which required states to develop evaluation systems that “differentiate effectiveness using multiple rating categories that take into account data on student growth . . . as a significant factor” (Race to the Top application, D[2][ii], p. 34). Race to the Top guidance also indicates that multiple measures of evaluating teacher performance should be used, a belief that is echoed by the research and policy communities. Multiple measures of teacher outcomes allow for a more comprehensive view of a teacher's effectiveness based on a variety of evidence. Although summative student achievement data are relevant, data on teacher performance are most useful

for targeting professional development and specifically addressing areas in which growth is needed.

According to Goe and Holdheide (2011), multiple measures:

- Strengthen teacher evaluation.
 - Provide a more complete picture of teachers' contributions to student learning.
 - Contribute to greater confidence in the results of teacher evaluations.
 - Provide more information about collaboration for student success.
- Contribute to teachers' professional growth.
 - Create opportunities for teachers to learn from their colleagues.
 - Provide teachers with greater insights into how their instruction affects student learning.
- Set the stage for improved teaching and learning.
 - Offer more complete evidence about students' learning growth, particularly in nontested subjects and grades.
 - Provide more complete evidence of learning growth for English language learners (ELLs) and students with disabilities.
 - Contribute to a more comprehensive view of students' strengths and areas in which they need improvement.



TQ CENTER RESOURCE

Guide to Evaluation Products

<http://www3.learningpt.org/tqsource/GEP/>

This guide can be used by states and districts to explore various evaluation methods and tools that represent the “puzzle pieces” of an evaluation system.

The guide includes detailed descriptions of more than 75 teacher evaluation tools that are currently implemented and tested in districts and states throughout the country.

The following information is provided for each tool:

- Research and resources
- Information on the teacher and student populations assessed
- Costs, contact information, and technical support offered

There are many potential measures of teacher performance that a state or district could use as part of the evaluation process. Measures of student growth provide specific feedback as to whether a student has progressed as expected in the course of a year. Potential measures include the following:

- Value-added models
- Other growth models
- Other measures (e.g., curriculum-based measures)
- Student learning objectives
- Subject specific tests

Although evidence of teacher effectiveness can be demonstrated, in part, through student growth measures, such measures are limited in distinguishing evidence of instructional quality. These measures are

better able to capture teacher practice, identify learning needs, and guide professional growth. Potential measures include the following:

- Observation instruments
- Performance rubrics
- Portfolios/evidence binders
- Teacher self-assessments
- Parent/student surveys

Each measure has its inherent strengths and weaknesses (See Little, Goe, & Bell, 2009). Likewise, each measure could fulfill a particular evaluation system purpose. Therefore, measure selection is dependent on the overall purpose of the evaluation system. For instance, if the purpose of the system is to improve teacher capacity and collaboration, the selected measures might

include an assortment of measures that provide evidence of instructional quality.

Table 2 reviews potential teacher evaluation goals and identifies the measurement types that are most appropriate to meet those goals. Research and policy has not suggested a particular number of measures that should comprise an evaluation “system”; however, policy does indicate that evidence of student learning should be a “significant” component within teacher evaluation. Hence, a measure of student growth is necessary to provide the “hard” data that effective instructional practices (as demonstrated through evidence of instructional quality measures) lead to student growth (as demonstrated through student growth model measures).

Table 2. Matching Measures to Specific Purposes

Purpose of Evaluation of Teacher Effectiveness	Value-Added	Classroom Observation	Analysis of Artifacts	Portfolios	Teacher Self-Reports	Student Ratings	Other Reports
Find out whether grade-level or instructional teams are meeting specific achievement goals.	X						
Determine whether a teacher's students are meeting achievement growth expectations.	X		X				
Gather information in order to provide new teachers with guidance related to identified strengths and shortcomings.		X	X	X			X
Examine the effectiveness of teachers in lower elementary grades for which no test scores from previous years are available to predict student achievement (required for value-added models).		X	X	X			X
Examine the effectiveness of teachers in nonacademic subjects (e.g., art, music, and physical education).		X		X		X	X
Determine whether a new teacher is meeting performance expectations in the classroom.		X	X	X		X	X
Determine the types of assistance and support a struggling teacher may need.		X	X		X	X	
Gather information to determine what professional development opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	X	X			X		X
Gather evidence for making contract renewal and tenure decisions.	X	X					X
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive pay (rewards).	X	X					
Gather information on a teacher's ability to work collaboratively with colleagues to evaluate needs of and determine appropriate instruction for at-risk or struggling students.				X	X		X
Establish whether a teacher is effectively communicating with parents/guardians.				X			X
Determine how students and parents perceive a teacher's instructional efforts.						X	
Determine who would qualify to become a mentor, coach, or teacher leader.	X	X	X	X			X

Reprinted from *A Practical Guide to Evaluating Teacher Effectiveness* by O. Little, L. Goe, and C. Bell. Copyright © 2009 National Comprehensive Center for Teacher Quality, p. 16.

Adoption of particular measures can be guided by the following factors:

- Evaluation system purpose
- Strength of measures
- Application to all student populations and teaching contexts
- Human and resource capacity strengths and limitations

Evaluation System Purpose

As mentioned previously, goal selection guides all decisions in the design process, particularly in measure selection. Systems designed with higher stakes (e.g., personnel dismissal and renewal decisions) point to measures that are technically defensible (e.g., valid and reliable); whereas, systems designed to improve teacher capacity point to measures of instructional quality. Frequent reflection on the evaluation system's purpose will help direct measure selection.

Strength of Measures

All measures have their own inherent strengths and weaknesses. Not all measures are equally useful nor equally valid and reliable. Measures should be selected based on the following:

- Ability to accurately measure student progress

- Demonstrated impact on student achievement
- Demonstrated impact on teacher practice

Federal priorities (Secretary's Priorities for Discretionary Grant Programs, 2010) provide guidance on student growth measures stipulating that such measures must:

- Be rigorous.
- Measure progress between two points in time.
- Be comparable across classrooms.

At the same time, these measures must be valid and reliable for their intended purposes. In other words, the measure or assessment must accurately and fairly measure what the student is supposed to learn, whether the student learned the material, and how results can be attributed to individual teachers (Herman, Heritage, & Goldschmidt, 2011). Existing potential measures of student growth are not yet likely to meet all these criteria; therefore, stakeholders should factor the measure's strength in terms of the technical adequacy of the instrument as measurement selection is being considered. Likewise, measuring teacher practice through observations or a review of classroom artifacts requires trained raters so that the scores teachers receive are not dependent on who observes them or analyzes artifacts. Demonstrated validity and

reliability within such measures also should guide the selection process. Appendix A provides an overview of measures including descriptions, research base, strengths, and cautions.

Application of Measures to All Student Populations and Teaching Contexts

Applicability to all teaching contexts and student populations also should be considered in the measure selection process. A measure's validity and reliability with all teachers, student populations, and local contexts play an important role in maintaining implementation fidelity and yielding valid and useful results. For example, implementing teacher evaluation systems in rural districts may be more challenging.



TQ CENTER RESOURCE

Alternative Measures of Teacher Performance (Policy-to-Practice Brief)

http://www.tqsource.org/pdfs/TQ_Policy-to-PracticeBriefAlternativeMeasures.pdf

This Policy-to-Practice Brief introduces five current examples of measures of teacher performance. The goal is to assist regional comprehensive centers and state education agencies in building local capacity to incorporate the use of alternative measures of teacher performance into the overhaul of state evaluation systems—especially in states with looming legislative deadlines.

These districts may lack the financial and human resources to implement a system with fidelity, which will likely result in less management support and fewer resources for professional learning. Likewise, the increasing diversity of our nation's classrooms is another factor to consider in measure selection. For example, certain measures may not be appropriate or yield useful information for teachers of students with disabilities, ELLs, or gifted students. In a recent TQ Center Research & Policy Brief, Holdheide, Goe, Croft, and Reschly (2010) address the following specific challenges in evaluating teachers of at-risk populations and measuring student growth in these populations:

- Statewide assessment results may be unavailable (e.g., students working toward alternative standards) or not viable.
- Learning trajectories may be different for students with disabilities and ELLs.
- The “ceiling effect” for gifted students may prevent adequate measurement of student growth.
- Attribution of student growth when multiple teachers are responsible for instruction and observation of teacher practice with multiple teachers in the classroom can be complicated.

Investigation into how measures apply to all teachers and contexts may increase the overall validity and reliability of measures. States need to consider these specific challenges and, if chosen, help districts develop feasible solutions to ensure successful implementation.

Human and Resource Capacity Strengths and Limitations

Each potential measure has associated expenses that need to be factored into the decision-making process. Likewise, some measures require more human capacity than others. Both human and resource capacity strengths and limitations need to be considered in the selection of measures. Implementing measures without regard to the demands they place on teachers, administrators, and others will likely yield results that lack validity or are not implemented with fidelity and thus fail to affect teacher performance and student learning.

Stakeholders may consider the following guiding questions for Component 3 during the measurement selection process.

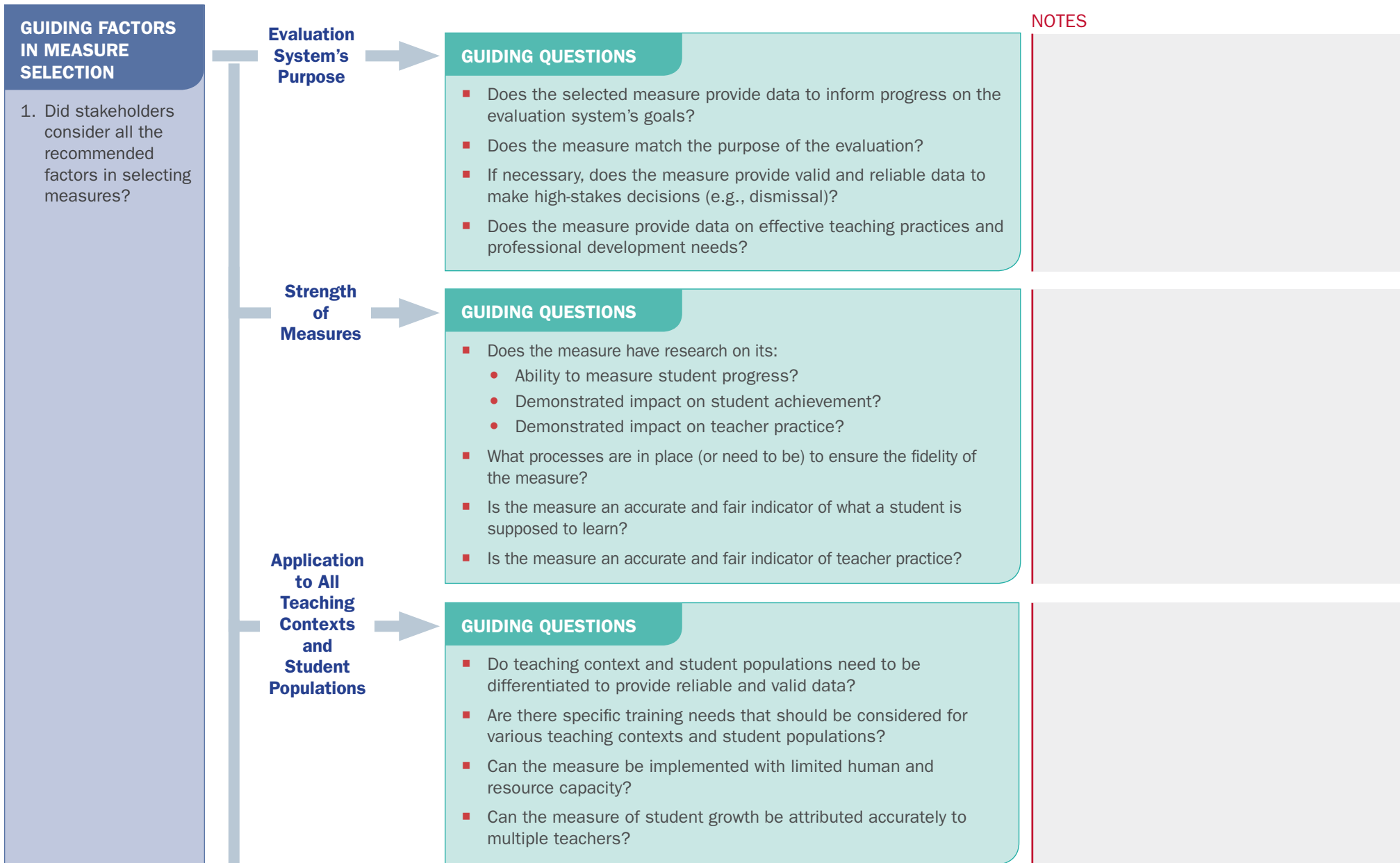


Challenges in Evaluating Special Educators and English Language Learner Specialists (Research & Policy Brief)
(<http://www.tqsource.org/publications/July2010Brief.pdf>)

This Research & Policy Brief offers the following recommendations for states and districts::

- Include special education and ELL administrators and teachers in the process of revamping/designing evaluation models.
- Identify a common framework that defines effective teaching for all teachers. Where appropriate, include differentiated criteria for special education teachers and ELL specialists.
- In addition to—or in the absence of—appropriate standardized assessment data, incorporate other concrete evidence of teachers’ contributions to student learning.

Selecting Measures

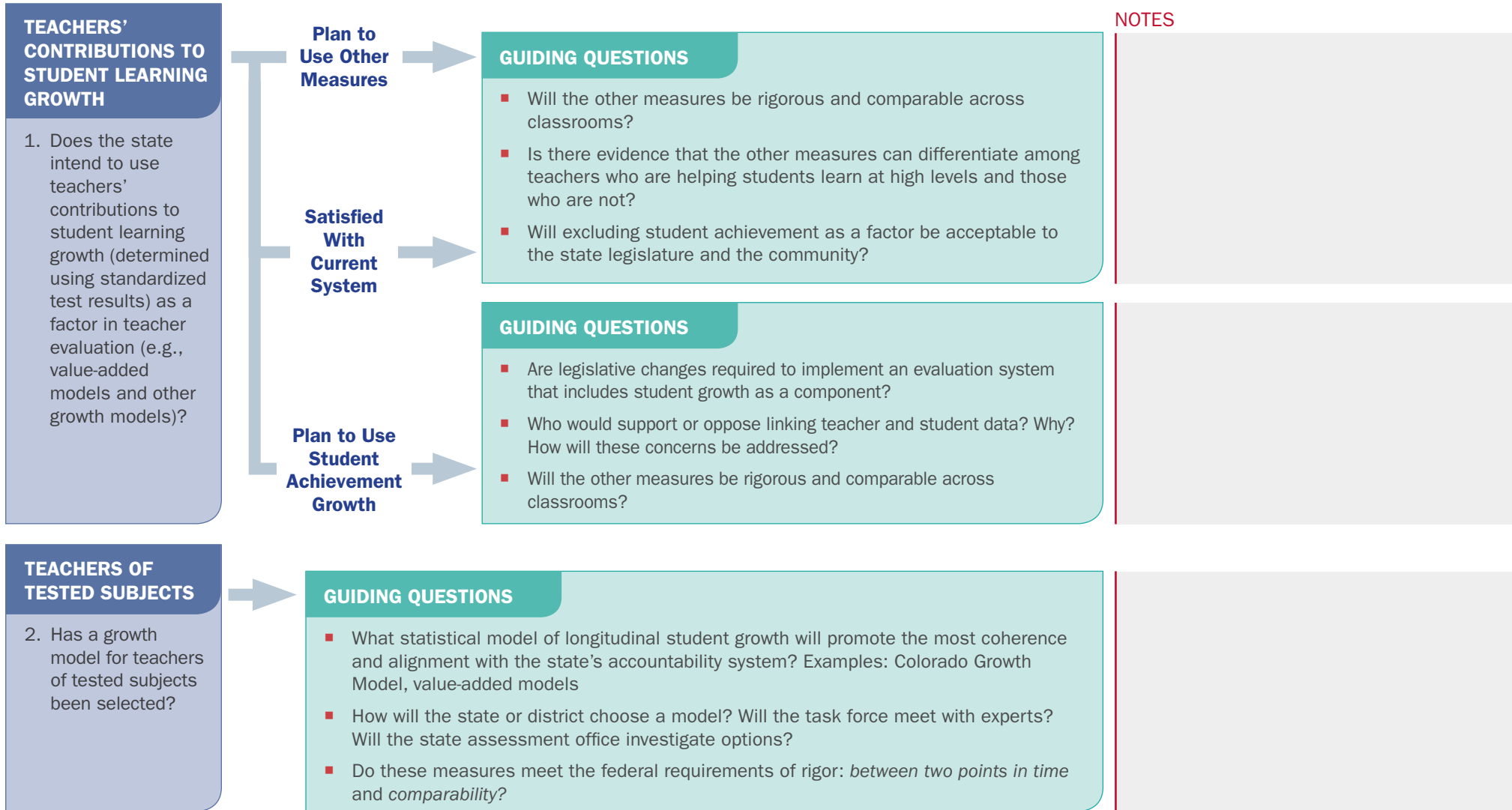


**Human
and
Resource
Capacity**

GUIDING QUESTIONS

- What human and resource capacity is necessary to implement the measure reliably and with validity?
- Can resources be pulled between and within districts to implement the measure?

Measuring Growth in Tested Subjects



PERCENTAGE OF RESULTS BASED ON GROWTH MODEL

3. Has the percentage of teacher evaluation results that will be based on the growth model been determined?

GUIDING QUESTIONS

- What percentage will be supported by the education community?
- What will the state define as significant?
- Is legislation necessary to determine the percentage?
- Are the assessments reliable and valid to support a significant portion of the evaluation to be based on student progress?

IDENTIFICATION OF TEACHERS

4. Have teachers for whom the growth model will be factored into evaluation results been identified?

Teacher Inclusion/Exclusion Criteria

GUIDING QUESTIONS

- Will all teachers of tested subjects be included?
- What is the minimum number of students required for a teacher to be evaluated with student growth (e.g., five students per grade/content area)?
- Are there certain student populations in which inclusion in value-added or other growth models may raise validity questions (e.g., students with disabilities, ELLs)?
- Can students working toward alternative assessments be included in the growth model?
- How will the state or district choose a model? Will the task force meet with experts? Will the state assessment office investigate options?

DATA LINKAGE

5. Can student achievement data be accurately linked to teachers (data integrity)?

Data Integrity

GUIDING QUESTIONS

- What validation process can be established to ensure clean data (e.g., teachers reviewing student lists, administrators monitoring input)?
- Can automatic data validation programs be developed?
- Are there certain student populations in which inclusion in value-added or other growth models is not appropriate (e.g., students with disabilities, ELLs)?

Teaching Context/
Extenuating Circumstances

GUIDING QUESTIONS

- Has the teacher attribution process been established for coteaching situations?
- How will teachers with high student absenteeism rates or highly mobile students be evaluated?
- Has a focus group been held with teachers to determine fair attribution?

DETERMINATION OF ADEQUATE GROWTH

6. Has a process been established to determine adequate student growth?

GUIDING QUESTIONS

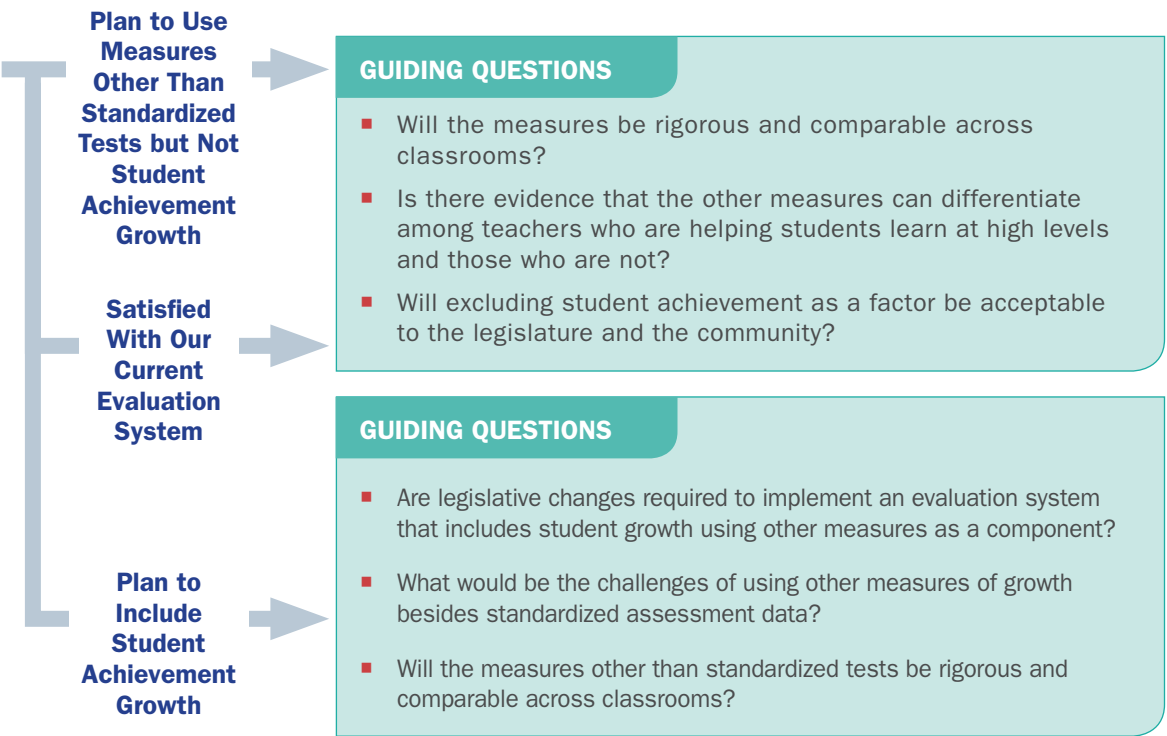
- What does the research suggest regarding the number of years teacher data should be collected in order to use it as part of teacher evaluation?
- Will the learning trajectory be different for at-risk, special needs, or gifted students?
- Has the ceiling effect been addressed?
- Will the use of accommodations affect the measure of student growth?
- Does this measure meet the federal requirements of rigor: *between two points in time and comparability*?

NOTES

Alternative Growth Measures in Tested and Nontested Subjects

MEASURES OTHER THAN STANDARDIZED TESTS

1. Does the state intend to use measures other than standardized tests to determine student growth (e.g., classroom-based assessments; interim or benchmark assessments; curriculum-based assessments; the Four Ps: projects, portfolios, performances, products)?



NOTES

IDENTIFICATION OF TEACHERS

2. Have the teachers who meet the criteria for use of measures other than standardized tests been identified?

GUIDING QUESTIONS

- Will all teachers (in both tested and nontested subjects) be evaluated with alternative growth measures? Only teachers of nontested subjects?
- Which teachers fall under the category of nontested subjects?
- Are there teachers of certain student populations or situations in which standardized test scores are not available or appropriate to utilize?
- Will contributions to student learning growth be measured for related services personnel?

NOTES

IDENTIFICATION OF MEASURES

3. Have measures to determine student learning growth been identified?

Content Standards

GUIDING QUESTIONS

- Do content standards exist for all grades and subjects?
- Is there a consensus on the key competencies students should achieve in the content areas?
- Can these content standards be used to guide selection and development of measures?

Measure Selection

GUIDING QUESTIONS

- Which stakeholders need to be involved in determining or identifying measures?
- What type of meetings or facilitation will stakeholder groups require to select or develop student measures?
- How will growth in performance subjects (e.g., music, art, physical education) be determined to demonstrate student growth?
- Will the state use classroom-based assessments, interim or benchmark assessments, curriculum-based assessments, and/or the Four Ps (i.e., projects, portfolios, performances, products) as measures?
- Are there existing measures that could be considered (e.g., end-of-course assessments, DIBELS, DRA)?
- Could assessments be developed or purchased?

FEDERAL REQUIREMENTS

4. Do these measures meet the federal requirements of rigor: *between two points in time and comparability*?

Validity and Reliability

GUIDING QUESTIONS

- Does the measure accurately and fairly measure what the student is supposed to learn?
- Does the measure assess what it is intended to assess?
- Can the measure accurately indicate levels of student growth in the course of a year?
- Can student growth be accurately linked to teachers' efforts?
- Are there appropriate assessments for all grades and all teachers, including special educators and ELL specialists?

RESEARCH

5. Are there plans to conduct research during implementation to increase confidence in the measures?

GUIDING QUESTIONS

- Are federal, state, or private funds available to conduct research?
- How will the content validity be tested?
- Can national experts in measurement and assessment be appointed to assist in conducting this research?

Guiding Questions

Observation Measures

MEASURE OF INSTRUCTIONAL QUALITY

1. Does the state intend to use measures other than observations as indicators of instructional quality?

GUIDING QUESTIONS

- If observations will not be used, how will the results from other measures be used to guide and strengthen teacher practice?
- Will the other measures be able to detect teacher strengths and weaknesses?
- Will the other measures be able to identify effective teaching practices?
- Will the other measures be able to identify professional development needs?

NOTES

RESEARCH BASE

2. Is there a research base for this observation tool?

GUIDING QUESTIONS

- Has the tool/instrument been piloted?
- Can results from the tool/instrument be correlated with improved student achievement?
- Have any research studies been conducted on this tool/instrument?

APPLICABILITY

3. Is the observation instrument applicable to all teachers and teaching contexts?

GUIDING QUESTIONS

- Is there any teacher population that requires differentiation in the observation process? For example, do teachers of special populations (e.g., special education students, ELLs) require different instruments and/or different observers?
- Will teachers serving in a coteaching capacity need to be observed with a different or modified tool, or will specialized training be required for evaluators to appropriately use the tool in these settings?
- Will teachers of specific content areas benefit from a more specialized tool that focuses on evidence-based practices in the content area?

PROCESS

4. Has the observation process been thoroughly specified?

Evaluators

GUIDING QUESTIONS

- Who will conduct the teacher observations (e.g., administrators, master teachers, peers)?
- Could expert teachers be appointed to conduct the observations?
- Will building administrators have the time and expertise to conduct the observations?
- Will more than one evaluator observe each teacher?

Frequency

GUIDING QUESTIONS

- How often will observations be required? Will it vary depending on teachers' levels of experience?

Training and Interrater Reliability

GUIDING QUESTIONS

- What training and/or certification will be required to qualify as an evaluator?
- How will the district or state ensure that evaluators can use the observation instrument with fidelity?
- How will the district or state ensure interrater reliability? During training? Over time?

Teacher Reflection

GUIDING QUESTIONS

- Will teachers have access to all observation forms and materials in advance?
- Will teachers' self-assessments on the instruments (to be compared to the evaluator's assessment) be part of the process?
- Will preobservation and/or postobservation conferences be conducted?
- How will the observation instruments support teachers in reflecting on their practice?

COMPONENT 4

Determining the Structure of the Evaluation System

When determining the structure of the system, stakeholders must consider the designated levels of performance; the frequency of evaluations, as applicable; and a number of other factors related to implementation. In designating the number and description of levels, states must ensure that the level designations (e.g., developing, proficient, exemplary) work for teachers at different experience levels. Likewise, the instruments must be sensitive enough to identify the appropriate level of reliability.

In addition, it is important that the frequency of evaluation is considered separately for each measure used. Classroom observations, for example, are often conducted several times throughout the year, whereas analyses of teacher artifacts may be performed at a different frequency. The teacher's level of performance or experience also may be a factor in determining the appropriate frequency of evaluation. Beginning teachers, or teachers with identified areas of weakness, may be evaluated more frequently than teachers who have reached exemplary or master status.

States may elect to mandate specific format requirements or allow for local flexibility. When making these determinations, states should consider implementation fidelity and reliability, local bargaining restraints, and resource limitations.

As mentioned previously, all measures are not equally reliable and useful. States also may want to consider the measure's strength in comparison with the other measures used within the evaluation system. Measures that have higher validity and reliability may be used with more confidence. The measure's weight within a system may be dependent on its validity, its impact on student achievement, the information it provides to help teachers improve their practice, or other considerations. In some scenarios, states may gradually increase the weight of a measure as confidence in the measure increases and technical rigor is enhanced. For instance, states may determine that current assessments have not been validated for the purposes of teacher evaluation. In this case, data need to be collected and analyzed and compared with other types of evidence to determine whether the results are valid. As the system is evaluated and results, which increase or decrease confidence in the measures, are obtained, the weights

may need to be revisited. The measure's weight also may be reflective of the evaluation system's goals. If collaboration between teachers is a priority, a rubric measuring teacher capacity to collaborate may be weighted more heavily. Or if the ultimate goal of the system is to increase teacher capacity to implement evidence-based practices, the observation instrument may carry more weight.



TQ CENTER RESOURCE

Teacher Evaluation Models in Practice
<http://www3.learningpt.org/tqsource/evalmodel/>

This interactive online resource responds to the need for detailed information about the design, implementation, and delivery of teacher evaluation models in practice in districts and states. It includes an overview of district evaluation models with links to their documentation, tools, training materials, and resources. It also contains lessons learned from an in-depth examination of district efforts by national experts in measurement and assessment.

Stakeholders might consider the guiding questions for Component 4 as they determine the structure of the evaluation system.

Guiding Questions

Determining the Structure of the Evaluation System

MULTIPLE MEASURES

1. Will the state promote or use multiple measures?

GUIDING QUESTIONS

- Will a single measure be sufficient in making defensible decisions regarding teacher effectiveness?
- Will a single measure accurately capture teacher capacity in terms of ability to elicit improved student achievement and implement evidence-based instructional strategies?

NOTES

WEIGHT OF MEASURES

2. Has the state determined the percentage (weight) of each measure in the overall teacher rating?

GUIDING QUESTIONS

- Will each measure be weighted differently depending on:
 - Its relation to student achievement?
 - Its reliability and validity?
 - Its face validity?
- Will the weight of each measure fluctuate depending on the level of reliability and validity that is proven over time?
- Will the weight of each measure vary depending on teaching discipline and context?

LEVELS OF PROFICIENCY

3. Have the levels of teaching proficiency been determined?

GUIDING QUESTIONS

- How many levels of proficiency can be explicitly defined?
- Can rubrics be developed to ensure fidelity?
- How often can data be generated?
- What implementation limitations should be considered (e.g., how frequently assessments can be conducted)?
- Will baseline data be analyzed prior to making decisions regarding teacher proficiency levels?

FAILURE TO MEET PERFORMANCE LEVELS

4. Have consequences been determined for failure to meet acceptable performance levels?



GUIDING QUESTIONS

- Are opportunities for teachers to improve going to be embedded in the evaluation cycle?
- Are the measures technically defensible to make personnel and compensation decisions?
- Will teacher supports be provided to assist teachers with unacceptable performance?
- How much time and assistance will be provided for a teacher to demonstrate improvement before termination is considered?
- Will teacher performance affect tenure?

NOTES

COMPONENT 5

Selecting and Training Evaluators

Most measures require some level of training. The amount of training required to implement the evaluation system is highly dependent on the type of measure being considered. For example, value-added measures of student growth would require training related to the technical aspects of the system and how the data can be interpreted. Observations would require a substantial investment in training for evaluators to ensure interrater reliability as well as training for teachers and administrators in using the results to inform practice. States need to consider their own human capacity strengths and limitations in making decisions about measurement types to ensure that implementation fidelity is maintained. Moreover, local capacity limitations should be considered. For example, it may be unrealistic to mandate an evaluation system that requires a large investment in training

raters if state and district budgets are tight. Districts may need flexibility in funding and implementing evaluation models with the resources they have.

Implementation fidelity is most important when the selected measures are dependent on human scoring with observation instruments or rubrics. Effective evaluator selection and training is essential if the integrity of the system is to be maintained, ensuring that the resulting scores are fair and defensible. Including targeted evaluator training with explicit decision rules and examples of evidence that would justify one performance rating over another may help with interrater reliability. Training, coupled with feedback and support, will likely lead to a high level of integrity.

Likewise, with measures dependent on personnel, evaluators may have difficulty when observing someone outside of their area of expertise. Most observation instruments (e.g., Charlotte Danielson's *Framework for Teaching*, CLASS, and others)

are designed to evaluate all teachers without regard to content area. However, trained evaluators with knowledge of specialist roles and subject-matter competence may be seen as more credible and pick up on nuances in instruction that other raters would miss. States could use mentors or teacher leaders with expertise in content areas as evaluators to ensure appropriate frequency, duration, and feedback related to content/discipline.

Stakeholders might consider the guiding questions for Component 5 during the evaluator selection and training process.

Guiding Questions

Selecting and Training Evaluators

PERSONNEL

1. Do the selected measures require trained personnel to use rubrics or other sources of documentation to determine the level of teacher effectiveness?



GUIDING QUESTIONS

- If personnel are not utilized to determine teacher proficiency, are there other personnel training needs (e.g., interpreting value-added scores, tracking progress-monitoring data)?

NOTES

--

TRAINING AND GUIDELINES

2. Will the state provide training or guidelines on evaluator/reviewer selection and training?

Selection →

GUIDING QUESTIONS

- What criteria will be used to select evaluators or reviewers?
- Who will be eligible to conduct the evaluations?
- Which personnel will conduct evaluations/approve student learning targets?
- Will the state require evaluators or reviewers to have content knowledge and/or experience in the subject area/level being evaluated?
- Could teacher-to-teacher evaluations or reviews be considered?

Training →

GUIDING QUESTIONS

- How will the state ensure implementation fidelity?
- Will the state offer specialized training for the evaluation of or review of specific content or specialty area teachers?
- To what extent will the training provide opportunities for guided practice paired with specific feedback to improve reliability?
- Will the state provide examples and explicit guidance in determining levels of proficiency and approval?

--

--

RETRAINING

3. Does the state have a system in place to retrain evaluators/reviewers if the system is not implemented with fidelity?



GUIDING QUESTIONS

- How will the state address personnel time limitation for conducting evaluations or reviews?
- If evaluators/reviewers are not implementing the system with fidelity, what mechanisms will be in place to retrain evaluators/reviewers?
- Will evaluators/reviewers be monitored regularly for checks in reliability?

COMPONENT 6

Ensuring Data Integrity and Transparency

Data infrastructure that can be used to collect, validate, interpret, track, and communicate teacher performance data will be necessary to inform stakeholders, guide professional learning, and assess the measures and the teacher evaluation system as a whole. The evaluation system goals can guide this development and influence the required data elements.

An integral step in this process is ensuring that the data are sound. Data integrity is crucial in all types of data-based decision making—whether making high-stakes personnel decisions and/or targeting

professional learning activities. Verifying and cleaning existing data and establishing means to collect the required data elements requires a thorough understanding of available and potential data sources. Therefore, collaboration between teachers (who know their students and their classrooms) and information technology personnel (who know the data) to structure the data collection will lead to greater accuracy.

Transparency of measures and resulting data is also a key factor in measure selection. Measures that provide real-time feedback, are accessible and easily understood, and have direct application to teacher practice are more likely to have an immediate impact on teaching and learning. If teachers and

administrators are expected to enter information into data portals, ensuring that these portals are user-friendly will be critical as states scale up evaluation efforts.

Stakeholders might consider the guiding questions for Component 6 to ensure data integrity and transparency.

Guiding Questions

Ensuring Data Integrity and Transparency

DATA INFRASTRUCTURE

1. Is the data infrastructure to collect teacher evaluation data established?

GUIDING QUESTIONS

- Does the state or district have the data infrastructure to link teachers to individual student data including unique identifiers for both teachers and students?
- Have the critical questions that stakeholders want the evaluation system to answer been identified? Will the data system collect sufficient information to answer them?
- Have information technology personnel been brought into the discussion?
- Do districts have the technology and human capacity to collect data accurately?

NOTES

DATA VALIDATION

2. Is there a data validation process to ensure the integrity of the data?

Validation

GUIDING QUESTIONS

- What validation process can be established to ensure clean data (e.g., teachers reviewing student lists, administrators monitoring input)?
- Have criteria been established to ensure teacher/student confidentiality?
- Can computerized programs be used/developed for automatic data validation?

Training

GUIDING QUESTIONS

- What training will personnel need to ensure accurate data collection?
- Which personnel at the state and district levels will require training to ensure accuracy in data entry and reporting?

REPORTING

3. Can teacher evaluation data be reported (aggregated/disaggregated) to depict results at the state, district, building, or classroom level?

Teacher Data

Student Data

GUIDING QUESTIONS

- Do administrators/teachers have access to the teacher evaluation data?
- Is there a system whereby teachers or administrators can make changes when errors are found?
- Is the data collection methodology/database easily understood and user-friendly?
- Have teachers been trained to extrapolate and use the data to inform teacher practice?
- Are administrators, teachers, and parents (as appropriate) trained in how to use the database and interpret teacher evaluation results?

NOTES

USE OF DATA

4. Is there a plan for how the teacher evaluation data will be used?

Data Sharing

Data Use

GUIDING QUESTIONS

- How frequently should teacher evaluation data be shared with the education community?
- What teacher evaluation data would be relevant, easily understood, and appropriate to share with the education community?
- Will administrators and teachers have access to the teacher evaluation data?
- How will evaluation results be shared with the community (e.g., website, press releases, town meetings)?

GUIDING QUESTIONS

- Will teacher evaluation data be used to inform changes in the teacher evaluation design?
- Will administrators, teachers, and parents (as appropriate) be trained in how to use the database and interpret teacher evaluation results?
- Will data be used to identify teachers in need of support and target professional learning?
- Will data be used to identify highly effective teachers and potential mentors?

Using Teacher Evaluation Results

Selecting Trigger Points for Action

If a state plans to use its evaluation system for personnel decisions, designations of when action will be triggered need to be determined and communicated to the teacher workforce. For example, if evaluation results are tied to teacher advancement, will the teacher need to achieve exemplary ratings for three consecutive evaluation cycles prior to promotion? Will achieving exemplary ratings during two of four cycles trigger advancement? If ameliorative action is indicated, in how many evaluation cycles will improvement be expected?

Targeting Professional Development

Using evaluation results to support professional learning is likely the most significant phase of the evaluation cycle. An evaluation system's capacity to reliably identify highly effective and ineffective teachers is important. However, ensuring that teacher ratings can reliably detect teacher strengths and weaknesses is essential for accurately targeting

professional development. Evaluation results can then be used to identify individual, school, and districtwide needs; target professional learning; gauge teacher growth; and identify potential mentors. Providing job-embedded, ongoing, individualized professional learning and support is necessary for teacher evaluation to have positive impacts on teacher practice.

As professional development is incorporated into the evaluation cycle, stakeholders need to evaluate outcomes to determine whether the efforts have improved teaching practice. This process goes beyond a simple evaluation of the professional learning activity, moving toward a continual, longitudinal reflection and analysis of teacher participation, support, and outcomes related to student achievement. Investing in the technical infrastructure to collect, link, and analyze professional development and teacher evaluation results over time may improve the overall effectiveness of professional learning efforts.

Stakeholders might consider the guiding questions for Component 7 as they contemplate professional development needs.



TQ CENTER RESOURCE

Job-Embedded Professional Development: What It Is, Who Is Responsible, and How to Get It Done Well

(<http://www.tqsource.org/publications/JEPD%20Issue%20Brief.pdf>)

This issue brief provides specific recommendations for states to support high-quality job-embedded professional development (p. 10):

- “Help build a shared vocabulary.”
- “Provide technical assistance.”
- “Monitor implementation.”
- “Identify successful job-embedded professional development practices within the state.”
- “Align teacher licensure and relicensure requirements with high-quality job-embedded professional development.”
- “Build comprehensive data systems.”

Guiding Questions

Using Teacher Evaluation Results

TRIGGER POINTS FOR ACTION

1. Have trigger points for action using evaluation results been established?

GUIDING QUESTIONS

- Does the state intend to align evaluation results to human resource decisions?
- At what point will evaluation results warrant a promotion, dismissal, etc.?
- How many evaluation cycles will be used to ensure that opportunity for professional growth is provided?
- How will evaluation results be shared with teachers? When will teachers be notified of next steps toward professional growth or termination?

NOTES

EVALUATION CYCLE

2. Is professional development an integral component of the evaluation cycle?

GUIDING QUESTIONS

- Is a goal of the evaluation system to improve teacher capacity? If so, how will the evaluation system affect teacher practice?
- Will teachers identified as ineffective have sufficient opportunities and support to improve before termination is considered?
- Will personnel decisions be defensible if teachers were not provided an opportunity and the resources to improve?
- What resources, including time and personnel, are dedicated to teacher improvement?

EVALUATION RESULTS

3. Will teacher evaluation results be used to target professional development activities?

GUIDING QUESTIONS

- How will professional development opportunities be determined for teachers, schools, and the district?
- How will data obtained through the various teacher evaluation measures inform professional development offerings?
- How can the evaluation system be retooled to reliably detect teacher strengths and weaknesses?
- Can teacher evaluation results be used to identify teachers for roles such as mentor teachers, master teachers, and consulting teachers?

RESEARCH

4. Are professional learning activities provided in a manner that is supported in research?



GUIDING QUESTIONS

- What human and fiscal resources can be used to provide job-embedded professional development?
- Can teacher application and reflection be built into the professional learning activity?
- Are professional learning activities “job-embedded” or a one-time-only session?
- Do teachers have common planning times to reflect upon new practices?
- Can opportunities for teachers to observe effective teachers be provided?
- Will professional learning communities be established?

EVALUATION SYSTEMS

5. Are systems established to evaluate professional learning efforts?

Evaluating the Training

GUIDING QUESTIONS

- What mechanism will be established to ensure that participant feedback is obtained (e.g., training evaluation, follow-up survey)?
- What procedures will be established to ensure that active participation and application is an integral part of the professional development activity?

Reviewing the Outcomes

GUIDING QUESTIONS

- Can the evaluation measure(s) detect teacher growth as a result of professional development efforts?
- Can demonstrated teacher growth be correlated to improved student achievement?
- What mechanism will be established to follow up on teachers to ascertain whether teacher practice has been improved as a result of the professional learning efforts (e.g., follow-up survey/observation)?

Modifying the Process

GUIDING QUESTIONS

- Can the system identify which professional learning opportunities are/are not effective?
- Are changes in the evaluation system necessary to correlate teacher and student growth with participation in professional learning activities?
- How will results (e.g., evaluations and outcomes) be used to improve professional development offerings and strategies?

NOTES

COMPONENT 8

Evaluating the System

Systematically evaluating the performance of the evaluation model in terms of its goals and results and modifying its structure, processes, or format accordingly assures system efficacy and sustainability. States need to identify the factors that will determine whether the system is effective.

For example, the state and districts will want to know whether:

- Stakeholders value and understand the system.
- Student performance is improved.
- Teacher practice is affected.
- Teacher retention is improved.
- The system is implemented with fidelity.

States have used external and internal review processes to collect and analyze data or a combination of both. Surveys of teachers, administrators, and stakeholders may be valuable for this process. Ultimately, researchers should work closely with stakeholders to ensure that the design allows important questions to be answered.

Stakeholders might consider the guiding questions for Component 8 when determining the evaluation process for the system.

Guiding Questions

Evaluating the System

EVALUATION PROCESS

1. Has a process been developed to systematically evaluate the effectiveness of the teacher evaluation model?

GUIDING QUESTIONS

- How will the stakeholders know whether the new teacher evaluation model is effective?
- Has the model been piloted or are there plans to pilot the model prior to statewide or districtwide implementation?
- Is there a plan for securing stakeholder and participant feedback?
- Will research be conducted in conjunction with implementation to provide validation?
- Are the goals of the evaluation system a good measure of effectiveness?
- Will research be conducted to determine whether there is correlation between growth model scores and observation ratings?

NOTES

EFFECTIVENESS OUTCOMES

2. Have outcomes to determine the overall effectiveness of the evaluation system been established?

GUIDING QUESTIONS

- Have the stakeholders identified factors that should be considered in determining whether the evaluation system is effective (e.g., participant satisfaction, improved teacher practice, other improved student outcomes)?
- Are resources available to conduct an internal or external assessment of the evaluation model?
- Has the data infrastructure been established to track data over a period of time to determine teacher and student growth?
- Have explicit benchmarks or targets been established to determine effectiveness?
- In review of baseline data, what would be acceptable performance targets?

OTHER ASPECTS OF TEACHER QUALITY

3. Will other aspects of teacher quality that affect teacher performance be reviewed to determine whether they have been influenced by the evaluation system?



GUIDING QUESTIONS

- If the teacher evaluation plan includes modifications in tenure, promotion, or compensation, how will the state conduct research to determine the level of effectiveness on teacher retention and improved teacher capacity?
- Will the teacher evaluation plan include working in collaboration with teacher preparation programs to ensure that candidates are prepared with the competencies for which they will be held accountable when they begin teaching?
- Will data be collected on teacher effectiveness to determine whether effective teachers are equally distributed throughout the state—including both high-performing and low-performing schools?
- Will research be conducted to determine whether professional development efforts have resulted in improved teacher practice and student outcomes?

Conclusion and Recommendations

Designing a comprehensive teacher evaluation system in an effective and sustainable manner is a difficult process, especially with few research-based models to consider. States are charged with overseeing this process, which for many is unfamiliar territory because historically, evaluation in most states has been left up to districts. Using teacher evaluation to improve teacher practice in schools should be the ultimate goal of state and district efforts. Identifying areas in which teacher practice can be

improved and providing targeted professional learning opportunities to teachers should go a long way toward addressing the persistent achievement gaps in our nation's schools.

Too often, teacher evaluation is seen as a mechanism for enforcing personnel decisions rather than cultivating effective teaching. Adding to the challenges of creating comprehensive teacher evaluation systems is the relationship between state and district leaders and teachers. Building trust and ensuring collaboration toward common goals requires substantial resources, including time, patience, and resilience. To further the development of direct links between teacher

evaluation and instructional improvement, states and districts need to nurture an educational climate in which evaluation is not seen as punitive and teachers are highly invested in the process. The core of evaluation reform efforts should be human capacity building at all levels so that states, districts, and schools can identify and learn from top-performing teachers, support discouraged and less successful teachers, and continue to develop all teachers toward their full potential.

References

- American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009). Retrieved May 16, 2011, from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.txt.pdf
- Cincinnati Public Schools. (n.d.). *Teacher evaluation*. Cincinnati, OH: Author. Retrieved May 16, 2011, from <http://www.cps-k12.org/employment/tchreval/tchreval.htm>
- Colorado Department of Education. (2000). *Performance-based standards for Colorado teachers*. Denver, CO: Author. Retrieved May 16, 2011, from http://www.cde.state.co.us/cdeprof/download/pdf/li_perfbasedstandards.pdf
- Council of Chief State School Officers. (2011). *InTASC Model Core Teaching Standards: A resource for state dialogue*. Washington, DC: Author. Retrieved May 16, 2011, from http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf
- Danielson, C. F. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: ASCD.
- Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context* (NCES 2011-004). Washington, DC: National Center for Education Statistics. Retrieved May 16, 2011, from <http://nces.ed.gov/pubs2011/2011004.pdf>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved May 16, 2011, from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>
- Goe, L., & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects* (Research & Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved May 16, 2011, from <http://www.tqsource.org/publications/MeasuringTeachersContributions.pdf>
- Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Guidance for developing and selecting student growth measures for use in teacher evaluation*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved May 16, 2011, from http://www.swcompcenter.org/educator_effectiveness2/student-growth-measures-for-use-in-teacher-evaluationMARGARET_HERITAGE.pdf
- Holdheide, L., Goe, L., Croft, A., & Reschly, D. (2010). *Challenges in evaluating special education teachers and English language learner specialists* (Research & Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved May 16, 2011, from <http://www.tqsource.org/publications/July2010Brief.pdf>

-
- Learning Point Associates. (2010). *Evaluating teacher effectiveness: Emerging trends reflected in the state Phase 1 Race to the Top applications*. Naperville, IL: Learning Point Associates. Retrieved May 16, 2011, from http://www.learningpt.org/pdfs/RttT_Teacher_Evaluation.pdf
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective* (NCES 2005-003). Washington, DC: National Center for Education Statistics. Retrieved May 16, 2011, from <http://nces.ed.gov/pubs2005/2005003.pdf>
- Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved May 16, 2011, from <http://www.tqsource.org/publications/practicalGuide.pdf>
- National Board for Professional Teaching Standards. (2011). The standards [Website]. Arlington, VA: Author. Retrieved May 16, 2011, from http://www.nbpts.org/the_standards
- New York State Education Department. (2011). *Draft teacher and principal evaluation regulations*. Albany, NY: Author.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved May 16, 2011, from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- North Carolina Professional Teaching Standards Commission. (n.d.). *North Carolina professional teaching standards*. Raleigh, NC: Author. Retrieved May 16, 2011, from <http://www.ncpublicschools.org/docs/profdev/standards/teachingstandards.pdf>
- Ohio Department of Education. (n.d.). *Achieving more through educator quality* [PowerPoint presentation]. Columbus, OH: Author.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Secretary's Priorities for Discretionary Grant Programs, 75 Fed. Reg. 47,288 (proposed Aug. 5, 2010). Retrieved May 16, 2011, from <http://www2.ed.gov/legislation/FedRegister/other/2010-3/080510d.pdf>
- U.S. Department of Education. (2010). *Phase I Race to the Top application*. Washington, DC: Author. Retrieved May 15, 2011, from <http://www2.ed.gov/programs/racetothetop/application.doc>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved May 16, 2011, from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>

Appendix A. Summary of Measures

Measure	Description	Research	Strengths	Cautions
Classroom Observation	Used to measure observable classroom processes, including specific teacher practices, holistic aspects of instruction, and interactions between teachers and students. Can measure broad, overarching aspects of teaching or subject-specific or context-specific aspects of practice.	Some highly researched protocols have been found to link to student achievement, though associations are sometimes modest. Research and validity findings are highly dependent on the instrument used, sampling procedures, and training of raters. There is a lack of research on observation protocols as used in context for teacher evaluation.	<ul style="list-style-type: none"> Provides rich information about classroom behaviors and activities. Is generally considered a fair and direct measure by stakeholders. Depending on the protocol, can be used in various subjects, grades, and contexts. Can provide information useful for both formative and summative purposes. 	<ul style="list-style-type: none"> Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters. Classroom observation is expensive due to cost of observers' time; intensive training and calibrating of observers adds to expense but is necessary for validity. This method assesses observable classroom behaviors but is not as useful for assessing beliefs, feelings, intentions, or out-of-classroom activities.
Principal Evaluation	Is generally based on classroom observation, may be structured or unstructured; uses and procedures vary widely by district. Is generally used for summative purposes, most commonly for tenure or dismissal decisions for beginning teachers.	Studies comparing subjective principal ratings to student achievement find mixed results. Little evidence exists on validity of evaluations as they occur in schools, but evidence exists that training for principals is limited and rare, which would impair validity of their evaluations.	<ul style="list-style-type: none"> Can represent a useful perspective based on principals' knowledge of school and context. Is generally feasible and can be one useful component in a system used to make summative judgments and provide formative feedback. 	<ul style="list-style-type: none"> Evaluation instruments used without proper training or regard for their intended purpose will impair validity. Principals may not be qualified to evaluate teachers on measures highly specialized for certain subjects or contexts.
Instructional Artifact	Structured protocols used to analyze classroom artifacts in order to determine the quality of instruction in a classroom. May include lesson plans, teacher assignments, assessments, scoring rubrics, and student work.	Pilot research has linked artifact ratings to observed measures of practice, quality of student work, and student achievement gains. More work is needed to establish scoring reliability and determine the ideal amount of work to sample. Lack of research exists on use of structured artifact analysis in practice.	<ul style="list-style-type: none"> Can be a useful measure of instructional quality if a validated protocol is used, if raters are well-trained for reliability, and if assignments show sufficient variation in quality. Is practical and feasible because artifacts have already been created for the classroom. 	<ul style="list-style-type: none"> More validity and reliability research is needed. Training knowledgeable scorers can be costly but is necessary to ensure validity. This method may be a promising middle ground in terms of feasibility and validity between full observation and less direct measures such as self-report.

Measure	Description	Research	Strengths	Cautions
Portfolio	<p>Used to document a large range of teaching behaviors and responsibilities.</p> <p>Has been used widely in teacher education programs and in states for assessing the performance of teacher candidates and beginning teachers.</p>	<p>Research on validity and reliability is ongoing, and concerns have been raised about consistency/stability in scoring. There is a lack of research linking portfolios to student achievement. Some studies have linked NBPTS certification (which includes a portfolio) to student achievement, but other studies have found no relationship.</p>	<ul style="list-style-type: none"> • Is comprehensive and can measure aspects of teaching that are not readily observable in the classroom. • Can be used with teachers of all fields. • Provides a high level of credibility among stakeholders. • Is a good tool for teacher reflection and improvement. 	<ul style="list-style-type: none"> • This method is time-consuming on the part of teachers and scorers; scorers should have content knowledge of the portfolios. • The stability of scores may not be high enough to use for high-stakes assessment. • Portfolios are difficult to standardize (compare across teachers or schools). • Portfolios represent teachers' exemplary work but may not reflect everyday classroom activities.
Teacher Self-Report Measure	<p>Teacher reports of what they are doing in classrooms. May be assessed through surveys, instructional logs, and interviews. Can vary widely in focus and level of detail.</p>	<p>Studies on the validity of teacher self-report measures present mixed results. Highly detailed measures of practice may be better able to capture actual teaching practices but may be harder to establish reliability or may result in very narrowly focused measures.</p>	<ul style="list-style-type: none"> • Can measure unobservable factors that may affect teaching, such as knowledge, intentions, expectations, and beliefs. • Provides the unique perspective of the teacher. • Is very feasible and cost-efficient; can collect large amounts of information at once. 	<ul style="list-style-type: none"> • Reliability and validity of self-report is not fully established and depends on instrument used. • Using or creating a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings. • This method should not be used as a sole or primary measure in teacher evaluation.
Student Survey	<p>Used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by students.</p>	<p>Several studies have shown that student ratings of teachers can be useful in providing information about teaching; may be as valid as judgments made by college students and other groups; and, in some cases, may correlate with measures of student achievement. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only.</p>	<ul style="list-style-type: none"> • Provides perspective of students who have the most experience with teachers. • Can provide formative information to help teachers improve practice in a way that will connect with students. • Makes use of students, who may be as capable as adult raters at providing accurate ratings. 	<ul style="list-style-type: none"> • Student ratings have not been validated for use in summative assessment and should not be used as a sole or primary measure of teacher evaluation. • Students cannot provide information on aspects of teaching such as a teacher's content knowledge, curriculum fulfillment, and professional activities.

Value-Added Model	Used to determine teachers' contributions to students' test score gains. May also be used as a research tool (e.g., determining the distribution of "effective" teachers by student or school characteristics).	Little is known about the validity of value-added scores for identifying effective teaching, though research using value-added models does suggest that teachers differ markedly in their contributions to students' test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.	<ul style="list-style-type: none"> • Provides a way to evaluate teachers' contribution to student learning, which most measures do not. • Requires no classroom visits because linked student/teacher data can be analyzed at a distance. • Entails little burden at the classroom or school level because most data are already collected for NCLB purposes. • May be useful for identifying outstanding teachers whose classrooms can serve as "learning labs" as well as struggling teachers in need of support. 	<ul style="list-style-type: none"> • Models are not able to sort out teacher effects from classroom effects. • Vertical test alignment is assumed (i.e., tests essentially measure the same thing from grade to grade). • Value-added scores are not useful for formative purposes because teachers learn nothing about how their practices contributed to (or impeded) student learning. • Value-added measures are controversial because they measure only teachers' contributions to student achievement gains on standardized tests.
-------------------	---	--	---	--

Reprinted from *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*, by L. Goe, C. Bell, and O. Little. Copyright © 2008 National Comprehensive Center for Teacher Quality, pp. 16–19.



NATIONAL COMPREHENSIVE CENTER
FOR TEACHER QUALITY

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
877.322.8700 | 202.223.6690

www.tqsource.org

Copyright © 2011 National Comprehensive Center for Teacher Quality, sponsored under government cooperative agreement number S283B050051. All rights reserved.

This work was originally produced in whole or in part by the National Comprehensive Center for Teacher Quality with funds from the U.S. Department of Education under cooperative agreement number S283B050051. The content does not necessarily reflect the position or policy of the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.

The National Comprehensive Center for Teacher Quality is a collaborative effort of ETS, Learning Point Associates, and Vanderbilt University.

0040_05/11

About the National Comprehensive Center for Teacher Quality

The National Comprehensive Center for Teacher Quality (TQ Center) was created to serve as the national resource to which the regional comprehensive centers, states, and other education stakeholders turn for strengthening the quality of teaching—especially in high-poverty, low-performing, and hard-to-staff schools—and for finding guidance in addressing specific needs, thereby ensuring that highly qualified teachers are serving students with special needs.

The TQ Center is funded by the U.S. Department of Education and is a collaborative effort of ETS, Learning Point Associates, and Vanderbilt University. Integral to the TQ Center's charge is the provision of timely and relevant resources to build the capacity of regional comprehensive centers and states to effectively implement state policy and practice by ensuring that all teachers meet the federal teacher requirements of the current provisions of the Elementary and Secondary Education Act (ESEA), as reauthorized by the No Child Left Behind Act.

The TQ Center is part of the U.S. Department of Education's Comprehensive Centers program, which includes 16 regional comprehensive centers that provide technical assistance to states within a specified boundary and five content centers that provide expert assistance to benefit states and districts nationwide on key issues related to current provisions of ESEA.

LEARNING POINT Associates[®]
An Affiliate of American Institutes for Research[®]





NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Laura Goe, Ph.D.

609-734-1076

lgoe@ets.org

**National Comprehensive Center for
Teacher Quality**

1100 17th Street NW, Suite 500

Washington, DC 20036-4632

877-322-8700 > www.tqsource.org

Reform Support Network

Great Teachers and Leaders: State Considerations on Building Systems of Educator Effectiveness

Spring 2011

This document was produced under U.S. Department of Education Contract No. ED-ESE-10-0-0087 with ICF International. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service or enterprise mentioned in this publication is intended or should be inferred.

Introduction

On this point education research is clear – effective instruction matters. Teachers are the single most important school-level influence on student achievement.¹ So it is no surprise that, with 43 states and the District of Columbia adopting college- and career-ready Common Core State Standards, and 45 states and the District of Columbia involved in Race to the Top assessment consortia, states and districts are looking to ensure that they have a workforce that can deliver on rigorous student performance expectations. Recently, there has been an unprecedented focus across the nation on developing *systems of educator effectiveness* – cultivating highly-effective teachers and leaders by reexamining and realigning a range of policies and practices for recruiting, developing, retaining, and rewarding teachers and principals.

As part of this focus on systems of educator effectiveness, states and districts are rethinking the ways they *evaluate* teachers by improving the processes and the tools they use for assessing teachers, in particular by making *student* performance a significant criterion among multiple measures of *teacher* effectiveness.

While *performance-based teacher evaluations* are the focus of this paper, there are other crucial pieces to a full-fledged system of educator effectiveness. Evaluation of principals is a critical component. Professional development that is tailored to address the particular needs of individual teachers and principals, including those identified through performance-based evaluations, is also important. Some experts suggest it is critical that policies for promotion, tenure, compensation, and dismissal also be connected to performance-based evaluations. Teacher and leader recruitment policies, the structure and content of teacher and principal preparation programs, and the requirements for entry into the profession also have the potential to shape an overall system of educator effectiveness.

Why start with a focus on teacher evaluations as part of a system of educator effectiveness? Many experts argue that performance-based teacher evaluations – evaluations that include student achievement results as a significant measure of teacher performance, and also include meaningful, regular observations of classroom practice, and timely and detailed feedback to teachers – are an important foundation for a comprehensive and coherent system of educator effectiveness that aims to raise student achievement. Recent research by The New Teacher Project suggests that today, teacher effectiveness “is not measured, recorded, or used to inform decision-making in any meaningful way.” Their report, *The Widget Effect*, found that across the nation, teacher evaluations fail to differentiate performance. The result is that current teacher evaluations provide little data or information that could be used to give teachers the training and tools they need to be effective, better identify and meet individual professional development

¹ Hanushek, Eric A., and Steve G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100(2):267–71; Rockoff, Jonah. 2003. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," Harvard University. Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS) Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(1), 299-311. Wenglinsky, H. (2000, October). How teaching matters: Bringing the classroom back into discussions of teacher quality. Princeton, NJ: The Milken Family Foundation and Educational Testing Service.

needs, provide targeted intervention to help struggling teachers, or reward the accomplishments of effective teachers.

The report that follows is an effort by the Reform Support Network to broadly share the key questions, themes, and challenges related specifically to the development of performance-based teacher evaluation systems discussed by technical experts and Race to the Top state grantees during the first six months of the Reform Support Network's Teacher and Leader Effectiveness Community of Practice. The group has explored a number of issues critical to states making sound initial choices about overarching teacher evaluation systems design, student growth models, measuring performance in non-tested grades and subjects, and teacher observation instruments.

Some Design Principles for Effective Teacher Evaluation Systems

- 1. All teachers should be evaluated annually.*
- 2. Evaluations should be based on clear standards of instructional excellence that prioritize student learning.*
- 3. Evaluations should consider multiple measures, with emphasis on a teacher's impact on student academic growth.*
- 4. Evaluations should employ four to five rating levels.*
- 5. Evaluations should encourage frequent observations and constructive critical feedback.*
- 6. Evaluation outcomes must matter; evaluation data should be a major factor in key employment decisions.*

-The New Teacher Project

Background

Race to the Top

The \$4.35 billion Race to the Top Fund represents an unprecedented federal investment in reform. The initial grants are supporting eleven states and the District of Columbia in their efforts to implement comprehensive, coherent, statewide education reform across four key areas:

- Adopting standards and assessments that prepare students to succeed in college and the workplace;

- Building data systems that measure student growth and success, and inform teachers and principals how to improve instruction;
- Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and
- Turning around their lowest-performing schools.

The Race to the Top program has fundamentally redefined the education landscape in America by providing resources to states to lead comprehensive reform. A total of 46 states and the District of Columbia submitted bold, comprehensive Race to the Top plans; of the 35 states that applied and did not receive funding, many are still moving forward with those state plans.²

When it comes to great teachers and leaders, the 12 Race to the Top grantees are working to develop comprehensive systems of educator effectiveness by adopting clear approaches to measuring student growth; designing and implementing rigorous, transparent, and fair evaluation systems for teachers and principals; conducting annual evaluations that include timely and constructive feedback; and using evaluation information to inform professional development, compensation, promotion, retention, and tenure decisions.

Race to the Top states are not the only states tackling these issues. A 2010 review of state teacher policies by the National Council on Teacher Quality (NCTQ) finds that 21 states are now requiring annual evaluations of all teachers and 16 states are requiring that student achievement be incorporated into teacher evaluations.³

As states across the nation continue their focus on increasing effectiveness, the Reform Support Network is committed to making information about these efforts – from all states – widely available as a means of helping states to offer mutual support, lessons learned, expertise, and resources to aid one another on the road to reform.

Through Race to the Top, states are working to:

“design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that differentiate effectiveness using multiple rating categories that take into account data on student growth...as a significant factor.”

For more information on Race to the Top see:

<http://www2.ed.gov/programs/racetothetop>

² See “Race to the Top Has Unique Role to Play in Reforming Schools for the Future” at <http://www.ed.gov/blog/2010/09/race-to-the-top-has-unique-role-to-play-in-reforming-schools-for-the-future>.

³ See National Council on Teacher Quality. 2010. *Blueprint for Change: National Summary* at www.nctq.org/stpy.

State Plans for Improving Teacher and Principal Effectiveness Based on Performance

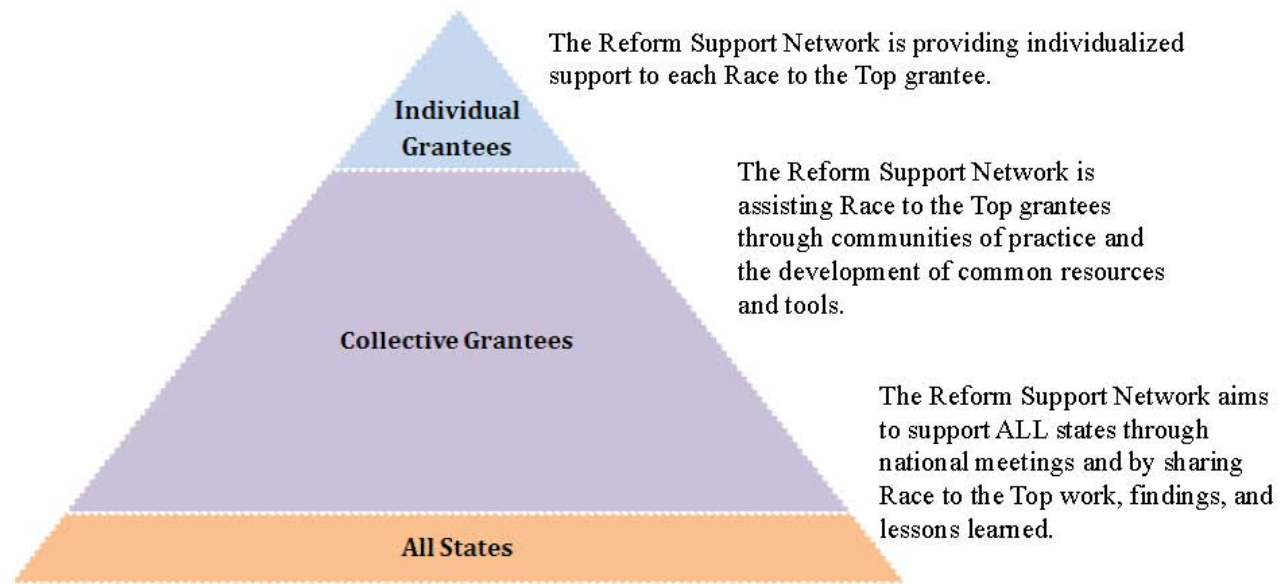
State Race to the Top applications were evaluated on the extent to which each state, in collaboration with its participating local educational agencies (LEAs), has a high-quality plan with ambitious yet achievable annual targets to ensure that participating LEAs:

- Establish clear approaches to measuring student growth and measure it for each individual student;
- Design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that (a) differentiate effectiveness using multiple rating categories that include data on student growth as a significant factor, and (b) are designed and developed with teacher and principal involvement;
- Conduct annual evaluations of teachers and principals that include timely and constructive feedback; as part of such evaluations, provide teachers and principals with data on student growth for their students, classes, and school; and
- Use these evaluations, at a minimum, to inform decisions regarding (a) developing teachers and principals, including by providing relevant coaching, induction support, and/or professional development; (b) compensating, promoting, and retaining teachers and principals, including by providing opportunities for highly effective teachers and principals to obtain additional compensation and be given additional responsibilities; (c) whether to grant tenure and/or full certification (where applicable) to teachers and principals using rigorous standards and streamlined, transparent, and fair procedures; and (d) removing ineffective tenured and untenured teachers and principals after they have had ample opportunities to improve, and ensuring that such decisions are made using rigorous standards and streamlined, transparent, and fair procedures.

- Race to the Top application criteria

Reform Support Network

The Reform Support Network is funded by the U.S. Department of Education to assist Race to the Top grantee states in implementing their comprehensive education reform plans. The Network is also committed to supporting all reform-minded states by widely sharing information on the kinds of education policies being adopted as part of Race to the Top.



The Reform Support Network's goal is to support Race to the Top by:

- Building capacity to execute and sustain reforms and continuously improve outcomes;
- Providing technical assistance to Race to the Top states;
- Facilitating collaboration across states; and
- Identifying and sharing promising and effective practices across states.

Communities of Practice

Communities of practice are “groups of people who share a concern, a set of problems, or passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis.”

-Wenger and Snyder

Teacher and Leader Effectiveness Community of Practice

As part of an effort to provide cross-grantee support, one of the strategies the Reform Support Network has adopted is to establish communities of practice (CoPs) to provide grantees with opportunities for cross-state learning and peer collaboration as well as to provide support to states from experts in the field. The purpose of the CoPs is to enhance state capacity for implementing reforms by providing for peer-to-peer learning, expert advice, model sharing, and collaboration on common needs.

In the fall of 2010, the Reform Support Network launched a Teacher and Leader Effectiveness Community of Practice to support RTT grantees in developing and implementing systems of educator effectiveness.

Initial work in this CoP has included collaborating to:

- Examine and develop practical approaches to measuring student learning using value-added models and student growth measures for the purpose of evaluating teacher performance;
- Explore the challenges related to ensuring rigor and comparability for measuring student growth in non-tested grades and subjects;
- Help states develop consistent, reliable, and appropriate teacher observation instruments for performance-based teacher evaluations; and
- Consider potential solutions and opportunities for state collaboration to address the challenges states face in designing and implementing comprehensive teacher and principal effectiveness systems.

This paper synthesizes some of the expert presentations and discussions among Race to the Top state grantees during six events listed below which occurred during the first six months of the Teacher and Leader Effectiveness CoP:

- November 10, 2010 webinar entitled “Getting the Math Right: Aligning Value-Added and Student Growth Models to State Policy Expectations,” which featured Dan Goldhaber, Ph.D.,

director of the Center for Education Data and Research (CEDR) and Daniel McCaffrey, Ph.D., a senior statistician at the RAND Corporation.

- November 17, 2010 webinar entitled “Non-Tested Grades and Subjects: Options for Measuring Student Growth,” featuring Robert Meyer, director of the Value-Added Research Center and professor at the Wisconsin Center for Education Research and William Slotnik, founder and executive director of the Community Training and Assistance Center.
- December 10-11, 2010 in-person convening on teacher and leader effectiveness, featuring presentations on teacher observation instruments and tools by a number of technical experts (see Appendix A), including Courtney Bell, Ph.D., Research Scientist at Educational Testing Service (ETS); Tim Daly, President of The New Teacher Project; and Dr. Tony Bryk, President of the Carnegie Foundation for the Advancement of Teaching.
- April 14, 2011 state peer webinar entitled “Measuring Student Learning in Educator Evaluation: Rhode Island’s Model Under Development,” led by state officials from Rhode Island.
- May 5, 2011 state peer webinar entitled “Value-Added and Student Growth Models: Operating Rules,” led by state officials from Tennessee.
- May 18-19, 2011 in-person meeting of the Reform Support Network’s Teacher and Leader Effectiveness Community of Practice, focused specifically on options and considerations related to measuring student growth in non-tested grades and subjects.

The considerations below aim to share broadly across *all* states some guiding questions and expert thinking on the early challenges and critical decisions facing states building performance-based teacher evaluations as part of systems of educator effectiveness. The topics include:

- 1) **Placing teacher evaluation design** in the context of state goals for framing a comprehensive educator effectiveness system;
- 2) **Choosing value-added and/or student growth models** to measure teacher impact on student achievement;
- 3) Developing **student growth measures in subjects and grades not covered by required statewide assessments**; and
- 4) **Choosing appropriate observation instruments** for teacher evaluations.

Considerations on Building Performance-Based Teacher Evaluations

Looking at the Big Picture: Framing the Design of Teacher and Leader Effectiveness Systems

Developing a system of educator effectiveness is a complex undertaking. There are many pieces that must fit together, from data systems for measuring student performance and linking teacher and student data, to teacher and principal evaluation, teacher and principal preparation, compensation, and professional development design.

According to Dr. Tony Bryk, President of the Carnegie Foundation for the Advancement of Teaching, and an expert on systems change, the first key step for a state is to develop clarity about goals, purposes, cost, and capacity, and to implement a comprehensive process that allows the state to engage stakeholders, develop a theory of action, plan and prepare for next steps.

Before focusing on any individual work streams or activities, it is beneficial for states to invest significant time and effort to identify and prioritize reforms that are most likely to improve the effectiveness of their teachers. This requires states to first consider the “big picture” view of the work and assess the ways that various state and district policies affecting teacher quality work together (or do not work together) to achieve a comprehensive educator effectiveness system.

Dr. Bryk observes that the rules and regulations now being developed by states and districts incorporate many technical details regarding the measurement of teacher performance through evaluations. These new rules and regulations will, in turn, shape practices in classrooms, schools, and district offices in profound and potentially unexpected ways. Success in improving student learning will depend on how effectively the field is able to understand and articulate the assumptions behind these rules and regulations, and integrate technical, regulatory, and practical considerations into a system that is continuously improving.

Bryk recommends that states consider the following questions as they begin a process of rethinking teacher evaluations in the context of a coherent and comprehensive educator effectiveness system:

- How do teacher evaluations **fit into the larger educator effectiveness system** or set of policies in place to ensure that the state recruits, develops, supports and retains highly-effective teachers?
- What are **the relationships, synergies, and dependencies** among teacher standards, teacher measures, teacher training, teacher qualifications, and policies that govern tenure, dismissal, promotion, compensation, and professional development? How coherent and aligned are these policies?

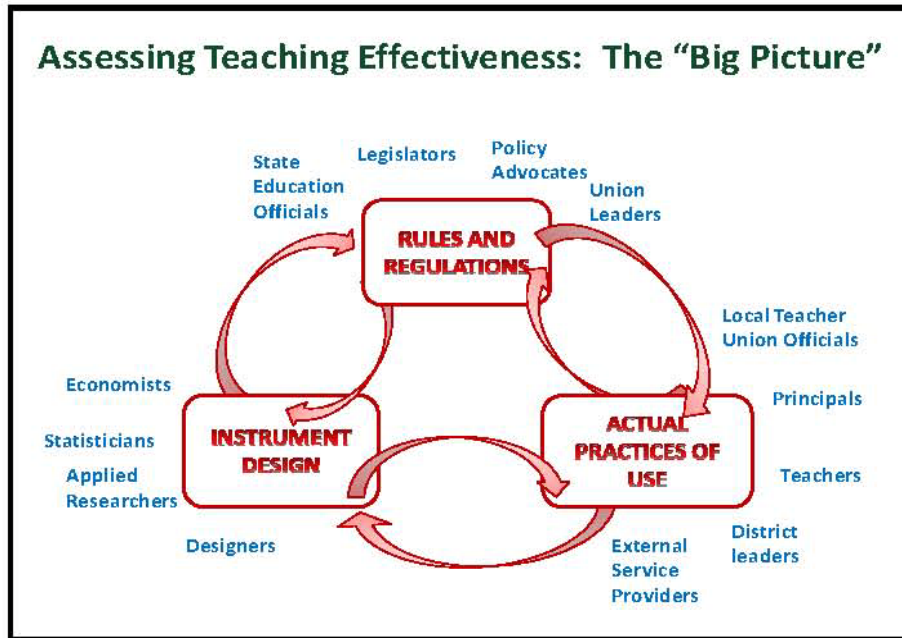
- What are the **desired outcomes** of the educator effectiveness system and what are the key policy levers that will lead to those outcomes?
- Given that the system is intended to do more than identify the lowest performing teachers, what are the **system strategies for improving teacher performance** more broadly?
- What do teachers, principals, and other leaders **believe about the ultimate purpose** of the teacher evaluations in particular and the state’s educator effectiveness system in general?

Experts note that it is critically important for states to provide time and opportunities for stakeholders to consider these important framing questions. With input from stakeholders on these issues, each state will be better prepared to articulate a theory of action that clearly describes its approach to developing performance-based teacher evaluations as part of a comprehensive system of educator effectiveness.

Framing Educator Effectiveness System Design

- *Maintain a “big picture” view of the work.*
- *Be clear how efforts to implement effectiveness systems and the need for better teacher evaluations are part of a larger effort to improve the professional and daily working life of teachers.*
- *Articulate the “theory of action” behind the state’s approach to teacher evaluation.*
- *Take an “improvement research orientation” to the work; that is, approach building teacher and leader effectiveness systems with flexibility and willingness to adjust policy based on experience, data, and feedback from research on reforms in action.*

Involve stakeholders – including teachers, state level education officials, legislators, union leaders, district and school leaders, experts and researchers – early and consistently.



Source: From a presentation titled, "Assessing Teaching: Seeing the Full Context," by Dr. Tony Bryk at the December 10-11, 2010 in-person convening on teacher and leader effectiveness.

Developing a theory of action. A state's answers to the framing questions above provide the framework for articulating a theory of action for how the state will design its teacher evaluations within the context of a broader educator effectiveness system. The theory of action unpacks the decisions for various aspects of the evaluation design – from purposes and goals to design and outcomes. Each of the decision points outlined along the way in the design and the development of teacher evaluations must align with a state's theory of action.

Framing teacher effectiveness. Experts broadly agree that policymakers and system developers also need to consider how ways they frame discussions of effectiveness will be received by teachers and school leaders and how new policies could affect their daily work. Bryk asserts that state and district policymakers "need to recognize that although challenges in the classroom and district are important, it is the school factors – the characteristics of the work environment – that have the most significant impact on whether a teacher remains in the profession. This is a critical consideration for educator effectiveness system design, given that many aspects of implementation occur in the school and affect teachers' work environments.

Adopting an improvement research orientation. Bryk also argues that it is important for states to take an "improvement research orientation" to the work of building teacher evaluations and educator effectiveness systems. States must also convey to all stakeholders, from the beginning, that development must be a process of continual improvement. It is important for stakeholders to acknowledge at the outset the importance of continuously reflecting on the progress of the work and making adjustments based on what is learned. Building mechanisms to evaluate progress from the beginning and setting a standard of continual improvement for the state's efforts may be crucial for success.

Assessing costs and capacity. Finally, Bryk recommends that states inventory and carefully consider the costs, resources, and capacity needed to make a performance-based teacher evaluation system and, more broadly, an educator effectiveness system, function well over time. This will require states, districts, and schools to identify new roles and responsibilities and plan for the intensive training and support that will be needed to carry out those new roles and responsibilities.

Focusing on Student Achievement: Choosing the Right Student Growth Model

Efforts to measure student growth and use data to inform education decision-making are not new. A few states have already implemented models to measure student growth for use in state accountability systems and adequate yearly progress determinations. But in designing and implementing performance-based teacher evaluations, many states now plan to go further, using measures of student growth as (a significant) one of multiple measures to inform their assessments of teachers and leaders.

With a big picture theory of action in place, there are numerous technical challenges that states face in developing teacher evaluations that are grounded in student performance. One of the key decision points that states face in developing performance-based teacher evaluations is choosing the method by which student achievement data are linked to individual teachers to make inferences about teacher performance and effectiveness.

In *Measuring Teacher Effectiveness Using Growth Models: A Primer*, the Reform Support Network discusses the important differences between value-added models and student growth percentile models.⁴ States must consider these differences carefully.

Value-added models (VAMs) are a specific type of growth model in the sense that they are based on changes in test scores over time. However, not all growth models are VAMs. VAMs specifically attempt to determine how specific teachers and schools affect growth in student achievement over time. VAMs are relativistic, addressing the question – to what extent can changes in student performance be *attributed to a specific school and/or teacher compared with that of the average school or teacher?* VAMs are complex statistical models that generally attempt to take into account student or school background characteristics in order to isolate the amount of learning attributable to a specific teacher or school. Teachers or schools that produce more than typical or expected growth are said to “add value.”

⁴ The text in this section is drawn from *Measuring Teacher Effectiveness Using Growth Models: A Primer*, which was prepared by the Reform Support Network and is available at: http://www.swcompcenter.org/educator_effectiveness2/Measuring_Teacher_Effectiveness_Using_Growth_Models.pdf

There are a variety of VAMs, but they can be categorized into the following major groups:

- **Gain score models:** measure year-to-year change by simply subtracting the prior year score from the current year score; typically the gains for all students for a given teacher are averaged.
- **Covariate adjustment models:** model current year test scores as a function of the prior year test scores and other student and classroom characteristics.
- **Layered models (including the persistence model):** model scores for multiple years in multiple subjects that may or may not include student background variables.

Student growth percentile models are another type of growth model that may be used to examine the contribution of teachers to student growth. In this model, a different type of statistical procedure is used to examine *changes in student achievement for individual students compared with other students*. This information is then aggregated to the teacher level to produce an estimate of the teacher’s impact on student learning.

Each of these approaches has different characteristics, advantages, and challenges. To date, there is no consensus among experts on the “best” model. Value-added analysis involves a series of steps from collecting and organizing data to deciding on a model and implementing and reporting on the results of that model. Each step is important and can have significant consequences for individual teachers and schools.

What experts do agree on is that selecting a model—whether VAM or student growth percentile model—involves states thinking carefully about what types of decisions will be made with the results and what model will provide the best information for these decisions. Not all models will necessarily produce equally useful information for each type of decision.

Ultimately, policymakers must ask themselves how credible and useful their model will be to their stakeholders and toward their student achievement goals. Issues states could think about in examining these models include the following:

- **Precision.** How precise is the model? How does the model account for error? How well will it differentiate among teachers? How likely is the model to misidentify teachers as highly-effective or ineffective (i.e., false positives or negatives)?
- **Validity.** What evidence is there that results from the model align with other measures of teacher effectiveness?
- **Data needs.** How much data and what type of data are needed to implement the model? Does the state have these data readily available? How does the model handle missing data? How does the model deal with student mobility?
- **Fairness and expectations for student achievement.** What factors or variables does the model include as controls? Why? How do the costs of including these controls (e.g.,

increasing model complexity, possible implications of differing expectations for different students, increasing possibility of more missing data) compare with the benefits (e.g., accounting for differences in student characteristics that are not attributable to teachers and schools, possibly improving precision)?

- **Stability and changes in estimates.** How many years of data are included for each estimate? Does the model estimate year-to-year change, or does it average information over multiple years? What are the tradeoffs in terms of precision and stability compared with the potential to see changes in estimates over time?
- **Comprehensibility of model and results.** How easy is the model to explain and describe to stakeholders, such as teachers? How are results typically presented, and how will they compare to measures of status or raw growth?
- **Cost.** How much will it cost to implement the model over time? Specifically, what are the upfront costs of development and the ongoing costs of system maintenance and improvement?
- **Ease of implementation and ownership.** What capacity (psychometric, software, etc.) is needed to implement the model? Can the state implement or “own” the model over time?
- **Alignment with other measures.** How does the model align with existing growth or value-added models in place in the state?
- **Usability.** How easily can data from the model be used along with other data to assess teacher or school leader practice?
- **Ability to aggregate to school level.** How can information from this model be used to help evaluate principals or other school leaders? Can the model or a different specification of the model estimate principal or school effects? How? What is the interpretation of these estimates?

With a strong understanding of the strengths, weaknesses, and trade-offs among various student growth models, states can then return to their basic theories of action to make choices about how to measure this key component of performance-based teacher evaluation. In addition, states need to begin to define the basic operating rules of the systems they are designing. To do this, states could:

Consider the policy outcomes. What decisions are the value-added or student growth models intended to inform? Different policy decisions, such as decisions about professional development, rewards, dismissal, tenure, or school accountability, require different data. States can backward map from their goals and their theory of action to help ensure that the purposes for which the state wishes to use data shape the specifications.

Engage stakeholders. Work with peers and technical experts to design a plan for engaging stakeholders in the selection of the model. Be transparent by documenting every step toward

decisions related to selection of a model so that decisions can be tracked, explained, and communicated publicly.

Ensure consistency and capacity. For assessment data that are comparable statewide, it is important that state and district data collections are consistent and that there is a capacity to generate the information needed for linking students and teachers. State trend data can also be used to inform decisions around cut-off points, and data from prior years can be useful in assessing the system before implementation. In addition, longitudinal data from other states can be used to inform decisions. States can also work with districts to gain insight into the system's potential strengths and limitations, especially in cases where the data will be used to inform decisions about individual teacher performance.

**A Tutorial on Assessing Student Performance and Growth
By the Value-Added Research Center**

For a useful and user-friendly tutorial on the similarities and differences among models for assessing student performance—attainment, growth, and value added—see Value-Added Research Center at <http://varc.wceruw.org/tutorials/>

The tutorial uses the analogy of gardeners growing trees to illustrate the different ways of looking at student achievement and how that information might be used to assess teacher performance. The tutorial could be a useful tool for engaging stakeholders in forums regarding how growth models can help inform student achievement and help build a system of educator effectiveness.

Take critical steps to ensure sustainability. It is important for states to plan and prepare for continual monitoring and refinement. Understanding state and local capacity is essential. Who will assume responsibility for managing different aspects of the measurement system? Who will be responsible for data quality and by what process will data quality be ensured?

Define the operating rules of the teacher evaluation system. What are the rules for using the chosen models, considering state and local data system capacity, and identifying the ways in which the models will contribute to measuring teacher effectiveness? For example, which students and what achievement data should count toward a teacher's performance evaluation? How much will growth measures count towards a teacher's overall evaluation? Will growth data for individual teachers be made public? How will student scores be attributed to teachers and schools? How will growth be measured for students who are highly mobile? How should the model account for student background characteristics?

Example: Some of Tennessee’s Operating Rules

Tennessee’s Value-Added Assessment System (TVAAS) has been in place since 1992. The state has statewide standardized Tennessee Comprehensive Assessment Program (TCAP) assessments in reading, math, science, and social studies in grades 3-8. In 2010, Tennessee implemented a new data dashboard and began training around TVAAS teacher effect scores with a plan for new annual teacher and principal evaluations for 2011-12. Operating rules include:

- *Teacher effect estimates are mandated by state statute;*
- *Annually, data from the TCAP tests, or their future replacements, will be used to provide an estimate of the statistical distribution of teacher effects on the educational progress of students within school districts for grades (3-8);*
- *A student must have been present for 150 days of classroom instruction per year or 75 days of classroom instruction per semester before that student's record is attributable to a specific teacher;*
- *The estimates of specific teacher effects on the educational progress of students will not be public record, and will be made available only to the specific teacher, the teacher's appropriate administrators as designated by the local board of education, and school board members; and*
- *Thirty-five percent of the evaluation criteria shall be student achievement data based on student growth data as represented by the TVAAS or some other comparable measure of student growth, if no such TVAAS data are available.*

For more information on Tennessee’s teacher evaluation plans, see
<http://www.tn.gov/firsttothetop/>.

Prepare a communication strategy. An effective communication strategy around incorporating student achievement into performance-based teacher evaluations will explain to all stakeholders the intent, goals, strengths, and limitations of value-added and student growth models. Most often, these models evaluate teachers on their contribution to student growth rather than overall proficiency, and recognize that there are factors contributing to student proficiency which are beyond a teacher’s control. Many experts argue that these models may be much fairer to teachers because they usually take a student’s background and prior performance into consideration – a fact often missing in recent debates on whether it is fair to judge teachers based on the performance of their students. States might consider conducting strategic planning with small groups of stakeholders to establish common agreements that can pave the way for constructive engagement with all stakeholders as work progresses.

Rigor and Comparability: Developing Growth Measures in Non-Tested Grades and Subjects

Value-added and other growth models are most readily suited to situations where standardized student assessment data are available. Statewide tested grades and subjects afford relatively large and robust data sets that can be used to measure changes in student academic achievement.

In most states, because of Elementary and Secondary Education Act (ESEA) assessment requirements, statewide data are readily available for many students and teachers in grades 3-8, as well as high school math and English/language arts.

However, even with these requirements, statewide standardized assessment data may not be available for more than half of the teachers in a given state. Thinking about the full complement of teachers – *including K-2, social studies, special education, non-core subject areas, and teachers of English Language Learners* – states face the challenge of how to develop fair, rigorous, and comparable measures of student growth and achievement that can be used to evaluate the performance of teachers for whom state standardized achievement data do not exist.

Given this, how should a state approach developing student growth measures in grades and subjects for which there are no statewide standardized assessments? When measuring student growth in “non-tested grades and subjects” (NTGS), other measures need to be used or developed.

The Reform Support Network’s *Measuring Student Growth in Non-Tested Grades and Subjects: A Primer* identifies three general approaches emerging from state and district practice as well as expert thinking in response to the challenge of measuring student learning in NTGS.⁵ It is important to note that these approaches are not mutually exclusive. It is likely that states and districts may want to use a variety of approaches to measuring student growth depending on the assessments available, the costs and benefits of each approach, and the contextual needs within the state. Examples of these approaches include:

- **Student learning objectives (SLOs)** are a participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction. SLOs can be based on standardized assessments, but they also may be based on teacher-developed assessments or other classroom assessments if they are “rigorous and comparable across classrooms.” When using SLOs, teachers set measurable expectations for student learning, usually in collaboration with their principal or other leader. SLOs can also be used in tested grades and subjects to help determine how predictive the measures of student growth are, and using them in all grades and subjects assures some comparability in methods.

⁵ The text in this section is drawn from *Measuring Student Growth in Non-Tested Grades and Subjects: A Primer*, which was prepared for the Reform Support Network and is available at http://www.swcompcenter.org/educator_effectiveness2/NTS__PRIMER_FINAL.pdf.

Student Learning Objectives: Rhode Island

In order to ensure that objectives are specific, measurable, and rigorous, evaluators must establish clear processes for setting them. Rhode Island, an example of a state using SLOs for NTGS, recommends that states consider how:

- District leadership and building administrators take time to establish a process of setting SLOs that ensures objectives are aligned to district and school goals.
- Processes should be established such that, at a minimum, teachers in a school who teach the same grade/subject have the same objectives and evidence (but may have different “targets” depending on their “baselines”).
- Eventually, teachers in different schools who teach the same content have similar objectives and comparable evidence.

Rhode Island also notes the importance of checks and balances when using SLOs to ensure that they are rigorous and evaluated objectively. This includes regular audits of principal evaluations of teachers and state training of a cadre of intermediary service providers (often experienced retired teachers and principals) to undertake evaluations.

- **New or existing measures of student growth** (including pre- and post-tests as well as performance and portfolio assessments) can be used to measure student growth in non-tested grades and subjects. These measures may include early reading assessments; end-of-course assessments; and benchmark, interim, or unit assessments. Other assessments may be developed at either the state or district level.

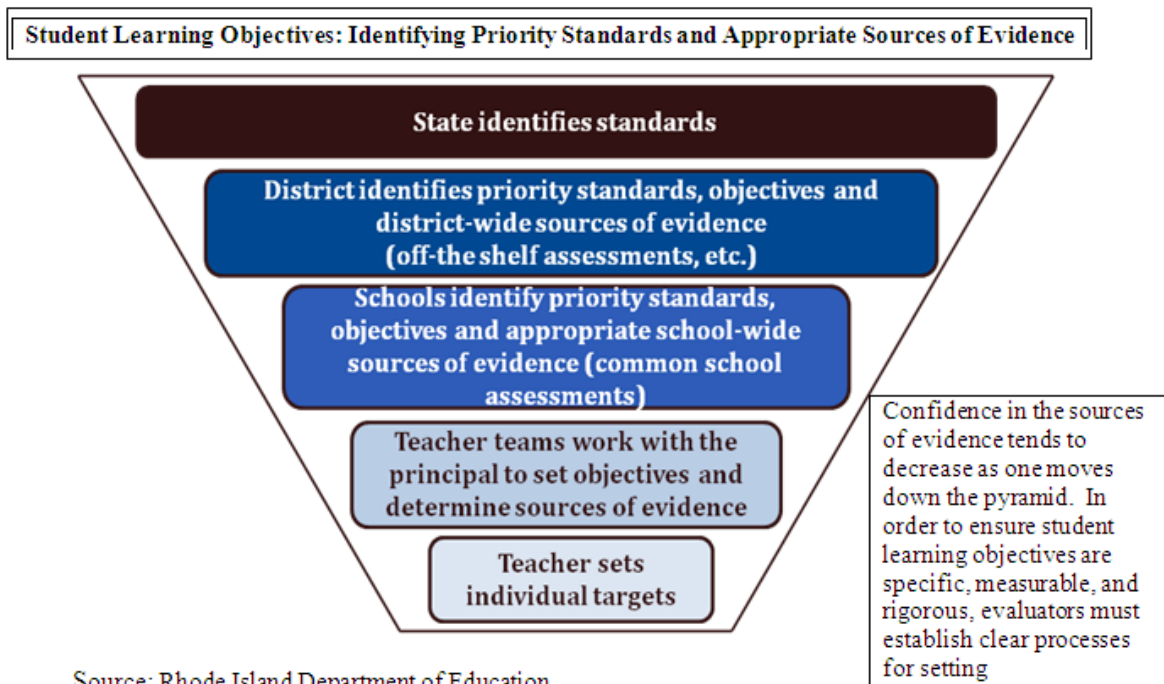
State discussions within the Teacher and Leader Community of Practice (CoP) make it clear that for some states, new options for measuring growth, such as those listed above, are critical. A number of states are interested in working towards new approaches to measuring student growth for grades K-2, social studies, special education and for English language learners. Options that states are exploring include development of early reading assessments aligned with the Common Core State Standards or progress monitoring protocols in the early grades. Some states see opportunities for collaboration across states in assessing middle school subjects such as social studies.

In each case, a goal for new assessment options is to increase the amount of comparable student learning data available for use in a broader system of educator effectiveness that differentiates and tailors professional development and improves student outcomes.

- **Measures of collective performance** assess the performance of the school, grade, instructional department, teams, or other groups of teachers. These measures can take a variety of forms including school-wide student growth measures, team-based collaborative achievement projects, and shared value-added scores for co-teaching situations.

As states consider selecting an approach to measuring teacher effectiveness in grades and subjects for which there are no statewide standardized assessment data, they are engaged in a process of weighing the costs and benefits of different options.

Similar to the process for considering value-added and growth models, states are examining the desired policy outcomes of the models selected; working with peers and technical experts to devise plans to engage stakeholders in the selection of the approaches for non-tested grades and subjects; defining operating rules; considering state and local data system capacities; identifying the ways in which the models will contribute to measuring teacher effectiveness; ensuring that state and district data collections are consistent; planning and preparing for continual monitoring and refinement; and preparing communication strategies that explain to educators and the public the intent, goals, strengths, and limitations of student growth modeling in non-tested subjects and grades.



In devising growth measures to assess teacher performance in subjects and grades for which no state standardized data are available, experts recommend that states consider:

- **Using existing assessment tools that are already available and appropriate** for this purpose. The Center for Educator Compensation Research is “in the process of developing a census of assessments being provided by all US states and selected districts. [Its] goal is to

capture innovative approaches these education agencies have taken to implement assessments in grades, subjects, and languages, not required under [ESEA].”⁶ This resource may provide useful information on available assessments that could be adopted or adapted for state or district use in current non-tested grades and subjects. Of course, states would have to carefully examine whether proposed assessments are aligned with state standards and serve their own goals.

- **Working strategically with vendors** to select or create a test bank of assessment items for current non-tested grades and subjects that are aligned with state standards and are appropriate for the purposes for which they will be used.
- **Identifying opportunities for collaboration** with other states, such as sharing item banks, determining best practices, and creating common assessments.
- **Engaging teachers in developing new assessments** for measuring student growth in non-tested grades and subjects. Many subjects have standards to guide assessment development, and national organizations often offer assessments for specific subjects. Teachers can bring a wealth of knowledge and ready examples to these discussions.
- **Incorporating predictors into the value-added or student growth model** for all grades and subjects. States may consider using secondary predictors such as college entrance exam scores or other tests, in addition to pre-tests, for a specific subject, to determine predicted growth.
- **Providing support for districts** on developing consistent measures or rigorous student learning objectives. States might consider developing models and prototypes of learning standards, sharing exemplars on their websites, as well as providing written guidance and decision-making tools to districts on the state’s standards for building student growth expectations for subjects and grades for which no statewide data are available.
- **Maintaining quality control** at the state level for how growth is measured in NTGS. Some states are requiring districts to submit their plans and methodology for developing growth measures for NTGS. By having a vetting process or state audit in place, states can help ensure that districts make good faith efforts to measure teacher performance in NTGS in a fair and reliable manner. States have some tools at their disposal for considering whether locally developed or proposed NTGS measures are defensible. The first is to use standardized statewide measures as a basis for comparison. To what degree do judgments about teacher performance in NTGS resemble the pattern in teacher performance on standardized statewide measures? Second, does the NTGS measure result in differentiation – does it result in a distribution of teacher performance which, at the very least, distinguishes between the best and worst teacher performance across the spectrum of teachers evaluated?

⁶ See <http://www.assessmentsurvey.wceruw.org/>.

**Teacher and Leader Community of Practice:
Considerations for Student Growth Measures in Non-Tested Grades and Subjects**

- *Are existing assessments consistent and comparable? Do they allow for measurement of student progress over time?*
 - *Should existing tests be used? Will additional assessments or new measures of student learning need to be developed?*
 - *What process will states use to ensure that locally developed measures of student growth are credible and reliable measures of teacher effectiveness?*
 - *What strategies need to be developed to ensure that there is meaningful engagement with stakeholders?*
 - *What are the costs associated with the various approaches?*
 - *What are the data capacity requirements for measuring growth in non-tested grades and subjects?*
-

Regardless of the strategies states pursue to measure growth in NTGS, experts emphasize that it is important for states to *prioritize*. It is important for states to develop fair and reliable student growth measures for which statewide assessment data *are* available. After that, states and districts can and should prioritize based on enrollment counts, number of teachers by subject and grade level, as well as the availability of consistent student achievement measures, to address the student achievement portion of teacher evaluations in non-tested grades and subjects.

Moving towards comparable measures of student growth to use in teacher evaluations for NTGS is important, and getting comparability within a district for teachers within the same grade and subject area is itself a substantial accomplishment – but it is not the only goal. Experts note the importance of ensuring that measures used to evaluate teachers are *rigorous* and *fair* for every teacher.

One consistent theme in state discussions on this issue is how important *transparency* is in developing measures of teacher performance – a point that is true for all aspects of teacher evaluations. In particular, where there may be critical questions about comparability, it is essential that states and districts can make clear cut cases for the fairness, the rigor, and the appropriateness of the measures chosen for evaluation.

Example: Delaware's Teacher Evaluation System

Delaware has had a statewide educator evaluation system since the 1980s. The state's current evaluation system, the Delaware Performance Appraisal System (DPAS) II, has been in use since 2008. It includes three versions, one for administrators, one for teachers, and one for specialists.

DPAS II for teachers and specialists is based on Charlotte Danielson's *Enhancing Professional Practice: A Framework for Teaching (2nd Edition)*, while DPAS II for administrators is based on the Interstate School Leaders Licensure Consortium's (ISLLC) standards for leaders.

For all educators, DPAS II defines standards for professional practice along five components: 1) planning and preparation, 2) classroom environment, 3) instruction, 4) professional responsibilities, and 5) student improvement. For each of the first four components, there is a set of four appraisal criteria. Each criterion has a rubric defining "unsatisfactory," "basic," "proficient," and "distinguished" performance.

Evidence for performance on components 1, 2, and 3 for teachers and specialists is gathered through observations by administrators trained in assessment. Evidence for performance on components 1, 2, and 3 for administrators is gathered through a survey completed by professional staff, the administrator's self-assessment on the ISLLC standards, and the assessor's survey data. For the fourth component, all educators complete a professional responsibilities form, which details their professional growth, communication with students, parents, and school colleagues, and their contributions to the professional community during the review period.

To receive a "satisfactory" rating for each of the first four components, an educator must receive a satisfactory ("basic," "proficient," or "distinguished") on at least three of the four criteria specified in the component.

Under Delaware's recently revised regulations, beginning in July 2011, a satisfactory rating for the fifth component (student improvement) means that the teacher has met the standard for student growth. That standard will represent an appropriate level of change in achievement data for an individual student between two points in time, as well as any other measures that are determined to be rigorous and comparable across classrooms.

Currently, assessments can result in summative ratings of "effective," "needs improvement," or "ineffective." Under the revised regulations, Delaware will add a fourth summative rating of "highly effective" in July 2011. Educators will be required to demonstrate satisfactory levels of student growth to receive an "effective" rating, and more than a year of student growth to receive a "highly effective" rating.

For more information on Delaware's Race to the Top plan, see
<http://www.doe.k12.de.us/rttt/>

Observing Teacher Practice: Choosing Appropriate Classroom Observation Instruments

Teacher evaluations have historically used some kind of classroom observation to obtain information about teacher effectiveness. Observations can provide important information that can be used to support professional growth, improve performance, and make decisions about compensation, employment, and other aspects of an educator effectiveness system. As states are assessing their current tools or considering new options, they should align the purposes and methods of observation with the expectations of their educator effectiveness system, and build in methods to assess inter-rater reliability among the individuals tasked with conducting teacher observations.

In choosing observation instruments to incorporate into performance-based teacher evaluations, experts in the field urge states to:

- **Clarify the purpose and objectives** of the teacher evaluation system and how observation instruments can help meet those objectives;
- **Evaluate the rigor, quality, and utility** of observation instruments under consideration;
- **Engage stakeholders**, including teachers and principals, in the design and selection of observation instruments; and
- **Provide professional development** to principals, teachers, and other raters to ensure that observation instruments are implemented with fidelity.

Teacher and Leader Community of Practice: Challenges Related to Observation Instruments

- *Ensuring that the observation instruments are aligned with the purposes and methods of the teacher evaluation system;*
 - *Ensuring that teacher evaluations are conducted in a consistent and reliable manner by evaluators;*
 - *Helping evaluators provide high-quality feedback to teachers that will help them improve their teaching; and*
 - *Establishing a common language on instructional practice to help district leaders develop consistent and effective professional development for teachers.*
-

Courtney Bell, Research Scientist for Educational Testing Service, recommends that states consider the research-based observation instruments already in use, including, for example:

Instrument	Developer(s)	Subject Area(s)	Grades
Framework for Teaching (FFT)	Charlotte Danielson	All	K-12
Classroom Assessment Scoring System - Secondary (CLASS-S)	Bob Pianta & Bridget Hamre	All	6-12
Mathematical Quality of Teaching (MQI)	Heather Hill, et al.	Math	4-12
Protocol for Language Arts Teaching Observations (PLATO)	Pam Grossman, et al.	English/Language Arts	4-12
Quality of Science Teaching (QST)	Ray Pecheone, et al.	Science	6-12

The table above includes some examples of existing instruments but is certainly not exhaustive. Bell provides several important indicators for assessing the quality of teacher observation protocols, applicable to any teacher observation instrument:

- There is a **clear articulation of score use**.
- There are **meaningful and observable differences** between score points.
- The **inferences required of the rater can be made reliably**, given training and support.
- There is **validity evidence** to support use of the instrument.
- **Teachers understand the scales** and score point distinctions.
- **Raters score consistently and accurately** at acceptable levels (~80%).
- Observational score and other quality indicator **comparisons make sense**.

Tim Daly, President of The New Teacher Project, also suggests that states consider existing instruments for observing teachers as part of performance-based teacher evaluations, but cautions against individual states or districts making modifications that alter the psychometric properties and validity of the instruments. He also suggests that in order for states to determine whether the observation instrument, its criteria, and its tools would contribute to accurate evaluation results, four key questions should be answered:

- **Is the instrument grounded in what matters to student achievement?** Does the instrument consider the classroom performance areas most connected to student outcomes, such as lesson objectives; strategies, activities, and delivery; physical environment;

classroom management and leadership; student engagement and real-time assessment; end of class assessment; and student mastery of lesson objectives?

- **What expectations does the instrument set?** Does the instrument set high performance expectations for teachers or outline only minimally acceptable performance?
- **Are the performance expectations for teachers unambiguous and precise?** That is, are the performance expectations clear enough that they leave little room for interpretation, telling observers exactly what to look for, or are they vague and general, leaving too much room for interpretation?
- **Is the instrument student-centered?** Does it require evaluators to look for direct evidence of student engagement and learning? Some observation tools focus only on the teacher's skills and behavior, without also including a focus on student response and impact, as well.

Experts also note that while implementing observation systems with fidelity requires significant time and expense, technology has the potential to ease costs and other challenges. Video databases, for example, are a potentially important emerging technology for evaluator and teacher training, evaluator (re)calibration, professional development, and principal workload management.

Conclusions and Looking Forward

Teacher evaluation is just one of several critical areas to consider in developing a coherent and aligned system for educator effectiveness –this paper only begins to explore some of the key issues and challenges states face in designing teacher evaluations that can identify varying levels of instruction and provide actionable information on improving teacher practice and ultimately, student achievement. Both state experience and expert advice on implementing performance-based teacher evaluations and systems of educator effectiveness suggest the following emerging guidelines:

- Ultimately, performance-based teacher evaluations are meant to be part of an overall educator effectiveness system **dedicated to improving student learning outcomes.**
- Educator effectiveness systems should be **built with the intention of improving individual and collective practice**, and facilitating the overall growth of the workforce of teachers and leaders. The information developed by and used in these systems should identify both strengths and weaknesses of individual teachers; provide rich information about students; be provided in a timely, user-friendly format to teachers and school leaders; and be used as the basis for policy and personnel decisions designed to improve student and school performance.
- As states move forward they are addressing **how teacher performance measures and instruments** align with teacher policy in other important aspects of a comprehensive system of educator effectiveness, including areas such as preparation, recruitment, tenure, promotion

policy, professional development, and policies focused on ensuring an equitable distribution of effective teachers across schools.

With so many states engaged in this complex and technical work simultaneously, the Reform Support Network has a special focus on helping Race to the Top states identify opportunities to collaborate with each other and on sharing those lessons broadly. In a recent meeting of the Teacher and Leader Effectiveness Community of Practice, a variety of Race to the Top states identified the following areas as priorities for potential state collaboration and solution design:

- Guidance on assessment design, especially for district, school, or classroom measures that might be used for evaluation in non-tested grades and subjects;
- Development of valid and reliable measures of student growth for kindergarten through second grade;
- Development of a cross-state item bank (particularly for non-tested grades and subjects); and
- Building a generic state framework to guide processes for developing and implementing student learning objectives.

In addition, the Reform Support Network is exploring ways that states can overcome resource barriers to reforms through cost-sharing and group purchasing strategies. Convening Race to the Top states to discuss solutions to common technical challenges is just one approach of the Teacher and Leader Community of Practice. The Network will continue to pursue opportunities to assist grantees in improving educator effectiveness and student achievement by providing resources that will help address challenges collaboratively and effectively. As the work progresses, the Reform Support Network will continue to listen actively to state needs and will regularly share key learnings and best practices with all states.

Appendices

Appendix A – List of Technical Experts Participating in the Teacher and Leader Effectiveness Community of Practice

Courtney Bell, Educational Testing Service
Tony Bryk, Carnegie Foundation for the Advancement of Teaching
Steve Cantrell, Bill & Melinda Gates Foundation
Tim Daly, The New Teacher Project
Ben Fenton, New Leaders for New Schools
Laura Goe, Educational Testing Service
Dan Goldhaber, Center for Education Data and Research, University of Washington
Brian Gong, National Center for the Improvement of Assessment and Harvard University
John Hussey, Battelle for Kids
Thomas Kane, Bill & Melinda Gates Foundation
Richard Laine, The Wallace Foundation
David Lussier, Austin Independent School District
Dan McCaffrey, RAND Corporation
Robert Meyer, Value-Added Research Center, University of Wisconsin
Richard Pennington, Scope Vision
Bill Slotnik, Community Training and Assistance Center (CTAC)
Chris Thorn, Value-Added Research Center, University of Wisconsin

Appendix B – Relevant Resources

- Community Training and Assistance Center. 2008. *CMS Student Learning Objective Guide*. Boston, MA: Author. Retrieved January 20, 2011, from http://gse.berkeley.edu/research/pace/reports/altcomp/Smith_Student_Learning_Guide.pdf.
- Community Training and Assistance Center. 2004. *Catalyst for Change: Pay for Performance in Denver Final Report*. Boston, MA: Author. Retrieved January 20, 2011, from <http://www.ctacusa.com/PDFs/Rpt-CatalystChangeFull-2004.pdf>.
- Goe, L., C. Bell, and O. Little. 2008. *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>.
- Holdheide, L. R., L. Goe, A. Croft, and D. J. Reschly. July 2010. “Challenges in evaluating special education teachers and English language learner specialists.” *TQ Research & Policy Brief*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/July2010Brief.pdf>.

- Little, O., L. Goe, and C. Bell. 2009. *A Practical Guide to Evaluating Teacher Effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/practicalGuide.pdf>.
- Milanowski, A. T., G. G. Henemann, and S. M. Kimball. 2009. *Review of Teaching Performance Assessment for Use in Human Capital Management*. Madison, WI: Strategic Management of Human Capital-Consortium for Policy Research in Education. Retrieved January 20, 2011, from <http://www.smhc-cpre.org/download/69/>.
- National Comprehensive Center for Teacher Quality. January 20, 2011. *Guide to Teacher Evaluation Products*. Retrieved from <http://www3.learningpt.org/tqsource/GEP/>.
- National Council on Teacher Quality. 2010. *Blueprint for Change: National Summary* at www.nctq.org/stpy.
- The New Teacher Project. 2010. Evaluation 2.0 at <http://tntp.org/publications/issue-analysis/view/teacher-evaluation-2.0/>.
- The New Teacher Project. 2009. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. See <http://widgeteffect.org/>.
- Steiner, L. October 2009. "Determining Processes That Build Sustainable Teacher Accountability Systems." *TQ Research & Policy Brief*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/October2009Brief.pdf>.
- Wenger, E.C., R. McDermott, and W.M. Snyder. 2002. *Cultivating Communities of Practice*. Cambridge, MA: Harvard Business School Press.